



Juliana Fernandes da Costa Macedo

**Um modelo Poisson-Lognormal para previsão da
quantidade IBNR via micro-dados**

Dissertação de Mestrado

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da PUC-Rio como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Prof. Álvaro de Lima Veiga Filho

Rio de Janeiro

Abril de 2015



Juliana Fernandes da Costa Macedo

**Um modelo Poisson-Lognormal para previsão da
quantidade IBNR via micro-dados**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre 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. Álvaro de Lima Veiga Filho
Orientador

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Cristiano Augusto Coelho Fernandes
Departamento de Engenharia Elétrica – PUC-Rio

Dr. Rodrigo Simões Atherino
JGP Global Gestão de Recursos

Prof. Eugenio Kahn Epprecht
Departamento de Engenharia Industrial – PUC-Rio

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico

Rio de Janeiro, 13 de Abril de 2015

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

Juliana Fernandes da Costa Macedo

Graduou-se em Ciências Estatísticas pela Escola Nacional de Ciências Estatísticas, no Rio de Janeiro – RJ (2010). Trabalhou como estagiária no Instituto Brasileiro de Geografia e Estatística (IBGE). Experiência profissional de mais de cinco anos em análise estatística de dados, em especial na área de pesquisa de mercado.

Ficha Catalográfica

Macedo, Juliana Fernandes da Costa

Um modelo Poisson-Lognormal para previsão da quantidade IBNR via micro-dados / Juliana Fernandes da Costa Macedo; orientador: Álvaro de Lima Veiga Filho. – 2015.

81 f. : il. (color.) ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2015.

Inclui referências bibliográficas

1. Engenharia elétrica – Teses. 2. IBNR. 3. Chain Ladder. 4. Triângulo Run-off. 5. Métodos sem triângulo. 6. Distribuição log-normal truncada. 7. Estimador de máxima verossimilhança. I. Veiga Filho, Álvaro de Lima. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

Agradecimentos

Primeiramente, agradeço a Deus por escutar minhas preces, por me dar força em todos os momentos de desespero, pela minha família e por tornar possível a conclusão deste trabalho e de mais esta etapa da minha vida. E por conceder saúde a mim e àqueles que eu amo para que possamos comemorar esta vitória juntos.

Ao meu orientador prof. Álvaro Lima Veiga pelos conhecimentos passados, pelo tempo dedicado, pelo apoio, pelos momentos de descontração e por todo auxílio necessário para realização deste trabalho.

À Luciene Souza por todo suporte, ajuda, material compartilhado, todas as dicas, dedicação de suas preciosas horas para me auxiliar com os códigos e dúvidas por email, skype, etc. Sua ajuda foi essencial para conhecimento do tema e realização desta dissertação.

Agradeço muito à minha irmã e meus pais por todo incentivo, carinho e apoio em todos os momentos da minha vida. A cada dificuldade que surgia eles sempre estavam ao meu lado me ajudando e acalmando. Foram compreensivos com minhas ausências e muitas vezes deixaram de sair aos fins de semana só para me fazer companhia em casa. Deolinda, Luciano e Tatiana vocês foram fundamentais para que eu realizasse essa conquista e são essenciais na minha vida. Amo muito vocês!

À minha família, tios, tias, primas (principalmente Camila e Amanda), primos e avós que sempre acreditaram em mim e compreenderam meu afastamento em alguns momentos. Em especial ao meu avô Gonzaga que é o maior incentivador da educação que eu conheço, um grande exemplo de vida e que me dá muito orgulho e apoio com suas palavras.

A todos os meus amigos que torceram por mim e me apoiaram todas as vezes, até quando eu não acreditava mais que seria possível. Nicole, muito obrigada por me acalmar em todos os momentos de desespero, por sofrer junto comigo e estar sempre ao meu lado; aos amigos que estudaram comigo na PUC (Alberto) e a maioria na ENCE e que vão estar para sempre na minha vida Nayara, Fabiano, Gustavo, e aos demais colegas que participaram desta jornada. Aos amigos que torceram mesmo que de longe e sempre me disseram que ia dar tudo certo Karine, Thaís, Juliana Dias, Paula e outros.

Aos amigos e colegas de trabalho pelo incentivo para entrar no mestrado, dias de home office, liberação do tempo e palavras de apoio e ajudas que tornaram esse trabalho possível: Thiago (preocupado e incentivador), Robertinha (me escutando e me apoiando), Raphael (sempre acreditando em mim e me botando pra cima), Talita (sempre cobrando e me dando força), Bruna, Patrícia, Erick, Alessandro, Roberta e demais que ajudaram direta ou indiretamente.

Agradeço aos professores Cristiano Augusto Coelho Fernandes, Eugenio Kahn Epprecht e doutor Rodrigo Simões Atherino por terem aceitado participar da banca.

À Seguradora Líder, responsável pela administração do DPVAT, pela concessão dos dados para minha utilização, tornando possível gerar os resultados apresentados.

Aos professores, ao pessoal do Departamento de Engenharia Elétrica, à todos os funcionários da PUC-Rio, e à própria PUC-Rio, pelo conhecimento, apoio, auxílio e infraestrutura proporcionados.

À todos que infelizmente não citei nominalmente, mas que contribuíram de forma direta ou indireta para este trabalho.

Resumo

Macedo, Juliana Fernandes da Costa; Veiga Filho, Álvaro de Lima (Orientador). **Um modelo Poisson-Lognormal para previsão da quantidade IBNR via micro-dados**. Rio de Janeiro, 2015. 81p. Dissertação de Mestrado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

O principal objetivo desta dissertação é realizar a previsão da reserva IBNR. Para isto foi desenvolvido um modelo estatístico de distribuições combinadas que busca uma adequada representação dos dados. A reserva IBNR, sigla em inglês para “Incurred But Not Reported”, representa o montante que as seguradoras precisam ter para pagamentos de sinistros atrasados, que já ocorreram no passado, mas ainda não foram avisados à seguradora até a data presente. Dada a importância desta reserva, diversos métodos para estimação da reserva IBNR já foram propostos. Um dos métodos mais utilizado pelas seguradoras é o Método Chain Ladder, que se baseia em triângulos run-off, que é o agrupamento dos dados conforme data de ocorrência e aviso de sinistro. No entanto o agrupamento dos dados faz com que informações importantes sejam perdidas. Esta dissertação baseada em outros artigos e trabalhos que consideram o não agrupamento dos dados, propõe uma nova modelagem para os dados não agrupados. O modelo proposto combina a distribuição do atraso no aviso da ocorrência, representada aqui pela distribuição log-normal truncada (pois só há informação até a última data observada); a distribuição da quantidade total de sinistros ocorridos num dado período, modelada pela distribuição Poisson; e a distribuição do número de sinistros ocorridos em um dado período e avisados até a última data observada, que será caracterizada por uma distribuição Binomial. Por fim, a quantidade de sinistros IBNR foi estimada por método e pelo Chain Ladder e avaliou-se a capacidade de previsão de ambos. Apesar da distribuição de atrasos do modelo proposto se adequar bem aos dados, o modelo proposto obteve resultados inferiores ao Chain Ladder em termos de previsão.

Palavras-chave

IBNR; Chain Ladder; Triângulo Run-off; Métodos sem triângulo; Distribuição log-normal truncada; Estimador de máxima verossimilhança.

Abstract

Macedo, Juliana Fernandes da Costa; Veiga Filho, Álvaro de Lima (Advisor). **A Poisson-Lognormal model to forecast the IBNR quantity via micro-data**. Rio de Janeiro, 2015. 81p. MSc Dissertation - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

The main objective of this dissertation is to predict the IBNR reserve. For this, it was developed a statistical model of combined distributions looking for a new distribution that fits the data well. The IBNR reserve, short for “Incurred But Not Reported”, represents the amount that insurers need to have to pay for the claims that occurred in the past but have not been reported until the present date. Given the importance of this reserve, several methods for estimating this reserve have been proposed. One of the most used methods for the insurers is the Chain Ladder, which is based on run-off triangles; this is a format of grouping the data according to the occurrence and the reported date. However this format causes the lost of important information. This dissertation, based on other articles and works that consider the data not grouped, proposes a new model for the non-aggregated data. The proposed model combines the delay in the claim report distribution represented by a log normal truncated (because there is only information until the last observed date); the total amount of claims incurred in a given period modeled by a Poisson distribution and the number of claims occurred in a certain period and reported until the last observed date characterized by a binomial distribution. Finally, the IBNR reserve was estimated by this method and by the chain ladder and the prediction capacity of both methods will be evaluated. Although the delay distribution seems to fit the data well, the proposed model obtained inferior results to the Chain Ladder in terms of forecast.

Keywords

IBNR; Chain Ladder; Run-off triangle; Triangle free methods; Truncated log-normal distribution; maximum-likelihood estimation.

Sumário

1 Introdução	12
1.1. Revisão bibliográfica	13
1.2. Organização da dissertação	15
1.3. Contribuições	16
2 Metodologia	17
2.1. Método <i>Chain ladder</i>	17
2.1.1. Triângulo <i>run-off</i>	17
2.1.2. Estimador <i>Chain ladder</i>	19
2.1.3. Medidas de qualidade do ajuste	20
2.2. Métodos para micro-dados	21
2.2.1. Abordagem de Weissner usando a distribuição log-normal truncada	22
2.2.2. Modelo conjunto para a quantidade e os atrasos	28
2.2.2.1. Apresentação das variáveis do modelo	30
2.2.2.2. Estimação dos parâmetros do modelo integrado quantidade/atraso por máxima verossimilhança	32
2.2.2.3. Estimador da quantidade IBNR	37
2.2.2.4. Distribuição da quantidade IBNR	38
3 Análise dos dados	39
3.1. Descrição dos dados	39
3.2. Análise descritiva dos dados	40
3.3. Análise gráfica da distribuição de atrasos e frequência de sinistros	41
4 Resultados	46
4.1. <i>Chain ladder</i>	46
4.2. Abordagem Weissner	48
4.3. Modelo conjunto quantidade/atraso	56
4.4. Comparação dos métodos	65
5 Conclusão	70
Referências Bibliográficas	72
Apêndice	75
A1- Máxima verossimilhança log-normal	75
A2- Quantidade IBNR por período de ocorrência	79

Lista de Figuras

Figura 1: <i>Triângulo run-off</i>	18
Figura 2: Triângulo <i>run-off</i> com as valores a serem estimados (sombreado cinza)	18
Figura 3: Reserva da seguradora por ano de calendário.....	19
Figura 4: Triângulo <i>run-off</i> acumulado	19
Figura 5: Processo de desenvolvimento de um sinistro.....	22
Figura 6: Histogramas dos atrasos no aviso para os 9 dias selecionados	43
Figura 7: Histogramas do log dos atrasos no aviso para os 9 dias selecionados...	44
Figura 8: Normal qqplot dos dias selecionados.....	45
Figura 9: Triângulo de desenvolvimento <i>Chain ladder</i>	47
Figura 10: Atraso estimado – Abordagem Weissner.....	49
Figura 11: Parâmetro μ estimado - Abordagem Weissner	50
Figura 12: Parâmetro σ estimado - Abordagem Weissner	50
Figura 13: Valor esperado do atraso estimado x média amostral por dia de ocorrência (dados até 2011)– Abordagem Weissner	51
Figura 14: Quantidade observada x Quantidade estimada por dia corte até 2008 - Abordagem Weissner	52
Figura 15: Quantidade observada x Quantidade estimada por dia corte até 2009 - Abordagem Weissner	52
Figura 16: Quantidade observada x Quantidade estimada por dia corte até 2010 - Abordagem Weissner	53
Figura 17: Quantidade observada x Quantidade estimada por dia até 2011 - Abordagem Weissner	53
Figura 18: Gráfico de médias mensais da quantidade observada x quantidade estimada por dia (dados até 2011) – Abordagem Weissner	54
Figura 19: Gráfico de dispersão <i>Ultimate</i> x Observado (até 2011) – Abordagem Weissner	55
Figura 20: Atraso estimado por corte – Modelo conjunto.....	56
Figura 21: Parâmetro μ estimado – Modelo conjunto.....	57
Figura 22: Parâmetro σ estimado - Modelo conjunto.....	58

Figura 23: Atraso observado e atraso médio esperado para o corte 2008 - Modelo conjunto.....	59
Figura 24: Atraso médio observado e esperado (corte 2009)- Modelo conjunto ..	59
Figura 25: Atraso médio observado e esperado (corte 2010)- Modelo conjunto ..	60
Figura 26: Atraso médio observado e esperado (corte 2011)- Modelo conjunto ..	60
Figura 27: Quantidade observada x <i>Ultimate</i> e Gama estimado até 2008 – Modelo conjunto.....	61
Figura 28: Quantidade observada x <i>Ultimate</i> e Gama estimado até 2009 – Modelo conjunto.....	62
Figura 29: Quantidade observada x <i>Ultimate</i> e Gama estimado até 2010 – Modelo conjunto.....	62
Figura 30: Quantidade observada x <i>Ultimate</i> e Gama estimado até 2011 – Modelo conjunto.....	63
Figura 31: Gráfico de dispersão <i>Ultimate</i> x Observado (até 2011) – Modelo conjunto	64
Figura 32: Gráfico de médias mensais da quantidade observada x quantidade estimada por dia (dados até 2011) – Modelo Conjunto.....	65
Figura 33: Gráfico de previsões 3 passos à frente com dados até 2008	67
Figura 34: Gráfico de previsões 2 passos à frente com dados até 2009	67
Figura 35: Gráfico de previsão 1 passo à frente com dados até 2010	68

Lista de Tabelas

Tabela 1: Estatísticas descritivas	40
Tabela 2: Dias selecionados para os gráficos	42
Tabela 3: Quantidade observada e atraso médio dias selecionados	43
Tabela 4 – Medidas dos erros de previsão <i>Chain ladder</i>	47
Tabela 5: Erros de previsão do método Weissner por corte	55
Tabela 6: Erros de previsão por cada corte e IBNR estimado para 2011	65
Tabela 7: Comparação erros de previsão e IBNR estimado	66
Tabela 8: Erros de previsão acrescentando resultados obtidos por Souza (2013)	66
Tabela 9: Intervalos de previsão de 90% x quantidades observadas	69

Introdução

Esta dissertação tem como objetivo utilizar micro-dados (ou sinistros individuais) para realizar a estimação de um tipo de reserva de seguradoras. O nosso foco será a quantidade IBNR, sigla em inglês para “*Incurred But Not Reported*”, que é dada pela soma dos sinistros que ocorreram, mas que ainda não foram avisados à seguradora até a data presente. A constituição desta reserva é obrigatória para as seguradoras e dá a garantia de que a seguradora será capaz de honrar seus compromissos com os segurados. A SUSEP (Superintendência de Seguros Privados) é o órgão regulador desta reserva, responsável pelo controle e fiscalização dos mercados de seguro, previdência privada aberta, capitalização e resseguro e valores.

Além do IBNR, existe a ocorrência IBNER (*incurred but not enough reported*), que representa os sinistros que já foram avisados, mas o seu pagamento ainda não foi realizado ou concluído. Essas quantidades juntas formam a chamada IBNR abrangente para um dado ano de exposição. Neste trabalho estimaremos somente a IBNR pura, que é a provisão de sinistros ocorridos e não avisados.

Como os sinistros ainda não foram avisados torna-se necessária a previsão de uma estimativa mensal ou anual para este valor. É importante obter uma previsão mais próxima do valor real, uma vez que previsões inadequadas afetam a gestão interna das seguradoras, assim como na confiança de investidores e interferência de órgãos reguladores (Friedland, 2010).

Diversos métodos para estimação da quantidade IBNR já foram propostos. Muitos deles são aplicados aos dados no formato de triângulo *run-off*, onde agregam-se as informações dos sinistros de acordo com as datas de ocorrência e pagamento ou aviso do sinistro, entre outras opções. Utilizando os dados neste formato, é possível aplicar um dos métodos mais usados pelos profissionais desta área, *Chain ladder*.

No entanto, esta forma de agrupar os dados faz com que informações dos sinistros individuais sejam perdidas e estas poderiam auxiliar na estimação da reserva IBNR (Parodi, 2013). Por isso, a ideia deste trabalho é testar uma

metodologia para o cálculo de reserva usando os sinistros individuais e comparar com o tradicional método *Chain ladder*.

1.1

Revisão bibliográfica

Muitos são os trabalhos com métodos estudados para estimar esta importante reserva. Segundo uma bibliografia organizada por Schmidt em 2012, existem quase 800 itens entre monografias, artigos científicos e alguns livros. A referência mais antiga citada neste arquivo vem do algoritmo desenvolvido por Thomas F. Tarbell em 1934.

Diversas características a respeito dos métodos mais tradicionais no setor de seguradoras foram abordadas por Friedland (2010), onde são citadas algumas vantagens e desvantagens, organização dos dados, criação do triângulo de desenvolvimento, enfoques técnicos e análises quanto às utilizações destes métodos. Devido a sua simplicidade e razoável acurácia, o *Chain ladder* é provavelmente o método mais popular para estimação de reservas IBNR (Mack, 1993) e muitas são as referências a este método.

England & Verrall (2002) consideram em seu artigo uma grande variedade de modelos estocásticos de reserva que reproduzem o tradicional *Chain ladder*, como por exemplo, métodos bayesianos, a técnica Bornhuetter-Ferguson, entre outros. Eles incluem também as previsões e erros de previsão dos modelos e mostram exemplos de aplicação dos métodos. Os autores afirmam que os métodos estocásticos têm poucos usuários e que provavelmente a razão principal é o fato dos métodos tradicionais serem suficientemente acurados para o cálculo da melhor estimativa de reservas de sinistro, sendo assim desnecessário o uso de outros métodos. A vantagem principal da forma estocástica sobre a determinística é a possibilidade de se ter informações sobre a incerteza do estimador pontual da reserva IBNR.

Muitos artigos criticam o método *Chain ladder* por sua natureza determinística, sem um modelo estatístico que possibilite a obtenção da variabilidade de suas previsões. Sendo assim, muitos autores tentaram atribuir um modelo estatístico ao *Chain ladder*, mas não obtiveram a mesma estimativa deste método. Mack (1994) diz que uma aproximação estocástica log linear para o

Chain ladder tem sido usada por muitos autores, mas que os modelos são diferentes e que a aproximação log linear é muito mais volátil que o *Chain ladder*. Mack (1993) propõe, após algumas suposições e sem associar nenhuma distribuição específica, um cálculo para o desvio padrão da estimativa obtida através do método *Chain ladder* com base nos estimadores originais do método e sendo referência em muitos artigos que calculam a variabilidade do *Chain ladder*.

Atherino (2005) utiliza um modelo baseado em espaço de estados proposto por De Jong & Zehnwirth (1983) que alinha a estimação da IBNR com a teoria de séries temporais e se serve das formulas de previsão e filtragem do filtro de Kalman (1960). Em outro trabalho, Atherino (2008) aplica uma ordenação distinta ao *triângulo run-off*, empilhando os dados em linhas, gerando uma série temporal univariada com dados faltantes (as reservas a serem estimadas) e aplica modelos em espaço de estado e filtro de Kalman a dados reais de seguradoras. Os modelos em espaço de estado utilizados por Atherino apresentaram resultados superiores ao método *Chain ladder* quanto ao poder preditivo.

Costa (2010) aplica aos dados o *Chain ladder* clássico, modelos de regressão linear com heterocedasticidade e modelos de espaço utilizando a ordenação dos dados supracitada (Atherino, 2008). Os resultados obtidos pela análise de regressão e modelos em espaço de estado foram superiores ao *Chain ladder*. No entanto, não foi possível identificar entre estas abordagens uma que se destacasse em relação a outra na estimação de reservas IBNR.

Kubrusly (2005) apresenta alguns métodos existentes para a estimação de reserva: o método *Chain ladder*, o modelo proposto por Hertig (1985) e Jong (2004), modelo proposto por Doray (1996) e apresenta um novo estimador que utiliza, além dos sinistros ocorridos e avisados, as apólices sem ocorrência de sinistro até o momento da estimação, buscando assim uma estimativa mais completa ao agregar mais informação.

A utilização dos dados não agregados ao invés dos em triângulos *run-off* é algo que vem sendo discutido em vários artigos. Taylor (2003) afirma que a necessidade de agrupar os dados vinha da limitação computacional do passado, quando não era possível analisar os dados no formato não agregado. Com a evolução da capacidade computacional, England e Verrall (2002) também acreditam que se torna questionável examinar os dados agregados e assim perder

informações importantes que poderiam ser utilizadas para produzir uma melhor estimativa e ajudar a associar uma distribuição a esta reserva.

Norberg (1993), inspirado em Karlsson (1976), Arjas (1989) e Jewell (1989) propõe um processo Poisson marcado não homogêneo: os sinistros seguem um processo homogêneo e uma marca aleatória é associada a cada sinistro, representando seu desenvolvimento da ocorrência até o encerramento final. Posteriormente, muitos trabalhos seguem sua proposta, inclusive Antonio e Plat (2012) que utilizam um processo Poisson marcado com posição dependente para modelar a processo de solicitação dos sinistros, onde um ponto é instante de tempo de ocorrência de um sinistro e a marca associada é a combinação do atraso no aviso e o desenvolvimento do sinistro. Baseando-se em Guiahi (1986), Weissner (1978) e Kaminsky (1987), Parodi combina modelos independentes: a frequência de sinistros ocorridos, distribuição de atrasos no aviso desses sinistros e severidade dos sinistros; obtendo a estimativa da quantidade IBNR com base no número de sinistros avisados até a data atual e produz a estimativa da distribuição das perdas totais IBNR. Para obter somente a quantidade IBNR pura, não é necessário modelar a severidade, mantendo o foco na modelagem da frequência de sinistros e de atraso no aviso de sinistros. Weissner (1978) baseia-se no método de máxima verossimilhança aplicada aos dados observados, para estimar a distribuição do tempo decorrido entre a ocorrência e aviso do sinistro. Dada a maior ocorrência de atrasos curtos, Weissner assume que os atrasos observados são uma amostra incompleta de uma distribuição imatura e tendenciosa.

1.2

Organização da dissertação

No Capítulo 2 encontram-se as metodologias abordadas neste trabalho; na seção 2.1 está apresentado o método de *Chain ladder* assim como a estrutura dos dados em *triângulo run-off*, que é o formato usado em muitos métodos tradicionais para cálculo de reserva, inclusive o *Chain ladder*. Além disso, são apresentadas algumas medidas de previsão de erro para as estimativas que serão obtidas pelos métodos. Na seção 2.2 estão apresentados os métodos que utilizam os dados individuais de sinistros (não agrupados em triângulo *run-off*) para estimação da quantidade IBNR: a abordagem de Weissner (1978) que utiliza uma

distribuição truncada para estimar distribuição de atraso entre a ocorrência do sinistro e aviso do mesmo e o método de máxima verossimilhança para estimar os parâmetros e que aqui será adaptada para a distribuição log-normal truncada; e a adaptação do modelo usado por Souza (2013) que trabalha a distribuição conjunta da distribuição de atraso no aviso, distribuição do total de sinistros ocorridos em um período e a distribuição da quantidade de sinistros ocorridos e avisados até a última data observada de um certo período. O Capítulo 3 possui a descrição e análise dos dados; o Capítulo 4 apresenta os resultados da aplicação das metodologias citadas anteriormente. As conclusões e sugestões de trabalhos futuros estão no Capítulo 5.

1.3

Contribuições

- Proposta de um modelo estatístico para estimar a quantidade IBNR utilizando sinistros individuais (micro-dados) através de uma distribuição conjunta, onde o número de sinistros segue uma Poisson e o atraso no aviso segue uma distribuição lognormal. Este modelo permite estimar e acompanhar o desenvolvimento dos parâmetros da distribuição de atrasos e da de quantidades, além de associar uma distribuição à quantidade IBNR, o que não é possível através do *Chain ladder*. Também é possível associar outras variáveis ao modelo a fim de obter melhores resultados.

- Desenvolvimento de um método de estimação por máxima verossimilhança.

- Apresentação de um exercício de comparação com o método de *Chain ladder* baseado em dados reais.

2

Metodologia

Neste Capítulo serão expostas as metodologias utilizadas neste trabalho. Primeiramente apresenta-se o método *Chain ladder* que é baseado no triângulo de desenvolvimento e depois é apresentado o método aplicado diretamente aos micro-dados, sem uso do triângulo *run-off*.

2.1

Método *Chain ladder*

O uso do triângulo *run-off* provou ser uma poderosa arma para atuários e permite que se tenha uma visão geral de como o montante total de sinistros se desenvolve ao longo dos anos. O *Chain ladder* é o principal método que utiliza o triângulo *run-off* e é considerado um *benchmarking* na estimação de reserva IBNR.

2.1.1

Triângulo *run-off*

Este formato de dados é muito utilizado em diversos métodos de estimação de IBNR e por muitos atuários. Ele apresenta as informações sobre a data de ocorrência do sinistro e a data de aviso que são agrupadas em uma matriz de dupla entrada, onde as linhas se referem ao período de ocorrência do sinistro e nas colunas está o desenvolvimento do sinistro, neste caso dado pelo atraso entre a ocorrência e o aviso do sinistro. Os dados podem ser apresentados em frequência diária, mensal, trimestral, anual, etc conforme a necessidade dos usuários.

Ano de ocorrência (t)	Desenvolvimento (d)				
	0	1	2	...	J-1
1	$Q_{1,0}$	$Q_{1,1}$	$Q_{1,2}$...	$Q_{1,J-1}$
2	$Q_{2,0}$	$Q_{2,1}$	$Q_{2,2}$...	
..		
J-1	$Q_{J-1,0}$	$Q_{J-1,1}$			
J	$Q_{J,0}$				

Figura 1: Triângulo *run-off*

Na Figura 1, as quantidades $Q_{i,j}$ referem-se a quantidade de sinistros ocorridos no ano i mas que foram avisados j períodos de tempo depois e consequentemente, quando $J=0$ tem-se que não houve atraso no aviso no período. Por exemplo, o valor $Q_{1,2}$ refere-se ao número de sinistro que ocorreram no período de tempo 2 e que levaram 1 período de tempo para ser avisado.

A estimação de IBNR é dada pelo preenchimento da parte de baixo do triângulo (parte destacada na Figura 2), isto é, pela estimação da quantidade que já ocorreu, mas ainda não foi informada.

Ano de ocorrência (t)	Desenvolvimento (d)				
	0	1	2	...	J-1
1	$Q_{1,0}$	$Q_{1,1}$	$Q_{1,2}$...	$Q_{1,J-1}$
2	$Q_{2,0}$	$Q_{2,1}$	$Q_{2,2}$...	$Q_{2,J-1}$
..
J-1	$Q_{J-1,0}$	$Q_{J-1,1}$	$Q_{J-1,2}$...	$Q_{J-1,J-1}$
J	$Q_{J,0}$	$Q_{J,1}$	$Q_{J,2}$...	$Q_{J,J-1}$

Figura 2: Triângulo *run-off* com os valores a serem estimados (sombreado cinza)

O valor da quantidade IBNR da seguradora para um ano qualquer é dado pela soma dos valores da diagonal, como pode ser observado na Figura 3.

Ano de ocorrência (t)	Desenvolvimento (d)				
	0	1	2	...	J-1
1	$Q_{1,0}$	$Q_{1,1}$	$Q_{1,2}$...	$Q_{1,J-1}$
2	$Q_{2,0}$	$Q_{2,1}$	$Q_{2,2}$...	$Q_{2,J-1}$
...
J-1	$Q_{J-1,0}$	$Q_{J-1,1}$	$Q_{J-1,2}$...	$Q_{J-1,J-1}$
J	$Q_{J,0}$	$Q_{J,1}$	$Q_{J,2}$...	$Q_{J,J-1}$

Figura 3: Reserva da seguradora por ano de calendário

Outra forma de utilização do *triângulo run-off* é com os dados acumulados por linha, onde $A_{i,j} = \sum_{k=0}^{j-1} Q_{i,k}$, ou seja, é a quantidade de sinistros ocorridos no período i e avisado em até j meses (Figura 4).

Ano de ocorrência (t)	Desenvolvimento (d)				
	0	1	2	...	J-1
1	$A_{1,0}$	$A_{1,1}$	$A_{1,2}$...	$A_{1,J-1}$
2	$A_{2,0}$	$A_{2,1}$	$A_{2,2}$...	
...	
J-1	$A_{J-1,0}$	$A_{J-1,1}$			
J	$A_{J,0}$				

Figura 4: Triângulo *run-off* acumulado

2.1.2

Estimador *Chain ladder*

A estimativa do *Chain ladder* resulta da estimação das quantidades desconhecidas do *triângulo run-off*, os valores ainda não avisados para cada ano de ocorrência até o ano de desenvolvimento corrente. Neste método usa-se o triângulo acumulado e assim, a estimativa de IBNR obtida por este método será dada pela última coluna deste triângulo.

Como foi mencionado anteriormente, as quantidades são conhecidas para $1 \leq i \leq J$ e $0 \leq j \leq J - i$. Fazendo a suposição de que cada sinistro demora no máximo J meses para ser avisado, o IBNR será dado por:

$$\text{IBNR} = \sum_{k=1}^J (A_{i,k} - A_{i,k-i}).$$

O método de *Chain ladder* assume que a razão $\frac{A_{i,j}}{A_{i,j-1}}$ é uma variável aleatória com média f_j . Sendo assim os fatores de crescimento entre as colunas $j-1$ e j independem da linha i e desta forma a estimação dos fatores pode ser realizada de diversas formas, entre elas: a média, máximo, mínimo, mediana ou alguma outra estatística. Neste trabalho será utilizado o estimador que Mack, 1994 provou ser não tendencioso e de variância mínima, dado pela equação abaixo:

$$\hat{f}_k = \frac{\sum_{i=1}^{J-k} A_{i,k+1}}{\sum_{i=1}^{J-k} A_{i,k}}, \quad 1 \leq i \leq J \text{ e } 1 \leq k \leq J-1. \quad (2.1)$$

Após a estimação dos fatores, pode-se realizar a estimação dos valores desconhecidos do triângulo acumulado, para isto esses fatores são aplicados ao último valor observado de cada linha.

$$\hat{A}_{i,j-i+1} = A_{i,j-i} \hat{f}_{i,j-i+1}, \quad 1 \leq i \leq J. \quad (2.2)$$

$$\hat{A}_{i,j} = \hat{A}_{i,j-i} \hat{f}_{i,j}, \quad 1 \leq i \leq J, \quad J-i+2 \leq j \leq J-1. \quad (2.3)$$

A quantidade IBNR estimada para cada período de ocorrência é $\hat{Q}_i = \hat{A}_{i,j} - A_{i,j-i}$ para $1 < i < J$. Esta quantidade de sinistros será usada para calcular o valor da reserva total, dada por $\hat{Q} = \sum_{k=2}^J \hat{Q}_k$.

2.1.3

Medidas de qualidade do ajuste

Buscando comparar a qualidade de ajuste dos métodos para a estimativa do IBNR, os dados serão separados: a primeira parte será utilizada para estimar os parâmetros e realizar previsões parciais dos valores de IBNR para comparar com os valores observados na segunda parte. As medidas de qualidade de ajuste calculadas são: MAE (*Mean Absolute Error*, em inglês, ou erro médio absoluto), MAPE (*Mean Absolute Percentage Error*, em inglês, ou porcentagem do erro médio absoluto) e RMSE (*Root Mean Square Error*, em inglês, ou raiz do erro médio quadrático).

Seja n o último período de tempo onde há informação de aviso de sinistro na amostra e A_{n+t} a quantidade de sinistros informados no período $n+t$, ocorridos e avisados até n . Com base no *triângulo run-off* obtêm-se as estimativas das

quantidades de sinistros em $n-1$ períodos de aviso posteriores a n . Mantendo uma quantidade razoável de dados, serão retirados os dados dos T últimos períodos, que serão usados na comparação com os dados previstos. Sob essas condições, dado um horizonte $t=1, 2, \dots, T$ com $T=1, \dots, n-1$, tem-se:

$$A_{n+t} = \sum_{i=t+1}^n Q_{i,n-i+t+1} \quad (2.4)$$

$$\hat{A}_{n+t} = \sum_{i=t+1}^n \hat{Q}_{i,n-i+t+1} \quad (2.5)$$

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |A_{n+t} - \hat{A}_{n+t}| \quad (2.6)$$

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \left| \frac{A_{n+t} - \hat{A}_{n+t}}{A_{n+t}} \right| \times 100 \quad (2.7)$$

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (A_{n+t} - \hat{A}_{n+t})^2} \quad (2.8)$$

Em casos de divergência, o critério mais relevante a ser utilizado será o MAE, pois mantém a escala dos dados originais.

2.2

Métodos para micro-dados

Segundo Antonio e Plat (2012), o uso de dados individuais de seguro melhora a qualidade das previsões, permite uma modelagem mais flexível do processo de desenvolvimento de sinistros, pois outras informações covariadas podem ser incluídas e evita os problemas encontrados quando os triângulos são usados, como: problemas de robustez, valores zerados ou quedas bruscas na quantidade de sinistros em alguma célula do triângulo, entre outros.

O processo de desenvolvimento de um sinistro pode ser ilustrado pela Figura 5¹, que mostra o tipo de informação disponível com os micro-dados. O sinistro ocorre em algum período de tempo t_1 , então ele é avisado à seguradora em t_2 (provavelmente com algum atraso em relação à ocorrência) e um ou alguns pagamentos são realizados (t_3 à t_{k-1}) até o encerramento (t_k). Em alguns casos, dependendo da natureza do pedido, o processo pode ser reaberto (t_{k+1}) e novos pagamentos (t_{k+2} ...) serão feitos até o encerramento final (t_{k+n}) do sinistro.

¹ A sigla RBNS presente na Figura significa, em inglês, “Reported But Not Settled”, que são os sinistros que foram reportados, mas não liquidados.

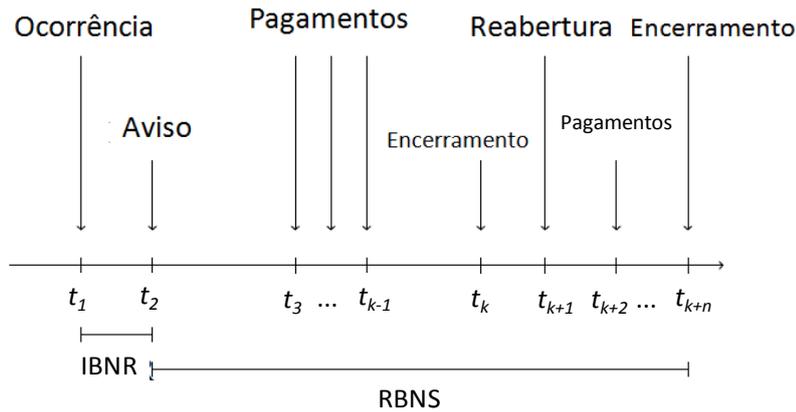


Figura 5: Processo de desenvolvimento de um sinistro

Grande parte dos métodos de cálculo de reserva que utilizam dados individuais segue a decomposição deste processo de desenvolvimento do sinistro desde a ocorrência, resolução (pagamentos) até a total finalização. De acordo com esse processo, o desenvolvimento de um sinistro pode ser modelado por uma distribuição de ocorrência de sinistros, uma distribuição de atraso na notificação do sinistro (tempo entre a ocorrência e o aviso), uma distribuição do desenvolvimento de pagamento de sinistros e uma distribuição de severidade.

A seguir será apresentada, primeiramente, a abordagem sugerida por Weissner (1978) para estimar a distribuição de atrasados, mas utilizando a distribuição log-normal truncada no lugar da exponencial. Depois, apresenta-se a proposta sugerida que é uma adaptação do modelo usado por Souza (2013), utiliza-se a mesma estruturação, mas adotando-se uma distribuição diferente para o atraso na notificação do sinistro, a distribuição log-normal.

2.2.1

Abordagem de Weissner usando a distribuição log-normal truncada

Weissner (1978) utiliza o método de máxima verossimilhança ao modelo que estima a distribuição do tempo entre a ocorrência e aviso do sinistro. Ele assume que os dados observados são originários de uma distribuição incompleta, imatura e tendenciosa da distribuição de atrasos, que prevalece atrasos curtos. E esta não é uma amostra aleatória, pois qualquer acidente com atraso maior que os observados no período estudado não consta na amostra.

Em seu trabalho, Weissner (1978) sugere que no lugar da distribuição exponencial, outras distribuições poderiam ser testadas, como por exemplo, a distribuição log-normal, a qual será testada neste trabalho. Ele usa os dados agrupados em um triângulo *run-off* e para cada período de tempo t , tem-se a quantidade de sinistros ocorridos naquele período (k) e os tempos de atrasos observados de cada sinistro ocorrido no período ($\tau_1, \tau_2, \dots, \tau_k$). Por exemplo, um sinistro que ocorreu em março e foi avisado em maio teve 2 meses de atraso no aviso ($\tau_i=2$). Supondo que a distribuição de atrasos é a log-normal com parâmetros μ e σ^2 desconhecidos, temos:

$$f_{\tau}(\tau/\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma\tau} e^{-\frac{1(\ln\tau-\mu)^2}{2\sigma^2}}, \tau > 0, \sigma^2 > 0 \quad (2.9)$$

Mas não se pode dizer que a distribuição de atraso no aviso está completa, pois para cada período de ocorrência t muitos sinistros podem ter ocorrido, mas ainda não terem sido avisados até a última data observada. Para cada período de ocorrência t só é possível observar atrasos inferiores a $n-t$, sendo n a quantidade máxima de períodos de tempo. Por exemplo, se os dados vão de janeiro a dezembro, teríamos $n=12$ e se considerarmos o mês de abril como período de ocorrência ($t=4$), o atraso máximo possível seria $T=n-t=8$ meses. Sendo assim, esses atrasos observados não seguem a distribuição log-normal, mas sim a distribuição log-normal condicionada (truncada) ao atraso máximo possível $T=n-t$ que é dada por:

$$f_{\tau/\tau \leq T}(\tau/\mu, \sigma^2, T) = f(\tau/\mu, \sigma^2) / P(\text{atraso no aviso} \leq T)$$

$$f_{\tau/\tau \leq T}(\tau/\mu, \sigma^2, T) = \frac{1}{\sqrt{2\pi}\sigma\tau} e^{-\frac{1(\ln\tau-\mu)^2}{2\sigma^2}} / \Phi\left(\frac{\ln T - \mu}{\sigma}\right) \quad (2.10)$$

Onde $\Phi(\cdot)$ é a função acumulada da distribuição normal padrão $N(0,1)$.

A função de verossimilhança para os parâmetros μ e σ^2 , $L(\mu, \sigma^2)$, no período de ocorrência t é dada por:

$$L(\mu, \sigma^2; \tau_1, \dots, \tau_k) = \prod_{i=1}^k \left\{ \frac{1}{\sqrt{2\pi}\sigma\tau_i} e^{-\frac{1(\ln\tau_i-\mu)^2}{2\sigma^2}} / \Phi\left(\frac{\ln T - \mu}{\sigma}\right) \right\}$$

Onde k é a quantidade de sinistros ocorridos naquele período. (2.11)

E a função de log verossimilhança, $l(\mu, \sigma^2)$, é dada por:

$$l(\mu, \sigma^2; \tau_1, \dots, \tau_k) = \sum_{i=1}^k \left\{ -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma^2) - \ln(\tau_i) - \frac{1}{2} \frac{(\ln \tau_i - \mu)^2}{\sigma^2} - \ln \left[\Phi \left(\frac{\ln T - \mu}{\sigma} \right) \right] \right\} \quad (2.12)$$

Para obter os estimadores de máxima verossimilhança para μ e σ^2 é necessário fazer a derivada da função de log-verossimilhança em relação a cada um dos parâmetros. Desta forma, a derivada da log-verossimilhança em relação a μ é:

$$\frac{\partial l(\mu, \sigma^2; \tau_1, \dots, \tau_k)}{\partial \mu} = \sum_{i=1}^k \left\{ \frac{(\ln \tau_i - \mu)}{\sigma^2} - \frac{\partial}{\partial \mu} \ln \left[\Phi \left(\frac{\ln T - \mu}{\sigma} \right) \right] \right\} \quad (2.13)$$

O segundo termo da equação acima requer alguns cálculos para ser obtido, os quais estão explicitados no Apêndice A1. Rearranjando os termos da expressão e igualando a derivada à zero, obtém-se a equação do estimador de máxima verossimilhança, onde não é possível isolar o parâmetro μ :

$$\mu = \frac{\sum_{i=1}^k \ln \tau_i}{k} + \sigma \left[\frac{\phi \left(\frac{\ln T - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)} \right] \quad (2.14)$$

Trata-se de uma equação transcendental em μ e assim sendo a sua solução requer o uso de um processo iterativo de análise numérica, como o Newton-Raphson, ou outros métodos numéricos. Para encontrar o estimador de máxima verossimilhança de σ^2 , o mesmo procedimento deve ser feito. A derivada da log-verossimilhança em relação a σ^2 é:

$$\frac{\partial l(\mu, \sigma^2; \tau_1, \dots, \tau_k)}{\partial \sigma^2} = \sum_{i=1}^k \left\{ -\frac{1}{2\sigma^2} + \frac{1}{2} \frac{(\ln \tau_i - \mu)^2}{\sigma^4} - \frac{\partial}{\partial \sigma^2} \ln \left[\Phi \left(\frac{\ln T - \mu}{\sigma} \right) \right] \right\} \quad (2.15)$$

Os cálculos também podem ser encontrados no Apêndice A1. Então igualando a derivada a zero e depois arrumando os termos, temos as expressões abaixo:

$$\frac{\partial l(\mu, \sigma^2; \tau_1, \dots, \tau_k)}{\partial \sigma^2} = \sum_{i=1}^k \left\{ -\frac{1}{2\sigma^2} + \frac{1}{2} \frac{(\ln \tau_i - \mu)^2}{\sigma^4} + \frac{1}{2\sigma^2} \left(\frac{\ln T - \mu}{\sigma} \right) \frac{\phi \left(\frac{\ln T - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)} \right\} = 0$$

$$\sigma^2 = \frac{\sum_{i=1}^k (\ln \tau_i - \mu)^2}{k - k \left(\frac{\ln T - \mu}{\sigma} \right) \frac{\phi \left(\frac{\ln T - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)}}$$
(2.17)

Como no estimador de μ , não é possível isolar o parâmetro σ^2 e métodos numéricos devem ser utilizados para estimar seu valor.

Os valores de μ e σ^2 estimados não são apenas os parâmetros da distribuição truncada, mas também da distribuição log-normal completa.

Estimação da quantidade total

Usando esta distribuição de atraso no aviso, é possível estimar a proporção de ocorrências de um período t que vai ter atraso maior, menor ou igual a um certo número de períodos de tempo. Por exemplo, a proporção de sinistros ocorridos em t que serão avisados após T , pode ser estimada por:

$$P(\tau > T) = 1 - \Phi \left(\frac{\ln T - \mu}{\sigma} \right)$$
(2.17)

A quantidade de sinistros comunicados, K , será determinada pela proporção de sinistros que ocorreram até o tempo T vezes a quantidade total de sinistros para o período de ocorrência t , U_t :

$$K = \Phi \left(\frac{\ln T - \mu}{\sigma} \right) U_t$$

$$E(K) = \Phi \left(\frac{\ln T - \mu}{\sigma} \right) E(U_t)$$

$$E(U_t) = \frac{E(K)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)}$$
(2.18)

Como só há uma observação de K , o valor k_t , a quantidade total de sinistros estimada para o período t é dada por:

$$\hat{U}_t = \frac{k_t}{\Phi\left(\frac{\ln T - \mu}{\sigma}\right)} \quad (2.19)$$

A quantidade esperada de sinistros ocorridos em t , mas que ainda não foram avisados até a data atual T é dada por:

$$Q_t = \frac{k_t}{\Phi\left(\frac{\ln T - \mu}{\sigma}\right)} - k_t = k_t \left(\frac{1}{\Phi\left(\frac{\ln T - \mu}{\sigma}\right)} - 1 \right) \quad (2.20)$$

E a estimativa da quantidade de sinistros que será avisada em qualquer período de tempo, como por exemplo, que será avisada em $T+1$:

$$\hat{Q}_{t,T+1} = \{P(\tau > T) - P(\tau > T + 1)\} \hat{U}_t \quad (2.21)$$

Distribuição de atraso no aviso para todos os dias da amostra

Também é possível usar toda informação presente na amostra para estimar os parâmetros, isto é, ao invés de calcular o estimador apenas para o período t , calcular usando todos os períodos da amostra. Sejam as n variáveis k_1, k_2, \dots, k_n os totais de sinistros avisados para cada período $t=1,2,\dots,n$; seja $T_t=n-t$, as n variáveis representando o atraso máximo observável para cada período de tempo t e por fim $\tau_{t,i}$ o i -ésimo atraso observado no período de ocorrência t . As amostras de atraso $\tau_{t,i}$ seguem $f(\tau/\mu, \sigma^2, T_t)$, a distribuição log-normal de atrasos truncada do período de ocorrência t . Supondo independência entre elas, a função de verossimilhança para todos os dados será:

$$L(\mu, \sigma^2; \tau_1, \dots, \tau_k) = \prod_{i=1}^{k_1} f(\tau_{1,i}/\mu, \sigma^2, T_1) \prod_{i=1}^{k_2} f(\tau_{2,i}/\mu, \sigma^2, T_2) \dots \prod_{i=1}^{k_n} f(\tau_{n,i}/\mu, \sigma^2, T_n)$$

$$L(\mu, \sigma^2; \tau_1, \dots, \tau_k) = \prod_{t=1}^n \prod_{i=1}^{k_t} \left\{ \frac{1}{\sqrt{2\pi}\sigma\tau_{t,i}} e^{-\frac{1(\ln\tau_{t,i}-\mu)^2}{2\sigma^2}} / \Phi\left(\frac{\ln T_t - \mu}{\sigma}\right) \right\} \quad (2.22)$$

E a função de log-verossimilhança é dada por:

$$l(\mu, \sigma^2; \tau_{t,i}) = \sum_{t=1}^n \sum_{i=1}^{k_t} \left\{ -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma^2) - \ln(\tau_{t,i}) - \frac{1}{2} \frac{(\ln \tau_{t,i} - \mu)^2}{\sigma^2} - \ln \left[\Phi \left(\frac{\ln T_t - \mu}{\sigma} \right) \right] \right\} \quad (2.23)$$

Para obter os estimadores de máxima verossimilhança dos parâmetros μ e σ^2 , deriva-se a função de log-verossimilhança em relação a cada um dos parâmetros e iguala-se a derivada a zero.

A derivada em relação a μ é:

$$\frac{\partial l(\tau_{t,i}; \mu, \sigma^2)}{\partial \mu} = \sum_{t=1}^n \sum_{i=1}^{k_t} \frac{(\ln \tau_{t,i} - \mu)}{\sigma^2} - \sum_{t=1}^n \sum_{i=1}^{k_t} \frac{\partial}{\partial \mu} \left\{ \ln \left[\Phi \left(\frac{\ln T_t - \mu}{\sigma} \right) \right] \right\} \quad (2.24)$$

Igualando a derivada à zero, obtém-se o estimador para μ , onde não é possível isolar o parâmetro, de modo que sua estimativa tem que ser obtida através de métodos de análise numérica.

$$\mu = \frac{\sum_{t=1}^n \sum_{i=1}^{k_t} \ln \tau_{t,i}}{\sum_{t=1}^n k_t} + \sum_{t=1}^n \frac{\sigma k_t}{\sum_{t=1}^n k_t} \frac{\phi \left(\frac{\ln T_t - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T_t - \mu}{\sigma} \right)} \quad (2.25)$$

Fazendo as mesmas contas para obter o estimador para σ^2 , temos a derivada da função de log-verossimilhança em relação a σ^2 (equação 2.26) e, ao igualar a derivada a zero, obtemos uma expressão (equação 2.27) onde não é possível isolar o parâmetro σ^2 , que também será estimado por métodos numéricos.

$$\frac{\partial l(\tau; \mu, \sigma^2)}{\partial \sigma^2} = \sum_{t=1}^n \sum_{i=1}^{k_t} \left\{ -\frac{1}{2\sigma^2} + \frac{1}{2} \frac{(\ln \tau_{t,i} - \mu)^2}{\sigma^4} - \frac{\partial}{\partial \sigma^2} \left\{ \ln \left[\Phi \left(\frac{\ln T_t - \mu}{\sigma} \right) \right] \right\} \right\} \quad (2.26)$$

$$\sigma^2 = \frac{\sum_{t=1}^n \sum_{i=1}^{k_t} (\ln \tau_{t,i} - \mu)^2}{\sum_{t=1}^n k_t - \sum_{t=1}^n \left(\frac{\ln T_t - \mu}{\sigma} \right) \frac{\phi \left(\frac{\ln T_t - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T_t - \mu}{\sigma} \right)}} \quad (2.27)$$

Os cálculos de desenvolvimento de ambas as derivadas podem ser encontrados no Apêndice A1.

2.2.2

Modelo conjunto para a quantidade e os atrasos

O modelo proposto neste trabalho segue a estrutura utilizada por Souza (2013). Este modelo se inspira em três abordagens para estimar a reserva IBNR: Weissner (1978), Antônio e Plat (2012) e Parodi (2013).

Antônio e Plat (2012) desenvolvem a estrutura probabilística de Norberg (1993, 1999), especificando modelos estocásticos para a frequência que os sinistros ocorrem, o atraso no aviso, o processo de desenvolvimento e os pagamentos feitos durante o processo. Essas medidas permitem a quantificação da reserva IBNR.

Parodi (2013) utiliza dados fora do triângulo *run-off* e inspira-se no estado da arte da metodologia usada em apreçamento, onde 2 modelos independentes, um de frequência de sinistros e outro de severidade são desenvolvidos. Em sua aplicação, ele obtém a reserva estimada, assim como uma medida da variação desta previsão e uma distribuição para medida de reserva que considera mais realista do que a obtida pelo *Chain ladder*. Neste método, o processo é composto pela combinação de três componentes: o processo de frequência em que os sinistros ocorrem, a distribuição de atrasos no aviso de sinistros e uma distribuição da severidade dos sinistros. Esta estrutura exposta por Parodi inspira o modelo que será utilizado nesta dissertação, sendo o foco nos 2 primeiros componentes para obter a IBNR pura. Nele é estabelecida uma estrutura para a frequência que os sinistros ocorrem e atraso do aviso dos sinistros, mas não modela o processo de pagamento no nível de sinistros individuais como Haastrup e Arjas(1997) e Norberg (1993, 1999), e sim o montante de pagamento esperado.

Utiliza-se um processo Poisson como distribuição do total de sinistros por dia de ocorrência, baseando-se em Guiahi (1986), Weissner (1978) e Kaminsky (1987).

A modelagem de atrasos é baseada na abordagem de Weissner (1978) que foi apresentada na sessão anterior e que aqui será adaptada para a distribuição log-normal. Esta foi a primeira inspiração para a elaboração deste modelo conjunto, pois usa uma distribuição truncada para lidar com a limitação dos dados que vão até a data atual e faz a estimação por função máxima verossimilhança de distribuições truncadas. No entanto seu modelo não leva em consideração que a quantidade de sinistros já comunicados também é uma variável aleatória. A abordagem de Antonio e Plat não trata o fato dos dados serem truncados na data presente, e sim tratam a distribuição real como uma amostra viesada para atrasos curtos. Desta forma será alta a probabilidade relacionada a atrasos curtos nas distribuições de atrasos estimadas. A abordagem de Parodi já evita este problema, estimando a distribuição empírica dos dados, mas ajustando a tendência para atrasos mais curtos usando uma relação entre a distribuição completa e a truncada dos atrasos. Mas esta correção pode ser complicada dependendo da distribuição a ser adotada.

A solução adotada neste modelo conjunto foi: modelar a quantidade de sinistros ocorridos em t e comunicados até T_t por uma distribuição binomial; o tempo de atraso no aviso dos sinistros ocorridos em t por uma distribuição log-normal truncada e a quantidade total de sinistros ocorridos no período de tempo t por uma distribuição Poisson, obtendo assim diretamente a distribuição da quantidade de sinistros IBNR. Justifica-se a escolha das distribuições binomial e Poisson pela definição e conceitos das próprias distribuições, além de elas possuírem uma relação que auxilia os cálculos. E como distribuição de atrasos foi escolhida a log-normal, pois esta distribuição muitas vezes é associada ao tempo de vida de produtos e materiais, tempo de falha ou sobrevivência, tempo necessário para executar uma tarefa, distribuições que possuem assimetria positiva, pois apresenta uma maior frequência em valores menores.

2.2.2.1

Apresentação das variáveis do modelo

Para começar, serão definidos e explicitados as variáveis e parâmetros que serão utilizadas no modelo.

Seja d a variável que representa o período de ocorrência do sinistro, $d=1, 2, \dots, D$, onde D é o último período de tempo observado na amostra.

Considera-se uma janela de períodos de ocorrência de tamanho J e os períodos contidos nesta janela identificados pelo índice $t=1, \dots, J$, onde $J \leq D$, ou seja, o último período de ocorrência dessa janela será sempre observado na amostra. Desta forma, a janela pode conter intervalos de períodos de ocorrência de $d= 1, \dots, J$ até $d= D-J+1, \dots, D$. Esta janela é deslizante e vai passar por todos os dias de ocorrência observáveis a fim de obter novas estimativas de parâmetros do modelo proposto à medida que deslizamos a janela. Sejam ainda:

- T_t : atraso máximo observável no aviso dos sinistros ocorridos no período t ;
- N_t : número total de sinistros ocorridos no período t ;
- K_t : variável aleatória do número de sinistros ocorridos em t , avisados até T_t ;
- $\Gamma_t = (\Gamma_{t,1}, \dots, \Gamma_{t,N_t})$: vetor aleatório de todos os tempos de atraso dos sinistros ocorridos no período t ;
- $\Gamma_{I,t} = (\Gamma_{t,1}, \dots, \Gamma_{t,K_t})$: vetor aleatório de todos os tempos de atraso dos sinistros avisados relativos ao período t ;
- $\Gamma_{II,t} = (\Gamma_{t,K_t+1}, \dots, \Gamma_{t,N_t})$: vetor aleatório de todos os tempos de atraso dos sinistros a serem avisados relativos ao período t ;
- $\Gamma = (\Gamma_1, \dots, \Gamma_J)$: vetor aleatório de todos os tempos de atraso;
- $T = (T_1, \dots, T_J)$: vetor de todos os atrasos máximos observáveis;
- $N = (N_1, \dots, N_J)$: vetor aleatório dos totais de sinistros ocorridos por cada dia;
- $K = (K_1, \dots, K_J)$: vetor das quantidades de sinistros ocorridas em cada dia.
- n_t : número não observável de sinistros com origem na data t ;

- k_t : número observado de sinistros comunicados até T com origem na data t ;
- $\tau_t = (\tau_{t,1}, \dots, \tau_{t,N_t})$: vetor de todos os tempos de atraso dos sinistros ocorridos no dia t ;
- $\tau_{I,t} = (\tau_{t,1}, \dots, \tau_{t,K_t})$: vetor aleatório de todos os tempos de atraso dos sinistros avisados relativos no dia t ;
- $\tau_{II,t} = (\tau_{t,K_t+1}, \dots, \tau_{t,N_t})$: vetor aleatório de todos os tempos de atraso dos sinistros a serem avisados relativos no dia t ;
- λ : vetor de parâmetros da distribuição de atrasos;
- γ : vetor de parâmetros da distribuição do número de sinistros ocorridos em t , N_t ;

A distribuição conjunta das variáveis aleatórias de atraso no aviso do sinistro (Γ_t), número de sinistros avisados até o último dia observado (K_t) e total de sinistros ocorridos no dia t (N_t) é dada por:

$$\begin{aligned} f_{\Gamma_t, K_t, N_t}(\tau_t, k_t, n_t; \mu, \sigma^2, \gamma, T_t) &= f_{\Gamma_t / K_t = k_t, N_t = n_t}(\tau_t; \mu, \sigma^2, T_t) \\ &\quad \times f_{K_t / N_t = n_t}(k_t; \mu, \sigma^2, T_t) \\ &\quad \times f_{N_t}(n_t; \gamma) \end{aligned} \quad (2.28)$$

onde:

$f_{\Gamma_t / K_t = k_t, N_t = n_t}(\tau_t; \mu, \sigma^2, T_t)$ é a distribuição de atrasos dado a quantidade de sinistros avisados até T_t , k_t e o total de sinistros ocorridos em t , n_t ;

$f_{K_t / N_t = n_t}(k_t; \mu, \sigma^2, T_t)$ é a distribuição de probabilidades de se observar K_t sinistros observados até T_t ;

$f_{N_t}(n_t; \gamma)$ é a distribuição do total de sinistros ocorridos em t .

Assume-se que a variável aleatória N_t , o número total de sinistros ocorridos no período t , segue uma distribuição Poisson com parâmetro γ e sua função de probabilidade é dada por:

$$f_{N_t}(n_t; \gamma) = \frac{\gamma^{n_t}}{n_t!} e^{-\gamma} \quad (2.29)$$

Onde $n_t = 0, 1, \dots$

Já a variável K_t condicionada a $N_t = n_t$ segue uma distribuição binomial com parâmetros n_t e $p(\mu, \sigma^2, T_t)$ que depende dos parâmetros μ e σ^2 e cuja função de densidade igual a:

$$f_{K_t/N_t=n_t}(k_t, n_t, p(\mu, \sigma^2, T_t)) = \frac{n_t!}{k_t!(n_t - k_t)!} p(\mu, \sigma^2, T_t)^{k_t} (1 - p(\mu, \sigma^2, T_t))^{n_t - k_t} \quad (2.30)$$

onde $k_t = 0, 1, \dots, n_t$ e $p(\mu, \sigma^2, T_t)$ é a probabilidade de sinistros serem avisados com atrasos menores ou iguais a T_t .

A variável aleatória Γ_t condicionada à $K_t = k_t, N_t = n_t$, que se refere à distribuição de atrasos, neste trabalho segue a distribuição log-normal, com função de densidade:

$$f_{\Gamma_t/K_t=k_t, N_t=n_t}(\tau_t; \mu, \sigma^2, T_t) = \frac{1}{\sqrt{2\pi\sigma\tau_t}} e^{-\frac{1(\ln\tau_t - \mu)^2}{2\sigma^2}}; \tau_t > 0, \mu > 0, \sigma^2 > 0 \quad (2.31)$$

2.2.2.2

Estimação dos parâmetros do modelo integrado quantidade/atraso por máxima verossimilhança

Se fosse possível observar todos os atrasos no aviso para os sinistros ocorridos em J dias, a função de verossimilhança generalizada para todos os dados observados em J dias é dada por (dado Γ_t, K_t, N_t):

$$\begin{aligned} L(\mu, \sigma^2, \gamma/\Gamma, K, N, T) &= \prod_{t=1}^J f_{\Gamma_t, K_t, N_t}(\tau_t, k_t, n_t; \mu, \sigma^2, \gamma, T_t) \\ &= \prod_{t=1}^J f_{\Gamma_t/K_t=k_t, N_t=n_t}(\tau_t; \mu, \sigma^2, T_t) f_{K_t/N_t=n_t}(k_t; \mu, \sigma^2, T_t) f_{N_t}(n_t; \gamma) \end{aligned} \quad (2.32)$$

Separando o primeiro termo em componentes observáveis, atrasos menores que T_t e não observáveis, atrasos maiores que T_t :

$$\begin{aligned}
L(\mu, \sigma^2, \gamma/\Gamma, K, N, T) &= \\
&= \prod_{t=1}^J f_{\Gamma_{I,t}/K_t=k_t, N_t=n_t}(\tau_{I,t}; \mu, \sigma^2, T_t) f_{\Gamma_{II,t}/K_t=k_t, N_t=n_t}(\tau_{II,t}; \mu, \sigma^2, T_t) \\
&\quad \times f_{K_t/N_t=n_t}(k_t; \mu, \sigma^2, T_t) f_{N_t}(n_t; \gamma)
\end{aligned} \tag{2.33}$$

A função de log-verossimilhança para os dados completos é dada por:

$$\begin{aligned}
l(\mu, \sigma^2, \gamma/\Gamma, K, N, T) &= \\
&= \sum_{t=1}^J \ln f_{\Gamma_t/K_t=k_t, N_t=n_t}(\tau_t; \mu, \sigma^2, T_t) \\
&\quad + \sum_{t=1}^J \ln f_{K_t/N_t=n_t}(k_t; \mu, \sigma^2, T_t) \\
&\quad + \sum_{t=1}^J \ln f_{N_t}(n_t; \gamma)
\end{aligned} \tag{2.34}$$

Utilizando como distribuição de atrasos no aviso, a distribuição log-normal:

$$f_{\tau_t}(\tau_{t,i}; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma\tau_{t,i}}} e^{-\frac{1(\ln\tau_{t,i}-\mu)^2}{2\sigma^2}} \tag{2.35}$$

E a distribuição condicionada (truncada) aos dados observáveis, ou seja, a atrasos menores ou iguais a T_t é dada por:

$$f_{\tau_t/\tau \leq T_t}(\tau_{t,i}; \mu, \sigma^2, T_t) = \frac{1}{\sqrt{2\pi\sigma\tau_{t,i}}} e^{-\frac{1(\ln\tau_{t,i}-\mu)^2}{2\sigma^2}} / \Phi\left(\frac{\ln T_t - \mu}{\sigma}\right) \tag{2.36}$$

Onde $\Phi(\cdot)$ é a função acumulada da distribuição normal padrão $N(0,1)$.

Seguindo os cálculos realizados por Souza (2013) na seção 3.4.3 de sua dissertação, a função de verossimilhança para os dados truncados e incompletos para $t=1,2,.. J$ dias do modelo proposto é:

$$\begin{aligned}
L(\tau_{t,i}; \mu, \sigma^2, \gamma) &= \prod_{t=1}^J \left\{ \prod_{i=1}^{k_t} \frac{f_{\tau}(\tau_{t,i}; \mu, \sigma^2)}{(1-F_{\tau}(T_t; \mu, \sigma^2))} \right\} \sum_{n=k_t}^{\infty} \frac{n!}{k_t!(n-k_t)!} (1-F_{\tau}(T_t; \mu, \sigma^2))^{k_t} (F_{\tau}(T_t; \mu, \sigma^2))^{n-k_t} \frac{\gamma^n}{n!} e^{-\gamma} \\
L(\tau_{t,i}; \mu, \sigma^2, \gamma) &= \prod_{t=1}^J \left\{ \prod_{i=1}^{k_t} f_{\tau}(\tau_{t,i}; \mu, \sigma^2) \right\} \frac{e^{-\gamma}}{k_t! \gamma^{k_t}} \sum_{n=k_t}^{\infty} \frac{(F_{\tau}(T_t; \mu, \sigma^2) \gamma)^{n-k_t}}{(n-k_t)!} \\
L(\tau_{t,i}; \mu, \sigma^2, \gamma) &= \prod_{t=1}^J \left\{ \prod_{i=1}^{k_t} f_{\tau}(\tau_{t,i}; \mu, \sigma^2) \right\} \frac{e^{-\gamma}}{k_t! \gamma^{k_t}} e^{F_{\tau}(T_t; \mu, \sigma^2) \gamma}
\end{aligned} \tag{2.37}$$

Onde $F_{\tau}(T_t; \mu, \sigma^2)$ é probabilidade do atraso ser maior que o atraso máximo observável T_t .

A função de log-verossimilhança a ser maximizada é:

$$l(\tau_{t,i}; \mu, \sigma^2, \gamma) = \sum_{t=1}^J \sum_{i=1}^{k_t} \ln f_{\tau}(\tau_{t,i}; \mu, \sigma^2) + \sum_{t=1}^J (-\ln k_t! + k_t \ln \gamma + F_{\tau}(T_t; \mu, \sigma^2) \gamma - \gamma) \tag{2.38}$$

Para obter o estimador de γ , primeiro temos que derivar a função de verossimilhança em relação à γ :

$$\frac{\partial l(\tau_{t,i}; \mu, \sigma^2, \gamma)}{\partial \gamma} = \sum_{t=1}^J \left(\frac{k_t}{\gamma} + F_{\tau}(T_t; \mu, \sigma^2) - 1 \right) \tag{2.39}$$

Igualando a derivada acima à zero, obtém-se o estimador de γ :

$$\hat{\gamma} = \frac{k.}{J - \sum_{t=1}^J F_{\tau}(T_t; \mu, \sigma^2)} \tag{2.40}$$

Onde $k. = \sum_{t=1}^J k_t$.

Atualizando a função de log-verossimilhança com a distribuição de atrasos log-normal, onde:

$$f_{\tau}(\tau_{t,i}; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma\tau_{t,i}}} e^{-\frac{1(\ln\tau_{t,i}-\mu)^2}{2\sigma^2}}$$

e

$$F_{\tau}(T_t; \mu, \sigma^2) = \left(1 - \Phi \left(\frac{\ln T_t - \mu}{\sigma} \right) \right)$$

Onde $\Phi(\cdot)$ é a função acumulada da distribuição normal padrão $N(0,1)$.

A log-verossimilhança da distribuição conjunta passa a ser:

$$\begin{aligned} l(\tau_{t,i}; \mu, \sigma^2, \gamma) &= \sum_{t=1}^J \sum_{i=1}^{k_t} \left\{ -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma^2) - \ln(\tau_{t,i}) - \frac{1}{2} \frac{(\ln \tau_{t,i} - \mu)^2}{\sigma^2} \right\} \\ &+ \sum_{t=1}^J \left(-\ln k_t! + k_t \ln \gamma + \left(1 - \Phi \left(\frac{\ln T_t - \mu}{\sigma} \right) \right) \gamma - \gamma \right) \end{aligned} \quad (2.41)$$

A derivada da função log-verossimilhança em relação a μ é:

$$\frac{\partial l(\tau_{t,i}; \mu, \sigma^2, \gamma)}{\partial \mu} = \sum_{t=1}^J \sum_{i=1}^{k_t} \frac{(\ln \tau_{t,i} - \mu)}{\sigma^2} - \sum_{t=1}^J \gamma \underbrace{\frac{\partial}{\partial \mu} \Phi \left(\frac{\ln T_t - \mu}{\sigma} \right)}_I \quad (2.42)$$

O termo I já foi calculado na conta da distribuição truncada da lognormal pelo enfoque de Weissner (Apêndice A1) e é dado por $I = -\frac{1}{\sigma} \phi \left(\frac{\ln T_t - \mu}{\sigma} \right)$, logo:

$$\begin{aligned} \frac{\partial l(\tau_{t,i}; \mu, \sigma^2, \gamma)}{\partial \mu} &= \sum_{t=1}^J \sum_{i=1}^{k_t} \frac{(\ln \tau_{t,i} - \mu)}{\sigma^2} + \sum_{t=1}^J \frac{\gamma}{\sigma} \phi \left(\frac{\ln T_t - \mu}{\sigma} \right) \\ \frac{\partial l(\tau_{t,i}; \mu, \sigma^2, \gamma)}{\partial \mu} &= \frac{k \cdot (\overline{\ln \tau} - \mu)}{\sigma^2} + \sum_{t=1}^J \frac{\gamma}{\sigma} \phi \left(\frac{\ln T_t - \mu}{\sigma} \right) \end{aligned} \quad (2.43)$$

onde $k = \sum_{t=1}^J k_t$ e $\overline{\ln \tau} = \frac{1}{k} \sum_{t=1}^J \sum_{i=1}^{k_t} \ln \tau_{t,i}$.

Igualando a derivada a zero, obtém-se o estimador de μ :

$$\mu = \overline{\ln \tau} + \frac{\sigma}{k} \gamma \sum_{t=1}^J \phi \left(\frac{\ln T_t - \mu}{\sigma} \right) \quad (2.44)$$

A derivada da log-verossimilhança em relação a σ^2 é:

$$\frac{\partial l(\tau_{t,i}; \mu, \sigma^2, \gamma)}{\partial \sigma^2} = \sum_{t=1}^J \sum_{i=1}^{k_t} \left\{ -\frac{1}{2\sigma^2} + \frac{1}{2} \frac{(\ln \tau_{t,i} - \mu)^2}{\sigma^4} \right\} - \sum_{t=1}^J \gamma \underbrace{\frac{\partial}{\partial \sigma^2} \Phi \left(\frac{\ln T_t - \mu}{\sigma} \right)}_{\text{II}} \quad (2.45)$$

O termo II já foi calculado na conta da distribuição truncada da lognormal por Weissner (Apêndice A1) e é dado por $\text{II} = \frac{1}{2\sigma^2} \left\{ -\left(\frac{\ln T_t - \mu}{\sigma} \right) \phi \left(\frac{\ln T_t - \mu}{\sigma} \right) \right\}$, então:

$$\begin{aligned} \frac{\partial l(\tau_{t,i}; \mu, \sigma^2, \gamma)}{\partial \sigma^2} &= \sum_{t=1}^J \sum_{i=1}^{k_t} \left\{ -\frac{1}{2\sigma^2} + \frac{1}{2} \frac{(\ln \tau_{t,i} - \mu)^2}{\sigma^4} \right\} + \sum_{t=1}^J \frac{\gamma}{2\sigma^2} \left(\frac{\ln T_t - \mu}{\sigma} \right) \phi \left(\frac{\ln T_t - \mu}{\sigma} \right) \\ \frac{\partial l(\tau_{t,i}; \mu, \sigma^2, \gamma)}{\partial \sigma^2} &= \sum_{t=1}^J \sum_{i=1}^{k_t} \left\{ -\frac{1}{2\sigma^2} \left(1 - \frac{(\ln \tau_{t,i} - \mu)^2}{\sigma^2} - \frac{\gamma}{k_t} \left(\frac{\ln T_t - \mu}{\sigma} \right) \phi \left(\frac{\ln T_t - \mu}{\sigma} \right) \right) \right\} \end{aligned} \quad (2.46)$$

Iguala-se a derivada acima a zero, para obter o estimador de σ^2 :

$$\begin{aligned} \sum_{t=1}^J \sum_{i=1}^{k_t} \left\{ 1 - \frac{(\ln \tau_{t,i} - \mu)^2}{\sigma^2} - \frac{\gamma}{k_t} \left(\frac{\ln T_t - \mu}{\sigma} \right) \phi \left(\frac{\ln T_t - \mu}{\sigma} \right) \right\} &= 0 \\ \sigma^2 &= \frac{\sum_{t=1}^J \sum_{i=1}^{k_t} (\ln \tau_{t,i} - \mu)^2}{\sum_{t=1}^J \sum_{i=1}^{k_t} \left\{ 1 - \frac{\gamma}{k_t} \left(\frac{\ln T_t - \mu}{\sigma} \right) \phi \left(\frac{\ln T_t - \mu}{\sigma} \right) \right\}} \end{aligned} \quad (2.47)$$

Nota-se que tanto para μ como para σ^2 não é possível isolar o parâmetro após igualar a derivada a zero (equações 2.44 e 2.47) para obter uma expressão analítica para os estimadores. Então eles serão estimados através do algoritmo de busca não linear `fmincon` do software Matlab, usando as derivadas acima para compor o gradiente a ser utilizado pelo algoritmo.

O parâmetro γ será estimado de forma iterativa seguindo os passos abaixo:

- 1- Escolhe-se um valor inicial para γ ;
- 2- A partir deste valor de γ fixo, as estimativas dos parâmetros da distribuição de atrasos são encontradas pelo algoritmo `fmincon`;
- 3- Calcula-se um novo valor para γ utilizando os parâmetros estimados no passo anterior;

4- Calcula-se a diferença entre este novo γ e o γ anterior, continuam-se as iterações para estimativa deste parâmetro (retornando ao passo 2), caso esta diferença seja maior que um valor determinado ε .

Para os últimos dias de ocorrência de sinistros temos poucos dados observados para ajuste das curvas de atraso. Além disso, os atrasos observados são muito pequenos, já que estão próximos à última data observada. Esses fatos atrapalham a estimação pelo método de máxima verossimilhança de uma distribuição que represente os atrasos que serão observados nesses dias. Souza (2013) baseia-se em Al-Athari (2008), que diz que o estimador de máxima verossimilhança do parâmetro da exponencial só existe quando a média amostral for inferior à metade do prazo do truncamento dos dados. Apesar de estarmos trabalhando com a distribuição log-normal, o mesmo critério foi utilizado e quando a média amostral da log-normal passa a ser superior à metade do truncamento, a estimativa dos parâmetros será igual ao último valor estimado.

2.2.2.3

Estimador da quantidade IBNR

Como foi dito anteriormente, a quantidade total de sinistros ocorridos em cada instante de tempo, N_d , segue uma distribuição Poisson com parâmetro γ_d . No entanto a distribuição de N_d não é conhecida, dado as informações observadas no último instante de tempo observado.

Sendo a quantidade N_d estimada pela esperança desta variável, dado as informações conhecidas, e sendo $(N_d - K_d) | \{N_d \geq k_d, K_d = k_d\} \sim \text{Poisson}(\gamma_d F_{\tau_d}(T_d; \lambda_d))$ tem-se:

$$\begin{aligned} \hat{N}_d &= E[N_d | N_d \geq k_d, K_d = k_d] \\ &= E[(N_d - K_d) | N_d \geq k_d, K_d = k_d] + E[K_d | K_d = k_d] \\ &= \hat{\gamma}_d F_{\tau_d}(T_d; \lambda_d) + k_d \end{aligned} \tag{2.48}$$

Onde $\hat{\gamma}_d$ é a estimativa de γ quando a janela de dados é composta pelos dias de ocorrência $t=d - J + 1, \dots, d$, com $J \leq d \leq D$. A demonstração da distribuição $(N_d - K_d) | \{N_d \geq k_d, K_d = k_d\}$ está no Apêndice A2.

A quantidade IBNR do período de ocorrência t será dada por:

$$\begin{aligned}\hat{Q}_d &= \hat{N}_d - k_d \\ &= \hat{\gamma}_d F_{\tau_d}(T_d; \lambda_d) + k_d - k_d \\ &= \hat{\gamma}_d F_{\tau_d}(T_d; \lambda_d)\end{aligned}\tag{2.49}$$

Então a quantidade IBNR total \hat{Q} é estimada por:

$$\hat{Q} = \sum_{d=1}^D \hat{Q}_d\tag{2.50}$$

Sejam τ_1 e τ_2 dois tempos de atraso quaisquer maiores que T , a quantidade de sinistros ocorridos e não avisados com atraso $\tau \in (\tau_1, \tau_2]$ será:

$$\hat{Q}_d = \hat{N}_d \left(F_{\tau_d}(\tau_1; \lambda_d) - F_{\tau_d}(\tau_2; \lambda_d) \right)\tag{2.51}$$

2.2.2.4

Distribuição da quantidade IBNR

Sendo a quantidade IBNR total, Q_d , igual a soma de variáveis independentes com distribuição Poisson, a distribuição da quantidade IBNR total, Q , também segue distribuição Poisson com média igual a soma das médias das variáveis independentes Q_d . Desta maneira, seu intervalo de previsão é diretamente calculado pelos percentis da Poisson com média e variância iguais à quantidade IBNR total estimada.

Da mesma forma, a quantidade IBNR por período de aviso também tem distribuição Poisson com média igual à média das variáveis Poisson que a formam. Esta distribuição será importante para calcular os intervalos de previsão da estimativa para avaliação da qualidade da previsão do modelo dos períodos que serão excluídos da amostra para teste. A distribuição da quantidade por período de aviso encontra-se no Apêndice A2.

3

Análise dos dados

Nesta seção será realizada a descrição e análise dos dados a serem utilizados para aplicar as metodologias citadas na seção anterior.

3.1

Descrição dos dados

Os dados utilizados neste trabalho são os dados de quantidade de sinistros do seguro DPVAT. O Seguro de Danos Pessoais Causados por Veículos Automotores de Via Terrestre, mais conhecido como Seguro DPVAT, existe desde 1974. É um seguro de caráter social que cobre danos pessoais (e não materiais) de vítimas de acidentes de trânsito ou seus beneficiários. Qualquer pessoa que for vítima de acidente causado por veículos automotores ou sua carga, independente da apuração de culpa, seja motorista, passageiro ou pedestre, tem direito ao seguro. As coberturas do seguro são três:

- Morte
- Invalidez permanente total ou parcial
- Reembolso de despesas de assistência médica e suplementares (DAMS)

Os recursos do Seguro DPVAT são financiados pelos proprietários de veículos, por meio de pagamento anual. As indenizações são pagas individualmente, não importando quantas vítimas estiverem envolvidas no mesmo acidente, assim fica difícil especificar a população exposta ao benefício. Além disto, grandes modificações são feitas na regulação deste seguro, entre elas, alteração do prazo de aviso e nos valores máximos de indenização de cada cobertura. O desenvolvimento de pagamento do seguro DPVAT está diretamente associado às alterações de limites máximos de indenização. Para que as indenizações continuem significativas em relação aos danos e causas cobertas, assim como fiquem de acordo com as oscilações de valorização da moeda, é que essas mudanças são realizadas. A mudança no prazo de aviso também influencia o comportamento dos reclamantes, afetando a velocidade de aviso do sinistro e o

valor médio da indenização. Então optou-se por modelar o desenvolvimento dos sinistros e calcular a quantidade IBNR e posteriormente associar a severidade a eles. Neste trabalho será calculada a reserva para os dados da cobertura morte e serão utilizados os dados de quantidade de aviso para o cálculo pelos métodos expostos.

3.2

Análise descritiva dos dados

Os dados utilizados possuem informações do dia da ocorrência do sinistro, o dia do aviso e quantidade de sinistros por dia e período de atraso. Esta informação tem início em 2001 e se desenvolve até dezembro de 2011, totalizando 4009 dias. Na Tabela abaixo está uma análise descritiva dos dados que estão agrupados em anos.

Tabela 1: Estatísticas descritivas

Ano de Ocorrência	Atraso Médio	Máximo Atraso	Desvio-padrão do Atraso Médio	Quantidade média por dia	Desvio-padrão da quantidade total
2001	176.5	4016	50.3	90.1	34.2
2002	176.3	3651	49.1	95.8	39.2
2003	194.1	3286	56.2	91.0	37.9
2004	224.0	2921	48.8	93.4	38.5
2005	189.9	2555	47.7	97.1	39.5
2006	172.6	2190	43.5	96.6	39.2
2007	166.1	1825	35.2	103.8	44.1
2008	133.2	1460	27.9	106.0	41.5
2009	115.6	1094	23.0	104.1	41.5
2010	96.4	729	16.5	113.0	45.5
2011	59.3	364	25.5	89.4	48.7
Total geral	154.0	3881	71.0	98.0	41.7

A Tabela 1 apresenta o número médio de dias de atraso, o atraso máximo observável em dias, o desvio padrão do atraso médio em dias, a quantidade média de sinistros ocorridos por dia da amostra e o desvio padrão da quantidade total de sinistros ocorridos por dia, todas essas informações separadas por ano e no total da amostra. Os anos que apresentaram maior atraso médio foram 2003, 2004 e

2005; em 2003 entrou em vigor o novo código civil (janeiro de 2003)² que reduziu o prazo para dar entrada no pedido de indenização de 20 anos para 3 anos. Percebe-se que, mesmo com esta alteração no prazo para dar entrada no pedido, o atraso médio ainda foi maior do que os outros anos, pois a alteração só é válida para os casos em que o tempo entre a ocorrência do sinistro e a solicitação seja inferior a 10 anos. E o atraso médio fica em menos de 60 dias em 2011, o que se deve principalmente a falta de observação de dados a partir de 2011.

O atraso máximo observado na amostra foi de 3881 dias, que é um pouco superior a 10 anos de atraso no aviso do sinistro. Há uma grande distância entre os atrasos máximos e os atrasos médios observados. Dado o truncamento dos dados, com o passar dos anos o atraso máximo vai se reduzindo, como era de se esperar, chegando à 364 dias (quase 1 ano) em 2011. O desvio padrão do atraso médio por dia apresenta maior valor em 2003 sendo de 56,2 dias, isto significa que a variação em relação a média de atrasos médios de todos os dias deste ano é de 56,2 dias. A quantidade média anual de sinistros informados por dia vai crescendo ao longo dos anos, mesmo com a ausência de dados observados após dezembro de 2011, o que sugere que este número deve ser ainda maior principalmente a partir de 2003 quando ocorreu a mudança no prazo máximo de aviso do sinistro. O desvio padrão da quantidade de atrasos avisados por dia é em torno de 41,7 sinistros no total da amostra.

3.3

Análise gráfica da distribuição de atrasos e frequência de sinistros

Buscando conhecer melhor a distribuição de atrasos, a fim de saber qual distribuição se assemelha mais aos dados, foi feita uma análise gráfica dos dados de atraso. Como não seria possível apresentar os gráficos de todos os dias observados, foram selecionados 9 dias da amostra de acordo com o intervalo de dados observados, de forma que estes dias estão bem distribuídos em toda a

² Para os acidentes anteriores a janeiro de 2003 os tribunais têm aplicado uma regra de transição: se entre a data em que está definido o dano pessoal que dá direito à cobertura do seguro (morte, invalidez ou reembolso de despesas) e a data de entrada em vigor do novo Código Civil (11/01/2003) houve o transcurso de um período superior a 10 (dez) anos, será aplicável o prazo de 20 (vinte) anos para o requerimento do DPVAT. Entretanto, se esse período é inferior a 10 (dez) anos, observar-se-á o novo prazo de 3 (três) anos, iniciando sua contagem a partir de 11/01/2003.

Fonte: Site da DPVAT.

amostra de 11 anos. Como os dados são truncados na última data observada e quanto mais antigo o dia da ocorrência, menor é o truncamento dos dados, o primeiro dia selecionado é o primeiro dia observado (01/01/2001), o que possibilita a observação de uma amostra mais completa e foram selecionados os demais dias seguindo um intervalo entre eles até o dia 3569 (09/10/2010) que é o dia de ocorrência mais recente e com a distribuição mais incompleta dentre os dias selecionados. A Tabela 2 a seguir apresenta os dias que foram selecionados e a identificação dos mesmos nos gráficos.

Tabela 2: Dias selecionados para os gráficos

Identificação do dia (y)	Data da ocorrência
1	01/01/2001
447	23/03/2002
893	12/06/2003
1339	31/08/2004
1785	20/11/2005
2231	09/02/2007
2677	30/04/2008
3123	20/07/2009
3569	09/10/2010

Na Tabela 3 encontram-se a quantidade observada e o atraso médio para os dias selecionados. Como foi possível observar no atraso médio por ano, o atraso médio vai diminuindo ao longo dos dias selecionados, como se esperava devido ao truncamento dos dados no ano de 2011. Já a quantidade observada em cada dia varia bastante, sendo o dia 09/10/2010 o dia com maior quantidade de sinistros observada, 194 ocorrências.

Tabela 3: Quantidade observada e atraso médio dias selecionados

	Quantidade observada	Atraso médio
01/01/2001	144	235.4
23/03/2002	120	217.0
12/06/2003	78	202.0
31/08/2004	67	255.3
20/11/2005	148	255.1
09/02/2007	96	180.8
30/04/2008	115	175.2
20/07/2009	83	123.7
09/10/2010	194	118.9

A seguir estão os histogramas dos atrasos dos avisos dos dias selecionados.

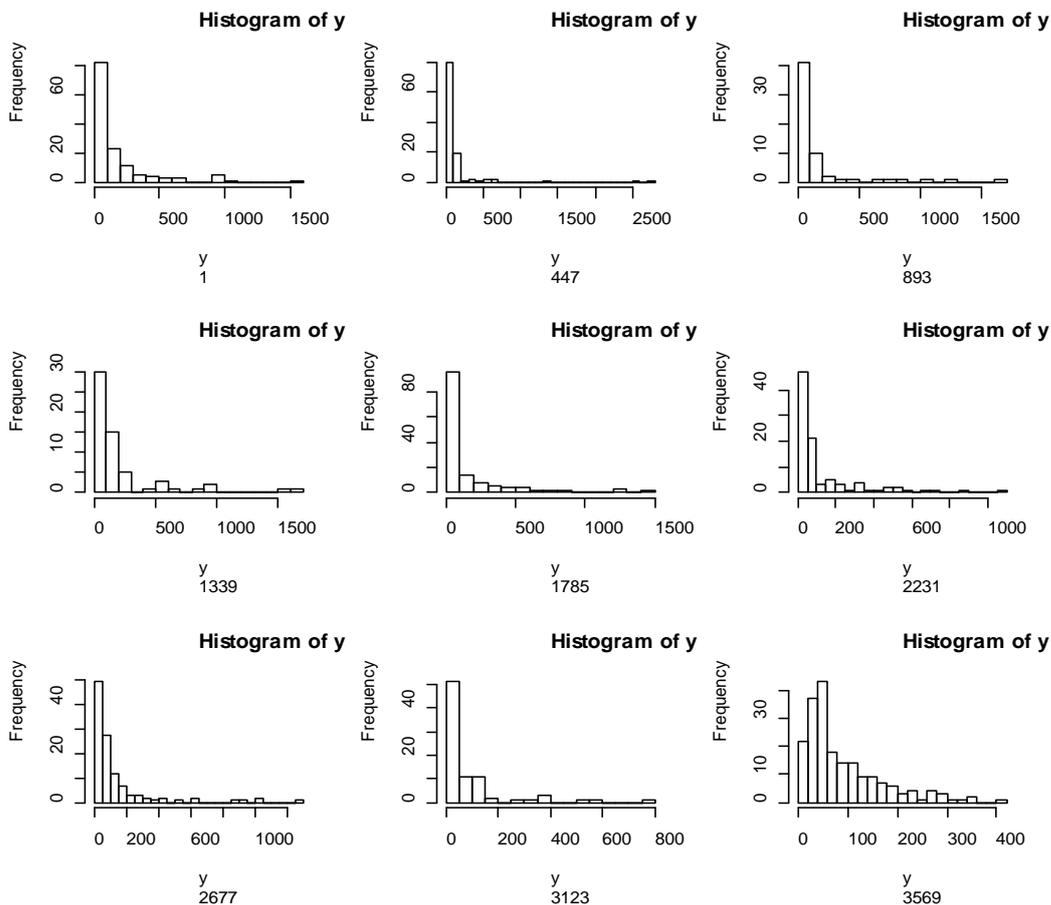


Figura 6: Histogramas dos atrasos no aviso para os 9 dias selecionados

Nos gráficos da Figura 6, pode-se considerar que em alguns dias a distribuição de atrasos possui uma semelhança com a distribuição exponencial,

especialmente os dias 1, 893 e 2677. Já a distribuição dos atrasos do último dia selecionado está mais distante da distribuição exponencial.

Agora será aplicada a função logarítmica aos dados para ver o modo como a distribuição se comporta após esta transformação.

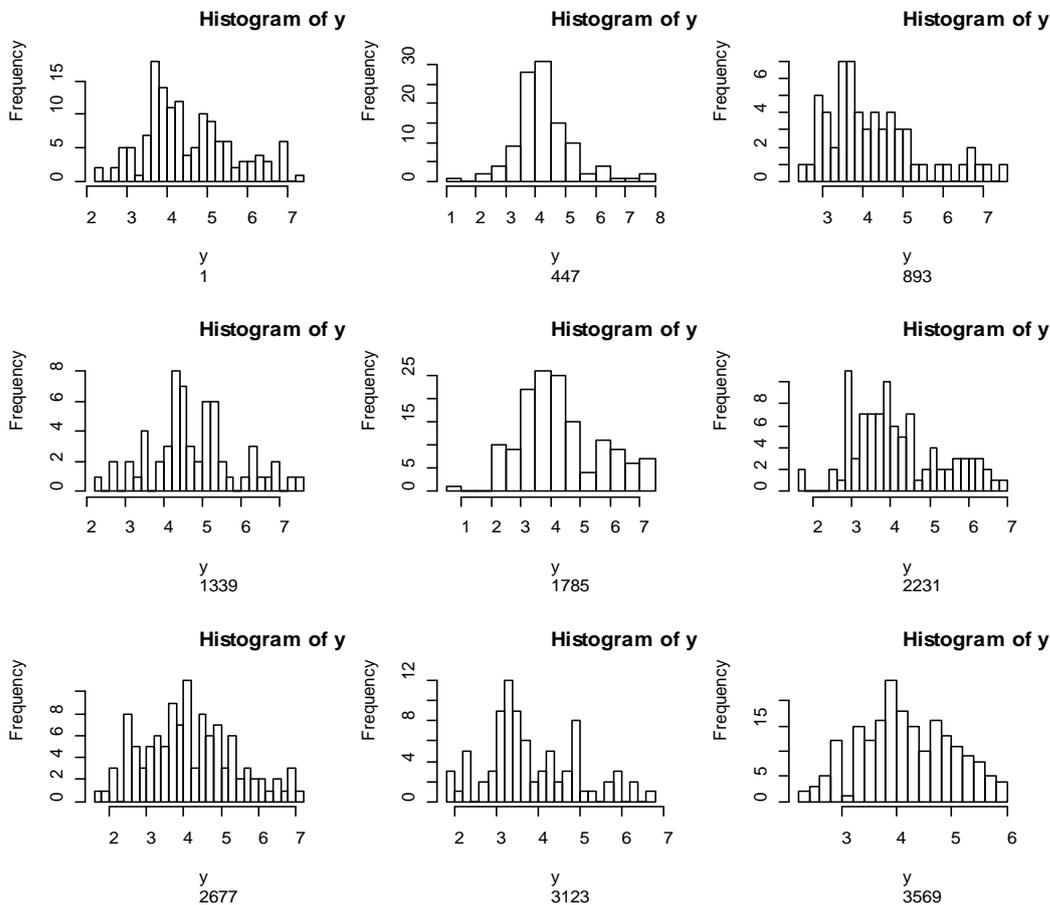


Figura 7: Histogramas do log dos atrasos no aviso para os 9 dias selecionados

Os gráficos acima não deixam claro qual distribuição os dados observados seguem. Com isso foi feito também o gráfico normal qqplot (quantil-quantil normal) que é uma ferramenta útil para checar a adequação da distribuição de frequência dos dados à distribuição normal. Se os dados estiverem próximos a linha que representa a distribuição teórica da distribuição normal, temos indícios de que a distribuição normal se aplica aos dados. Em alguns gráficos os dados estão bem próximos à distribuição normal, em outros devido à assimetria dos dados; aos dados serem discreto ou à presença de outlier (valores atípicos), as informações observadas se afastam da distribuição normal.

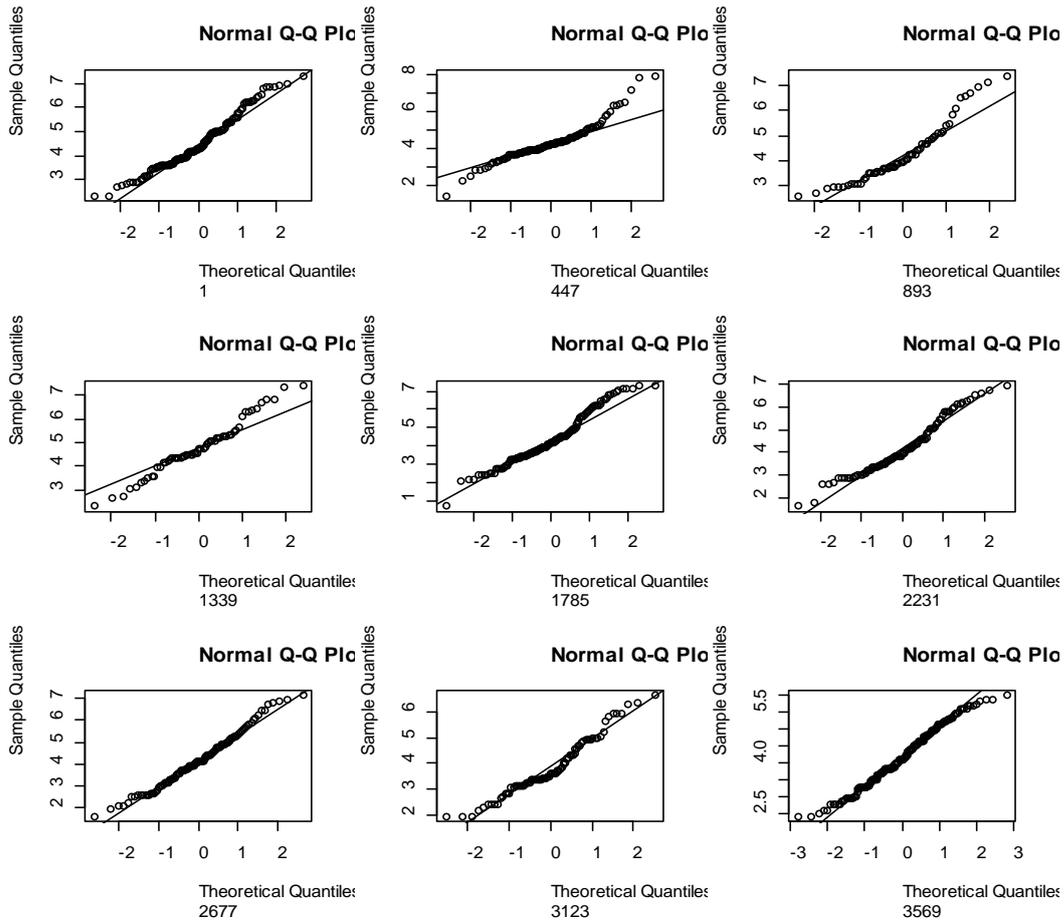


Figura 8: Normal qqplot dos dias selecionados

O fato de alguns dias (após aplicada a função logarítmica) apresentarem um comportamento semelhante a distribuição normal corrobora a escolha da distribuição log-normal como distribuição dos atrasos no aviso do sinistro.

4

Resultados

Neste Capítulo serão apresentados os resultados obtidos pelos métodos: o método *Chain ladder*, o modelo de Weissner para a distribuição de atraso no aviso usando a distribuição lognormal truncada e o modelo usando a distribuição conjunta do atraso no aviso (Lognormal) e quantidade de sinistros (Poisson).

Foram utilizados os dados do DPVAT para a cobertura morte, que possui dados diários com a informação dos sinistros avisados e data de ocorrência entre os anos de 2001 a 2011. Utilizando todos os dados (de janeiro de 2001 a dezembro de 2011) foi realizada a previsão da quantidade IBNR para o ano 2011. Para realizar as comparações com os valores observados e calcular os erros de previsão foram feitos três experimentos: primeiro foram removidos os três últimos anos da amostra (2009, 2010, 2011) e utilizando os dados até dezembro de 2008 foram realizadas as previsões 3 anos à frente, os resultados foram comparados com os valores observados nestes anos e os erros de previsão foram calculados; de forma similar, o segundo foi realizado removendo os dois últimos anos (2010, 2011) e utilizando os dados até dezembro de 2009 para realizar as previsões 2 passos à frente e cálculo de erros; e por último foi removido o último ano observado (2011) e com os dados até dezembro de 2010 foi realizada a previsão 1 ano à frente e o erro de previsão calculado.

4.1

Chain ladder

Primeiramente o método do *Chain ladder* foi aplicado aos dados a fim de obter a estimativa IBNR. Na figura a seguir está o triângulo de desenvolvimento usando os dados de janeiro de 2001 a dezembro de 2011. Como é possível observar no triângulo abaixo, a maior parte dos sinistros foi avisada dentro do mesmo ano.

Origem	Desenvolvimento										
	0	1	2	3	4	5	6	7	8	9	10
2001	21939	8584	1175	494	330	135	85	75	37	23	20
2002	23060	9269	1274	691	252	161	138	81	27	29	
2003	20969	8999	1776	751	306	216	108	60	46		
2004	19113	10993	2120	1118	447	224	97	79			
2005	22943	8868	1891	1082	395	159	103				
2006	23653	8248	1986	939	238	179					
2007	24029	10815	2005	791	261						
2008	26911	9708	1377	784							
2009	27179	9177	1657								
2010	29894	11353									
2011	31647										

Figura 9: Triângulo de desenvolvimento *Chain ladder*

O valor da quantidade IBNR estimada para 2011 encontrada foi 25164 sinistros que ocorreram, mas ainda não foram avisados. Foram também realizadas as previsões para os dados removidos da amostra nos três experimentos supracitados. Na Tabela 4 a seguir são exibidos os erros de previsão da quantidade de sinistros em cada experimento. Vale lembrar que as previsões realizadas pelo método foram comparadas aos valores observados para o cálculo do erro. As legendas ‘Até dez/2008’, ‘Até dez/2009’, ‘Até dez/2010’ se referem aos cortes realizados na base de dados e explicados anteriormente.

Tabela 4 – Medidas dos erros de previsão *Chain ladder*

	Até dez/2008	Até dez/2009	Até dez/2010	Previsão IBNR 2011
MAPE	25%	26%	9%	
MAE	930	1936	1237	
RMSE	1021	2232	1237	25164

Apesar de se tratar de uma previsão 3 passos a frente, as medidas de erro de previsão MAE (Mean Absolute Error, em inglês, ou Erro médio absoluto), MAPE (Mean Absolute Percentage Error, em inglês, ou Porcentagem do erro médio absoluto) e RMSE (Root Mean Square Error, em inglês, ou raiz do erro médio quadrático) para ‘até dez/2008’ apresentaram valores melhores que a previsão 2 passos a frente ‘até dez/2009’.

4.2

Abordagem Weissner

Ambos os modelos foram estimados considerando uma janela deslizante de 365 dias, isto é, os parâmetros da distribuição foram estimados para cada dia de ocorrência t , considerando na janela as ocorrências deste dia e as ocorrências dos 364 dias anteriores a t . Os modelos foram testados nas mesmas condições que o modelo de *Chain ladder*, utilizando os cortes nos dados para cálculo de erro e também utilizando todos os dados para realizar a previsão da quantidade IBNR para 2011.

Primeiramente serão apresentados os resultados obtidos através da abordagem de Weissner usando a distribuição lognormal como distribuição de atraso no aviso. A Figura 10 mostra o atraso médio esperado no aviso da ocorrência do sinistro por dia de ocorrência para cada corte de ano de aviso. As curvas dos diferentes cortes de ano apresentam comportamento muito similar entre si. Observa-se no gráfico que antes do dia de ocorrência 1365 (26/09/2004) começa uma tendência crescente nos atrasos estimados que resulta em um pico no gráfico por volta deste dia. Ao se comparar esta informação com a apresentada na Tabela 1, nota-se que realmente a partir de 2003 o atraso médio tem um aumento considerável em relação aos anos anteriores, sendo 2004 o ano com maior atraso médio observado, 224 dias. O ano de 2005 também apresenta uma média mais elevada e depois o atraso médio começa a decrescer. Ao final das curvas pode-se ver o truncamento dos valores estimados dos parâmetros que foi adotado aqui, por causa do limite de observação dos dados até 2011.

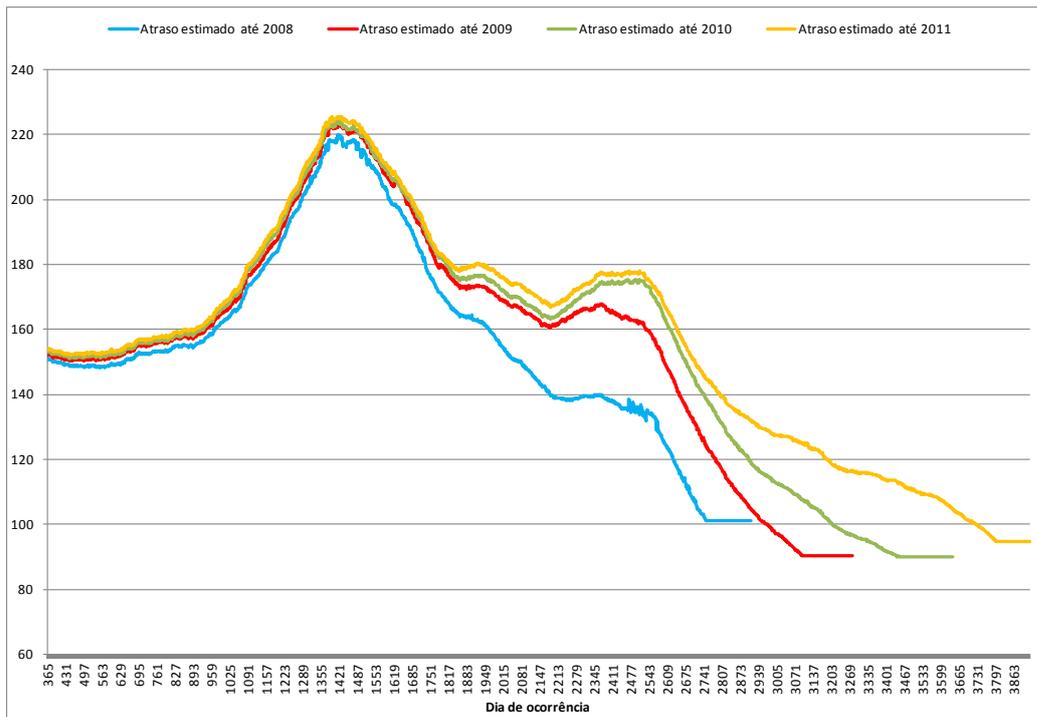


Figura 10: Atraso estimado – Abordagem Weissner

As Figuras 11 e 12 apresentam os gráficos com os valores estimados para cada dia de ocorrência dos parâmetros μ e σ , média e desvio padrão respectivamente, da distribuição de atrasos log-normal. O valor esperado (média) da distribuição log-normal é dado por $E(X) = e^{\mu+1/2\sigma^2}$, portanto depende de ambos os parâmetros. Os valores estimados de μ ficam em torno de 4.08 e 4.67 e, a estimativa de σ fica entre 0.93 e 1.29.

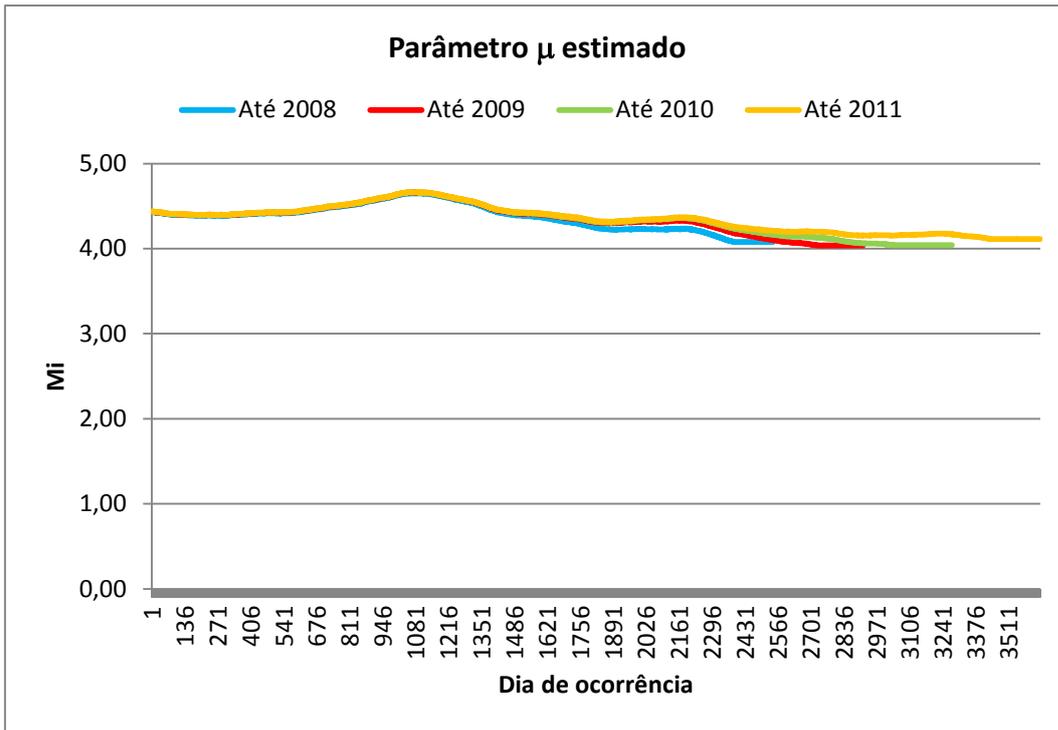


Figura 11: Parâmetro Abordagem Weissner

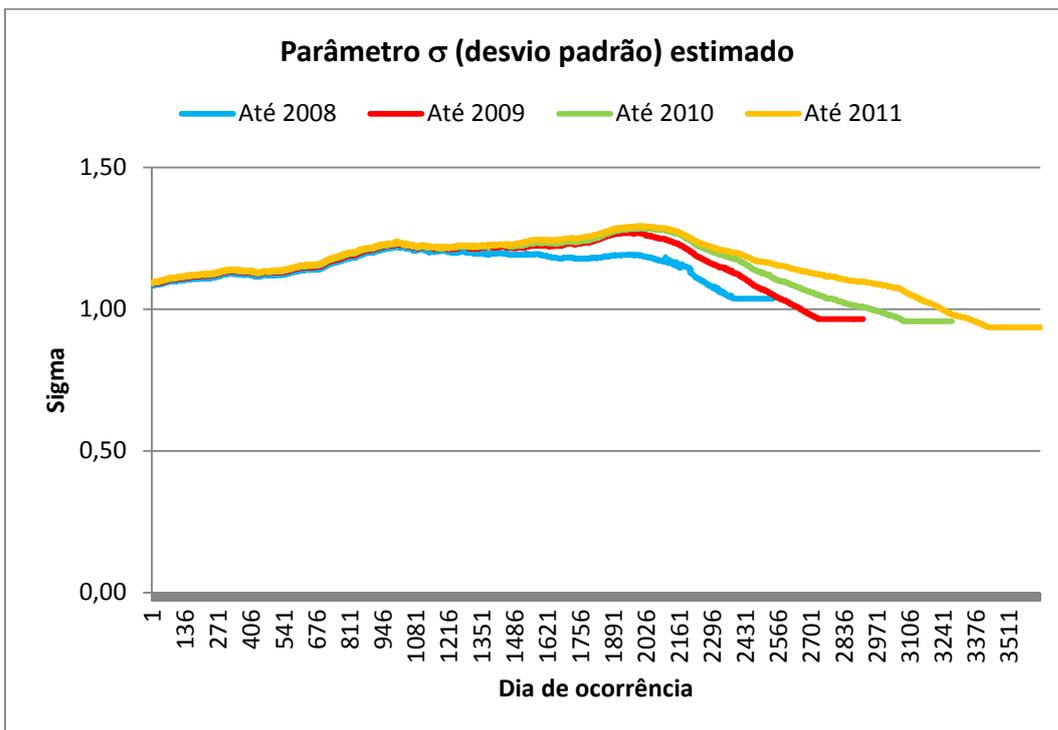


Figura 12: Parâmetro Abordagem Weissner

PUC-Rio - Certificação Digital Nº 1213316/CA

A Figura 13 apresenta o gráfico dos atrasos observados versus o atraso médio esperado para cada dia de ocorrência dos dados até 2011; o atraso médio esperado segue bem o padrão da distribuição de dados observados. Ao final do

gráfico para as estimativas serem mais confiáveis, elas foram repetidas e com isso o atraso médio não acompanha a diminuição que ocorre nos dados devido à proximidade com a última data observada.

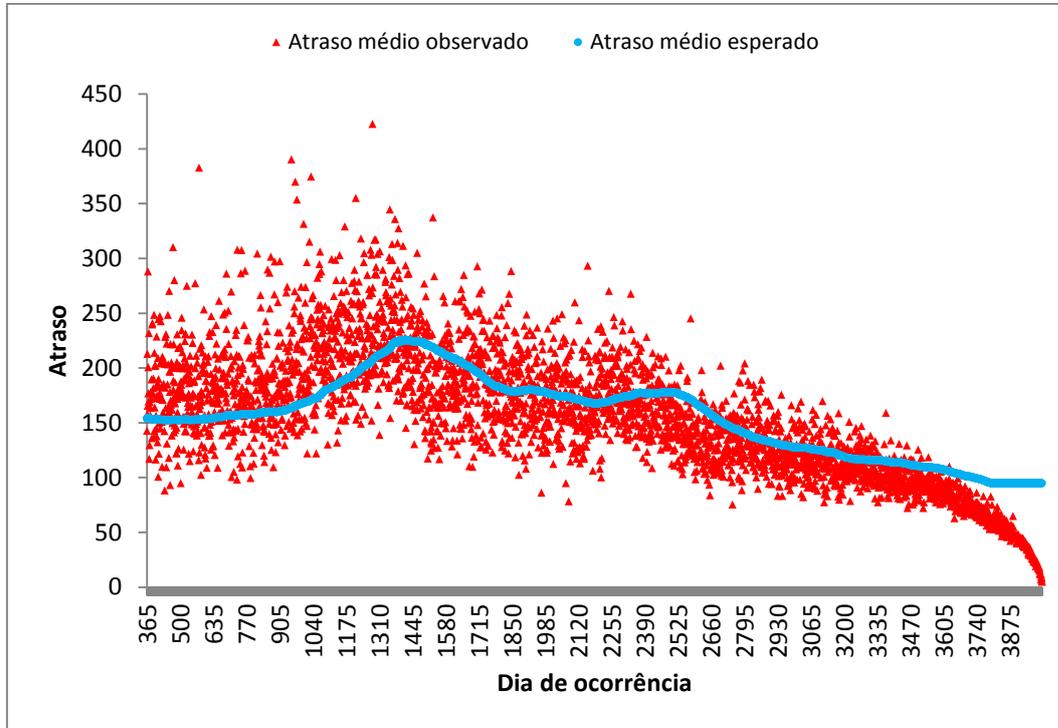


Figura 13: Valor esperado do atraso estimado x média amostral por dia de ocorrência (dados até 2011)– Abordagem Weissner

O *ultimate* é o valor esperado da quantidade total de sinistros ocorridos, o que inclui os sinistros já avisados e os que ainda serão avisados (IBNR). A seguir são apresentados os gráficos da quantidade estimada de sinistros (*Ultimate*) versus a quantidade observada na amostra. Os valores estimados estão muito próximos dos observados para todos os cortes, no final dos gráficos que é possível ver alguma diferença entre esses valores. No ano de 2010, o último valor estimado ficou muito alto e com isso alterou a escala do gráfico.

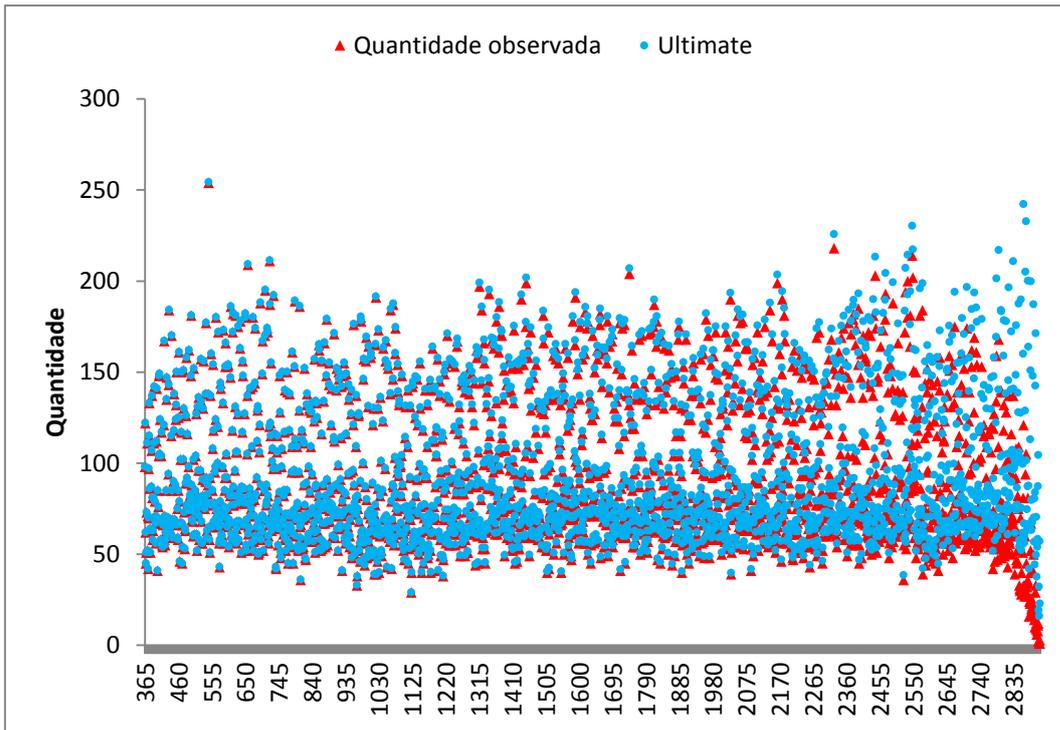


Figura 14: Quantidade observada x Quantidade estimada por dia corte até 2008 -
Abordagem Weissner

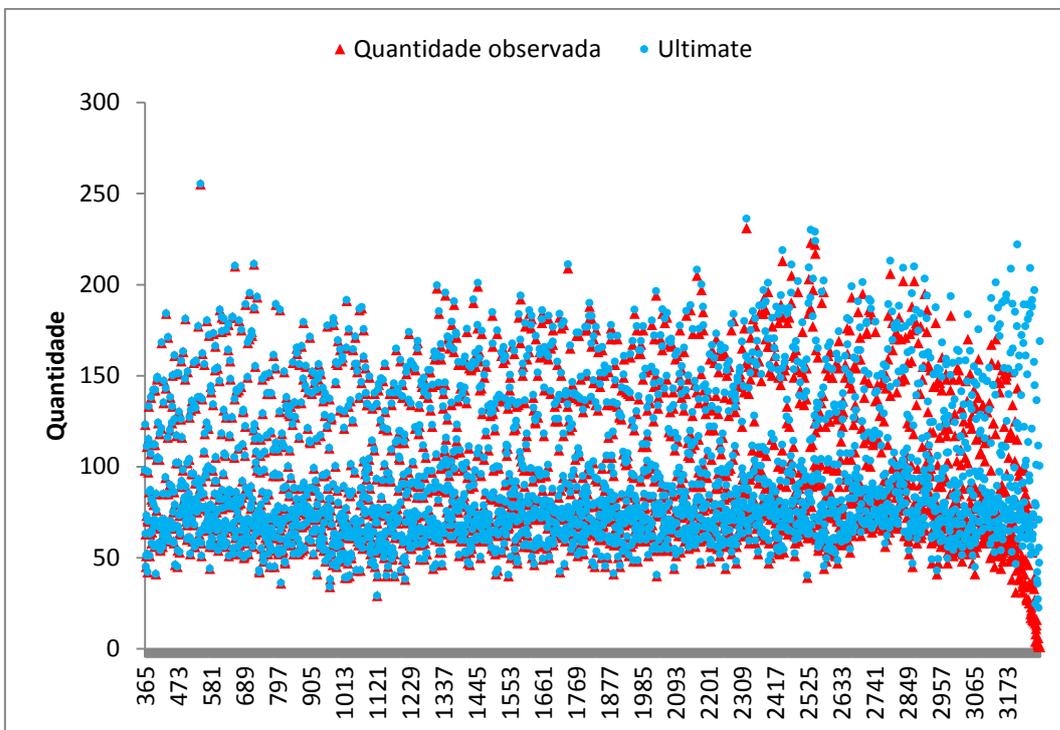


Figura 15: Quantidade observada x Quantidade estimada por dia corte até 2009 -
Abordagem Weissner

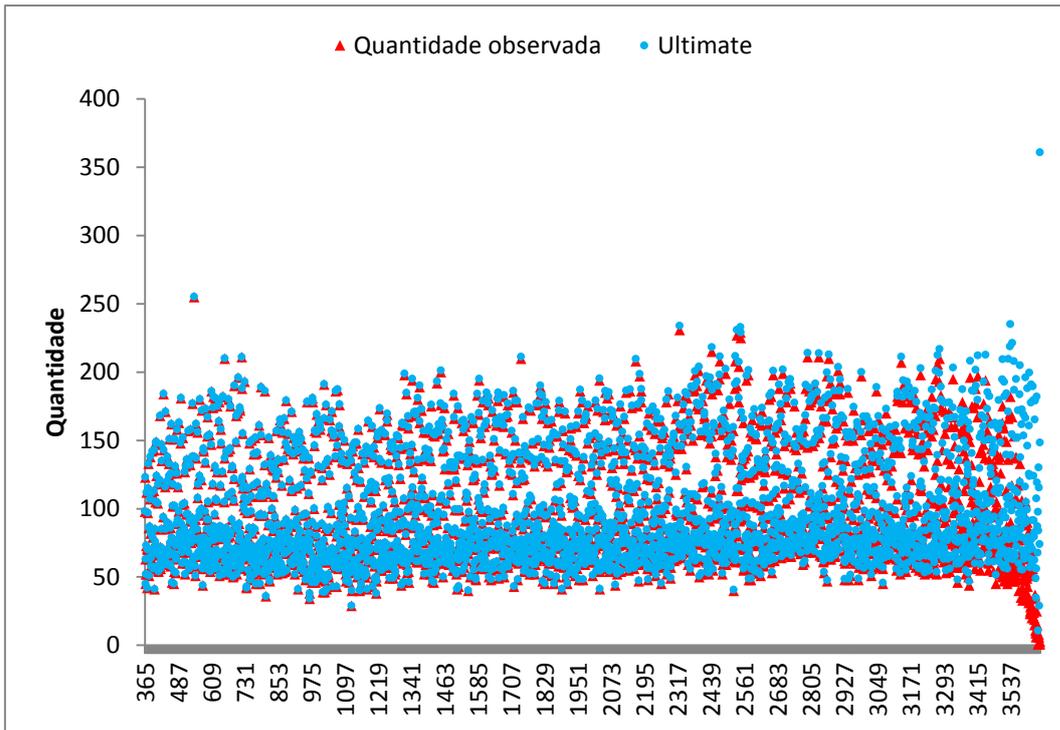


Figura 16: Quantidade observada x Quantidade estimada por dia corte até 2010 -
Abordagem Weissner

PUC-Rio - Certificação Digital Nº 1213316/CA

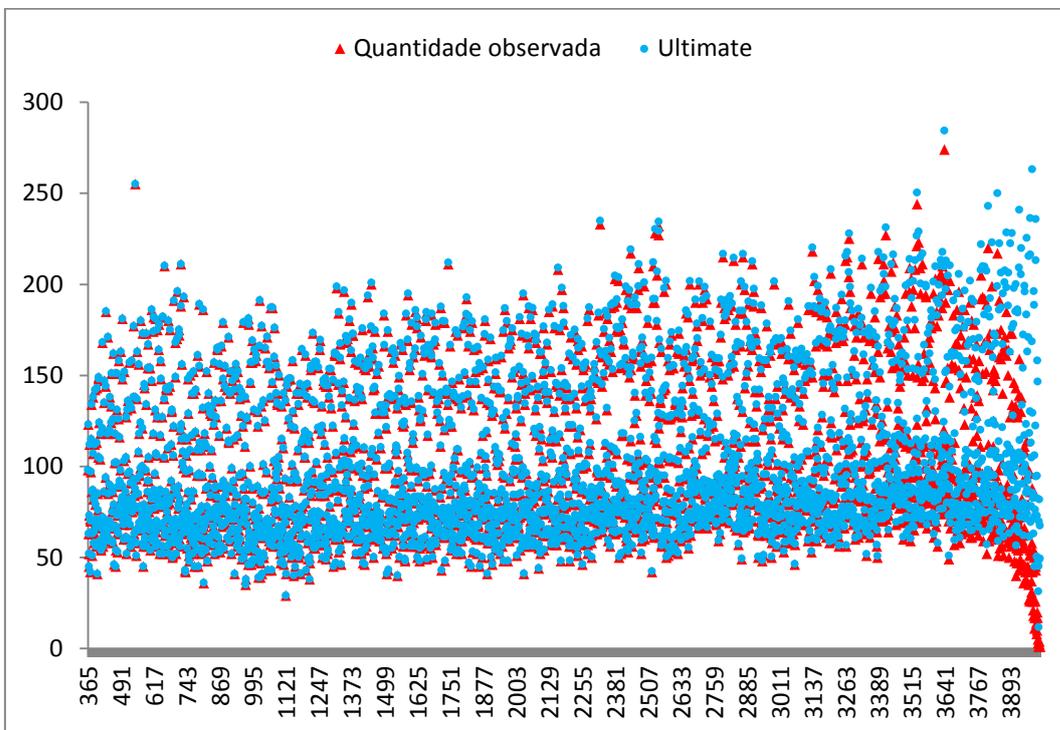


Figura 17: Quantidade observada x Quantidade estimada por dia até 2011 -
Abordagem Weissner

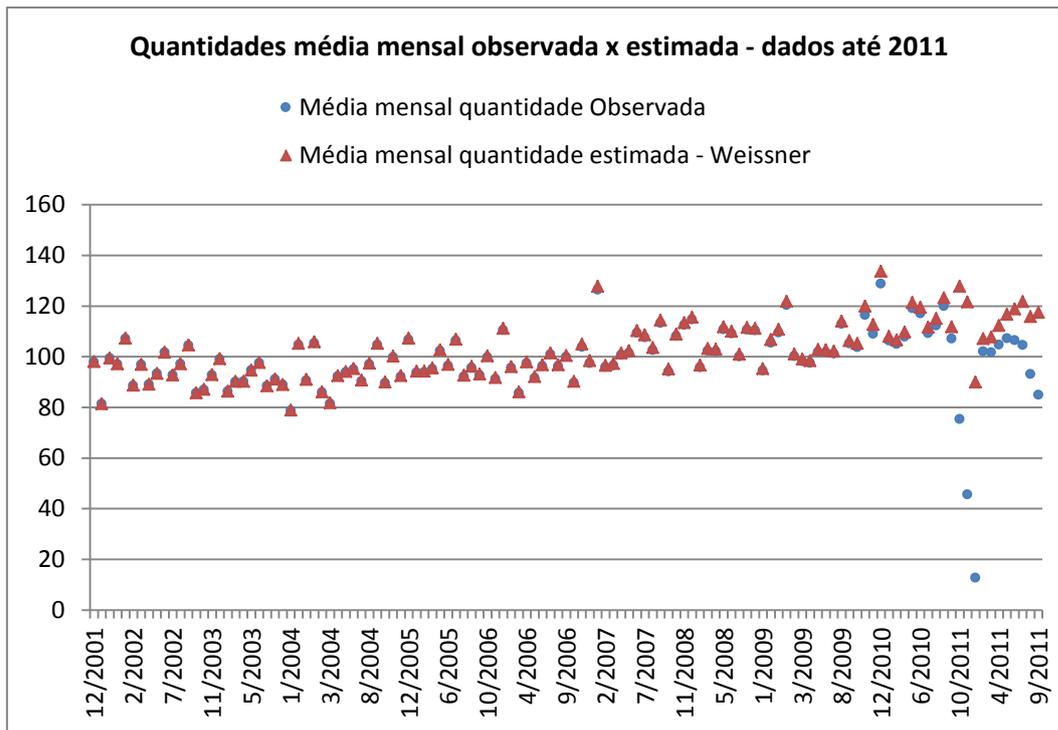


Figura 18: Gráfico de médias mensais da quantidade observada x quantidade estimada por dia (dados até 2011) – Abordagem Weissner

Além dos gráficos por dia de ocorrência dos sinistros para cada corte de dados, foi feito também um gráfico usando a média dos meses comparando as quantidades observadas e estimadas. Na Figura 18 está o gráfico considerando todos os dados até 2011, neste gráfico nota-se que a média mensal da quantidade estimada apresenta valores bem próximos à média mensal amostral, destoando apenas nos últimos meses do gráfico, principalmente nos 3 primeiros meses de 2011.

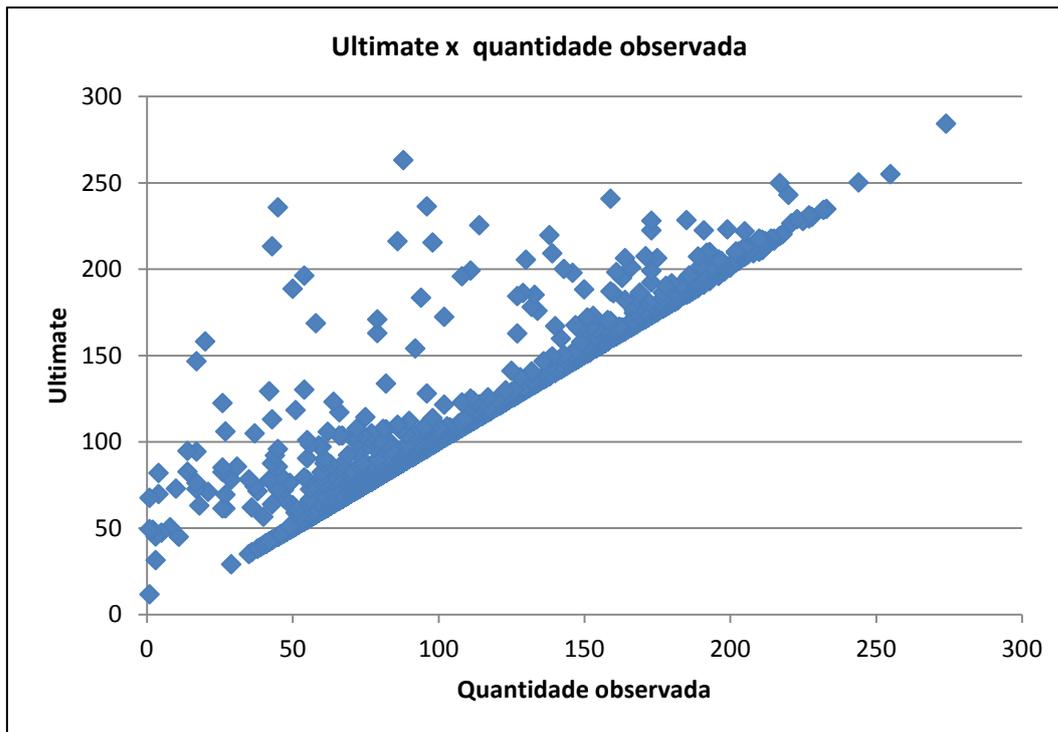


Figura 19: Gráfico de dispersão *Ultimate* x Observado (até 2011) – Abordagem Weissner

Também foi feito o gráfico de dispersão da quantidade estimada contra a observada. Nele fica mais claro de observar que os valores estimados são sempre maiores ou muito próximos dos observados.

Os erros de previsão gerados para cada corte são apresentados na Tabela 5. Para os 3 cortes os erros foram muito próximos, para o corte mais recente até dez/2010 este erro foi um pouco menor 26% o MAPE, no entanto o MAE e RMSE apresenta os maiores valores.

Tabela 5: Erros de previsão do método Weissner por corte

	Até Dez/2008	Até Dez/2009	Até Dez/2010	Previsão
MAPE	53%	46%	35%	IBNR 2011
MAE	2259	2421	5038	
RMSE	2671	2439	5038	

Com relação à previsão da quantidade IBNR usando todos os dados disponíveis até 2011, tem-se que o valor estimado foi de 11518 sinistros que ainda não foram reportados. Este valor é inferior ao estimado pelo *Chain ladder*, que foi

de 25164 sinistros. Mais para frente serão comparados os três métodos usados ao mesmo tempo.

4.3

Modelo conjunto quantidade/atraso

Como já foi dito, os parâmetros da distribuição foram estimados usando uma janela deslizante de 365 dias e foram realizadas os mesmos exercícios que para o método *Chain ladder* e a abordagem de Weissner.

O modelo proposto também leva em consideração que a quantidade de sinistros observada é uma variável aleatória com distribuição binomial que depende dos parâmetros da distribuição de atrasos, do atraso máximo observável e do parâmetro γ da distribuição do número total de sinistros ocorridos em cada dia. Os parâmetros das três distribuições foram estimados conjuntamente.

A Figura 20 apresenta os atrasos esperados por dia de ocorrência para cada um dos cortes utilizados.

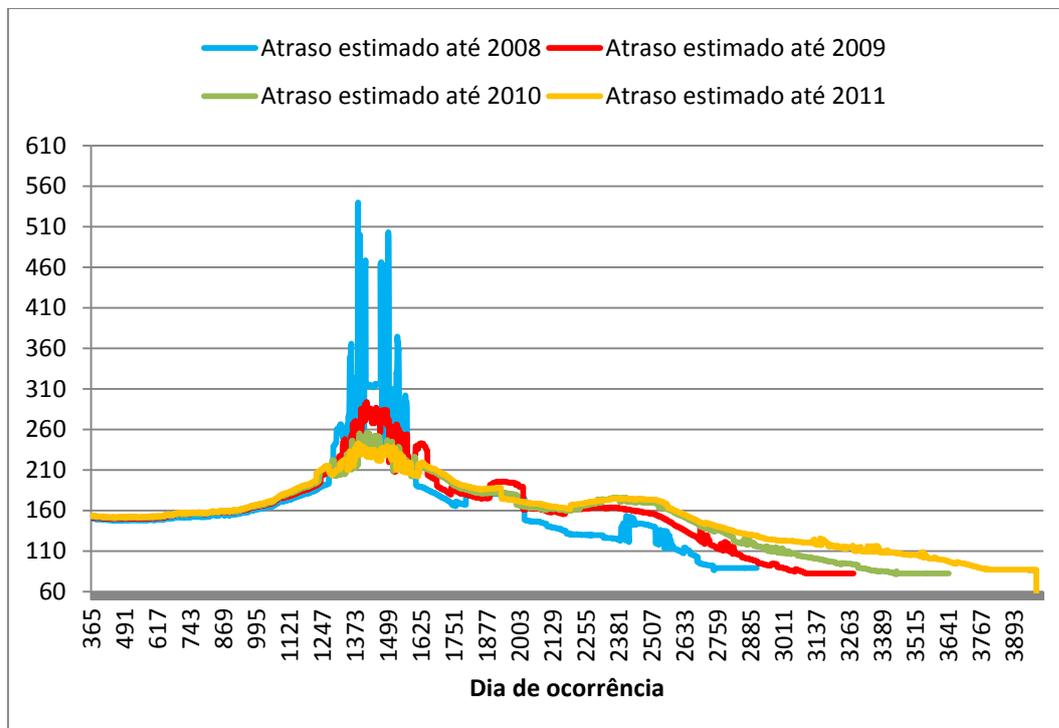


Figura 20: Atraso estimado por corte – Modelo conjunto

Apenas o corte ‘Até 2008’ possui alguns picos muito elevados em torno do dia 1300 (23/07/2004), um dos atrasos médios chega a quase 560 dias. Ao comparar com a Figura 10 do atraso médio estimado pelo método de Weissner, nota-se que neste período de tempo há um aumento no atraso médio, mas no caso do Weissner o valor do atraso médio não passa de 240 dias.

Na sequência, estão os valores estimados para os parâmetros μ e σ que são utilizados no cálculo do valor esperado da distribuição de atrasos. Vale lembrar que para a distribuição lognormal o valor esperado é dado por $E(X) = e^{\mu+1/2\sigma^2}$.

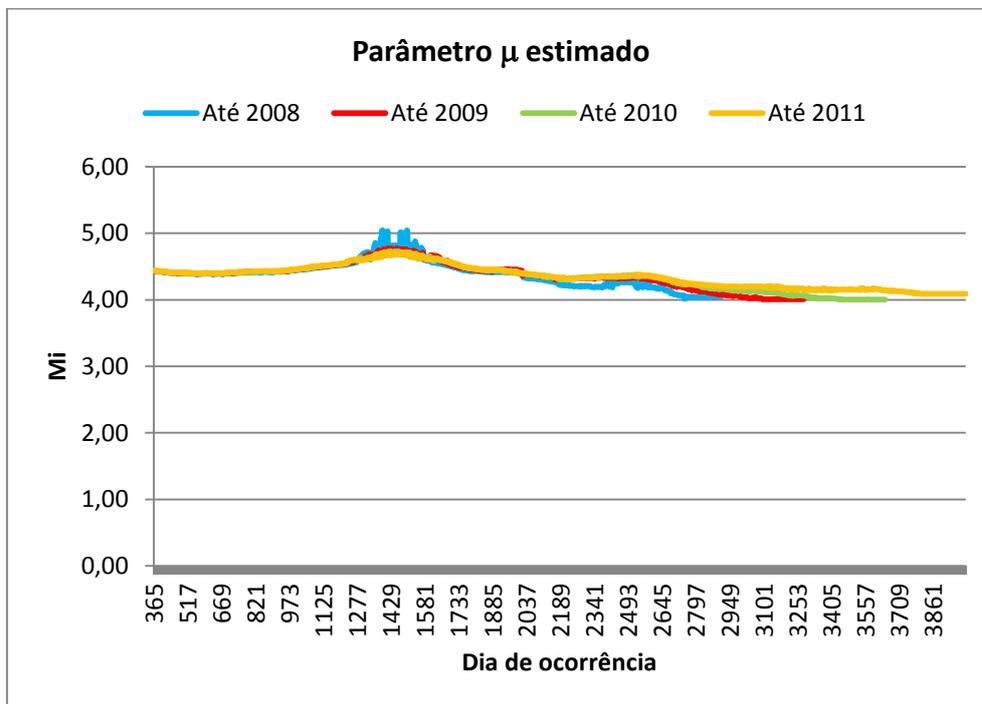


Figura 21: Parâmetro

Modelo conjunto

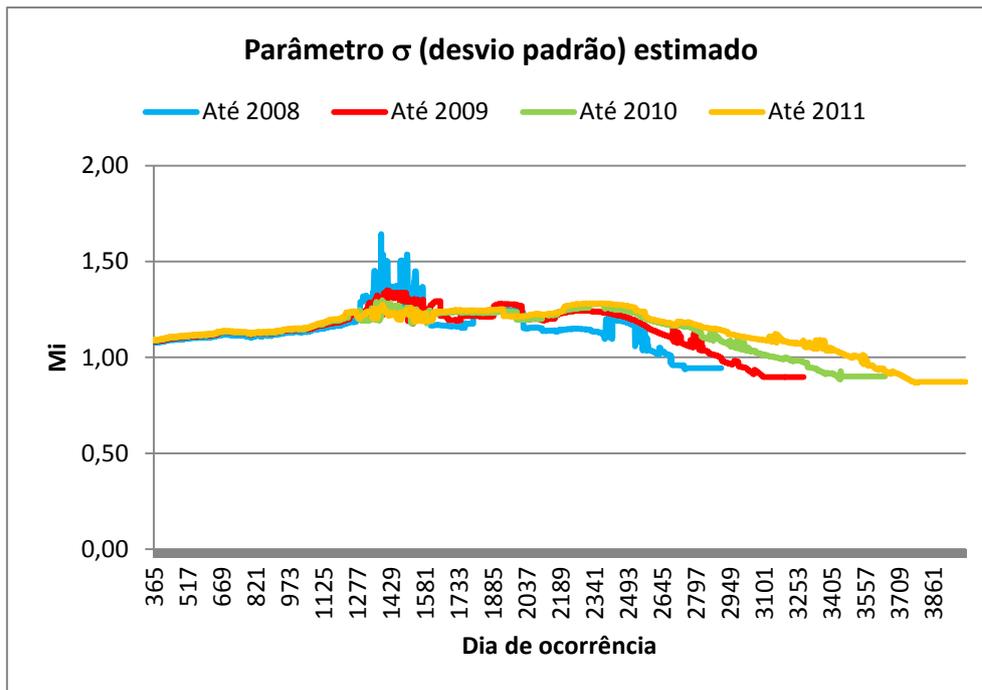


Figura 22: Parâmetro

Modelo conjunto

Ambos os parâmetros possuem uma variação maior no período onde estão os picos do atraso médio para o corte 2008.

Na Figura 23 está o gráfico de atraso médio versus os atrasos estimados para o corte 2008. É possível perceber que os parâmetros não estão bem estimados no período de maior atraso observado, pois só neste período da distribuição que o atraso médio estimado não se adequa bem à distribuição como nos demais períodos.

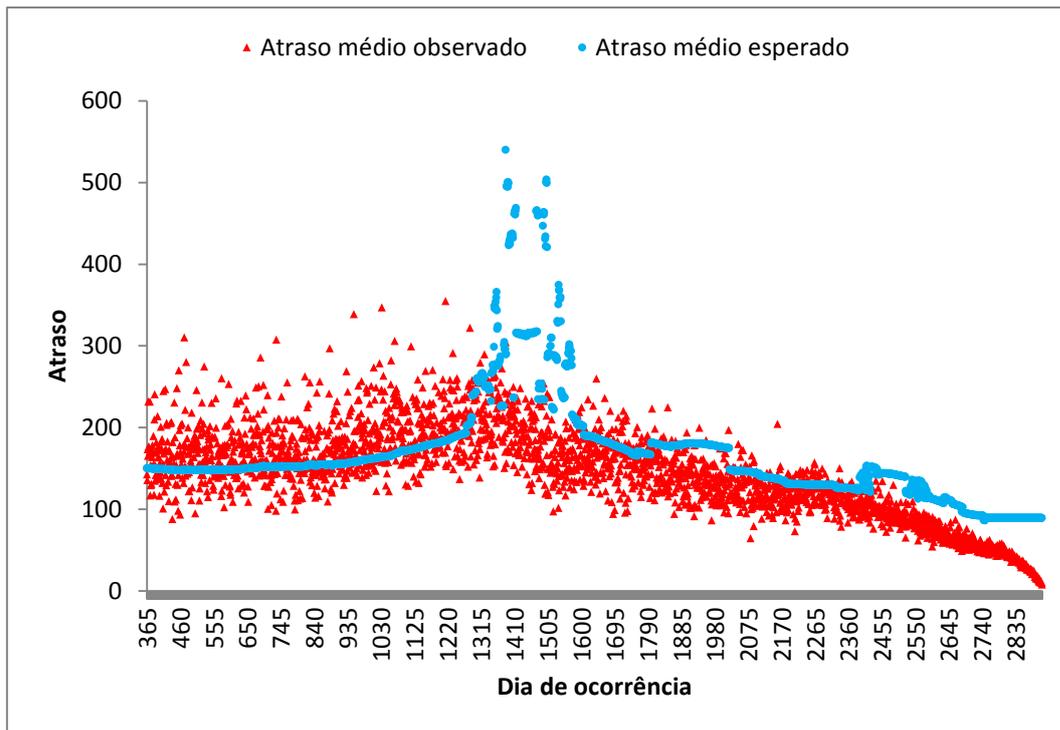


Figura 23: Atraso observado e atraso médio esperado para o corte 2008 – Modelo conjunto

É possível ver nas Figuras 24, 25 e 26 que para os outros cortes o atraso médio segue bem a distribuição de atrasos. Como já era de se esperar, conforme se possui mais dados, melhor se torna o ajuste do atraso médio aos dados observados.

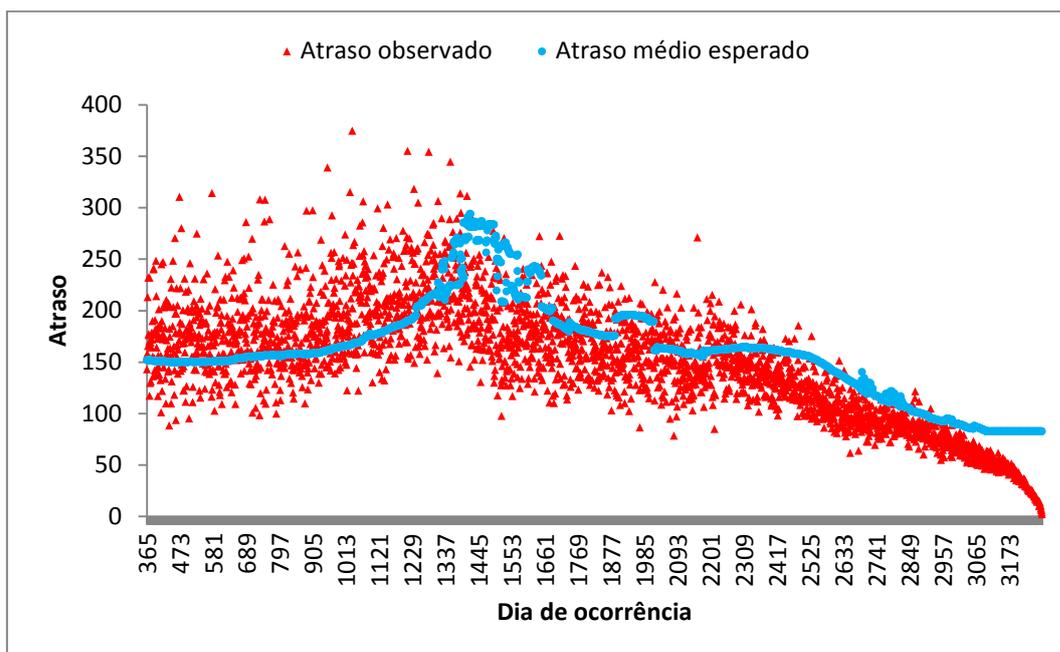


Figura 24: Atraso médio observado e esperado (corte 2009)- Modelo conjunto

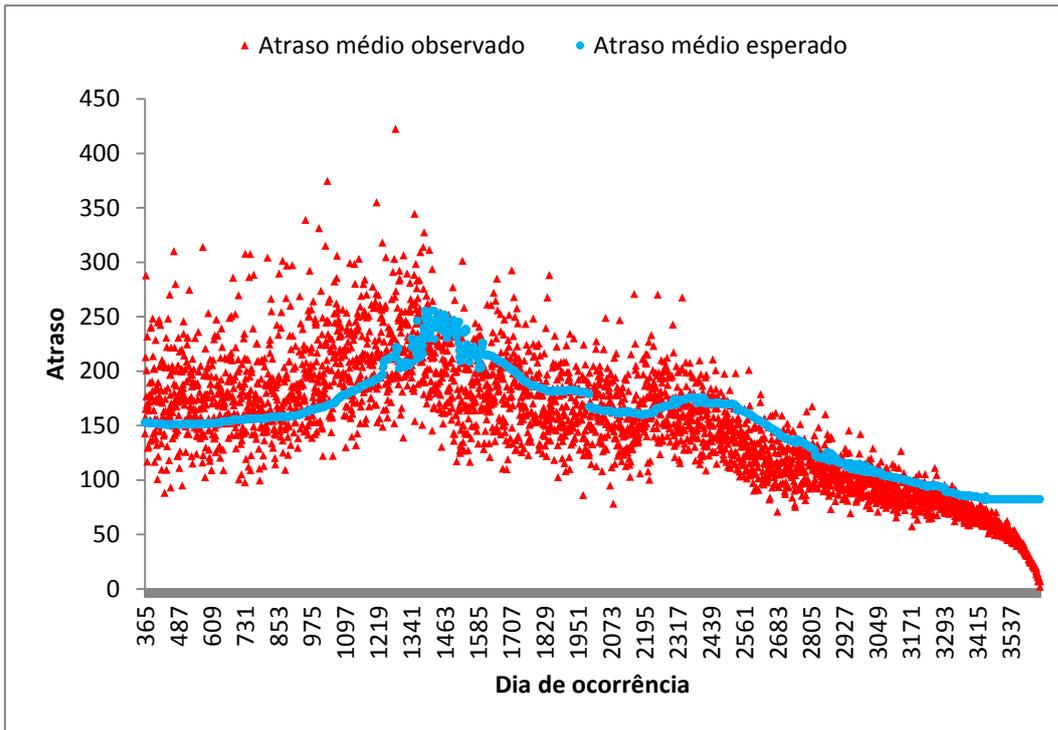


Figura 25: Atraso médio observado e esperado (corte 2010)- Modelo conjunto

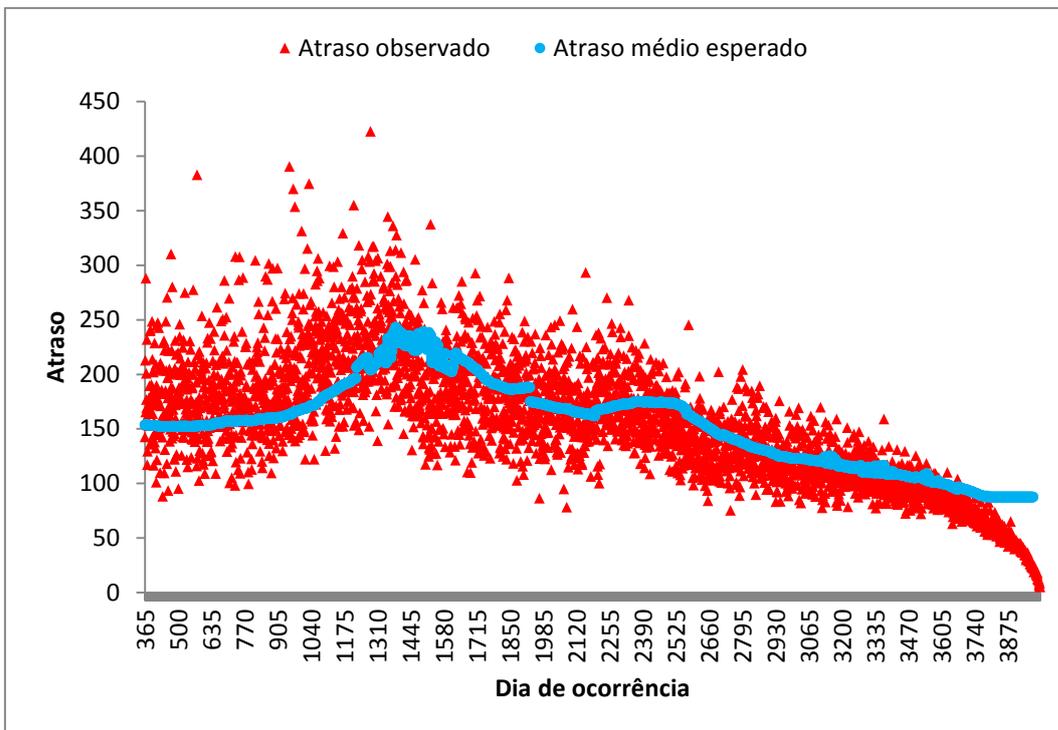


Figura 26: Atraso médio observado e esperado (corte 2011)- Modelo conjunto

Nas Figuras 27, 28, 29 e 30 temos as quantidades observadas versus as quantidades estimadas (*ultimate*) e o parâmetro estimado gama da distribuição da quantidade total de sinistros ocorridos para cada um dos cortes. Em todos os gráficos os valores estimados estão bem próximos das quantidades observadas, no final dos dados que o *ultimate* não acompanha a queda da quantidade observada, o que é esperado já que os dados são truncados na última data observada e essa diferença entre observado e estimado é a quantidade IBNR que se deseja estimar. O parâmetro gama estimado também está muito seguindo muito bem a distribuição de quantidades e se comporta como uma quantidade média da quantidade observada.

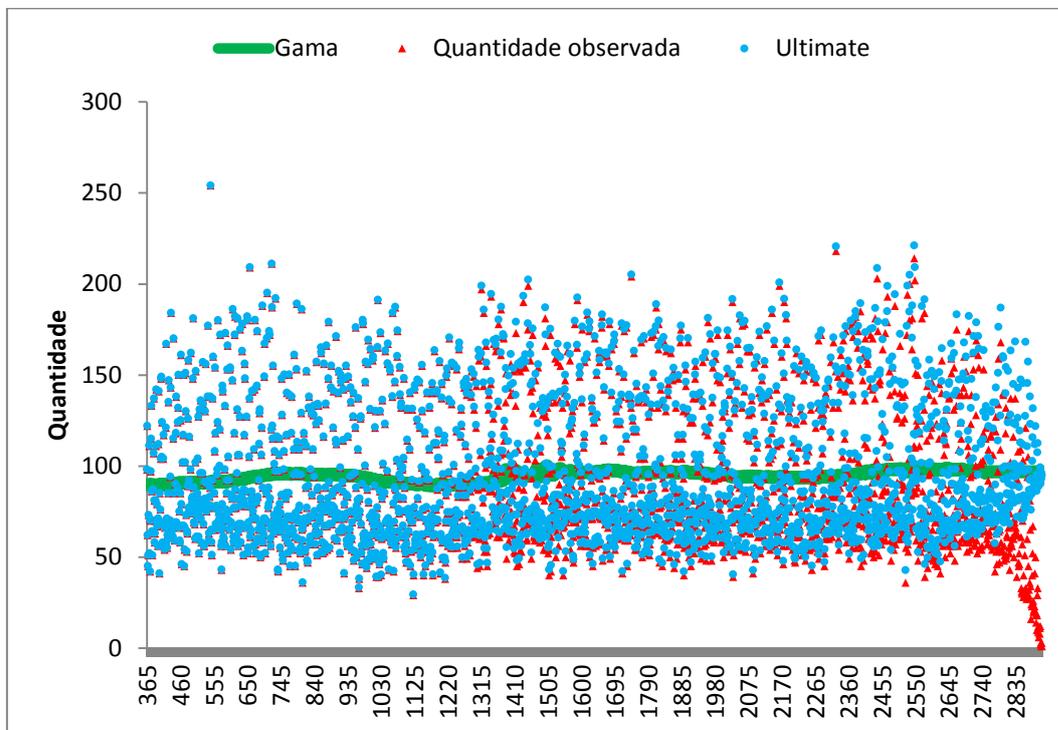


Figura 27: Quantidade observada x *Ultimate* e Gama estimado até 2008 –
Modelo conjunto

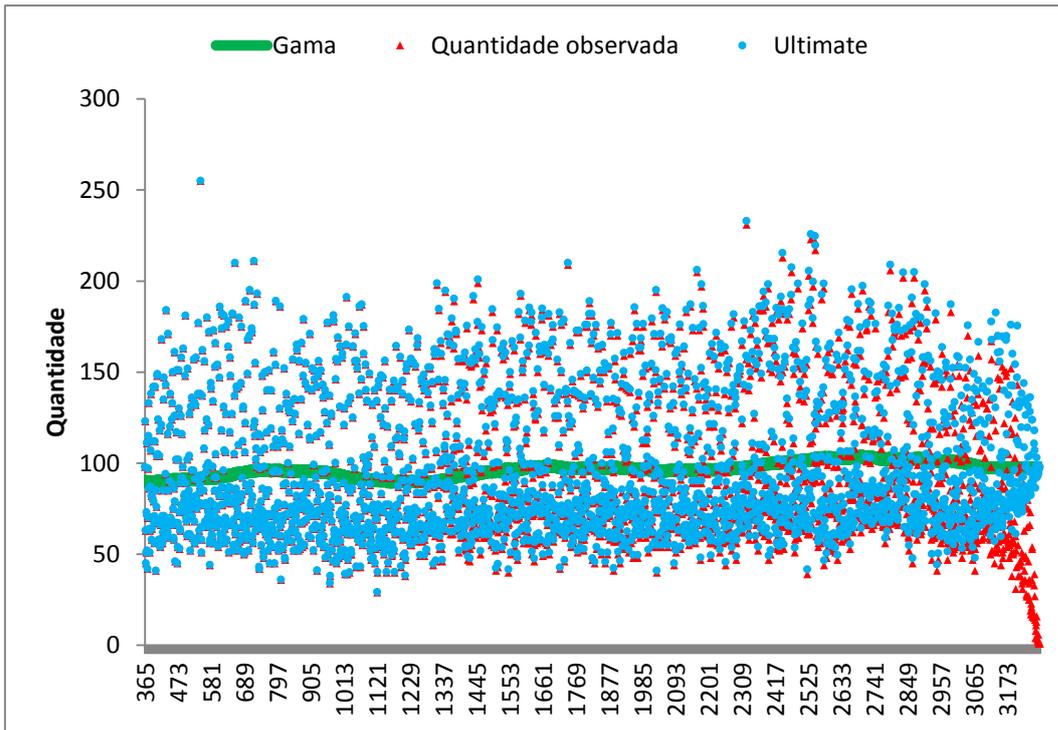


Figura 28: Quantidade observada x *Ultimate* e Gama estimado até 2009 – Modelo conjunto

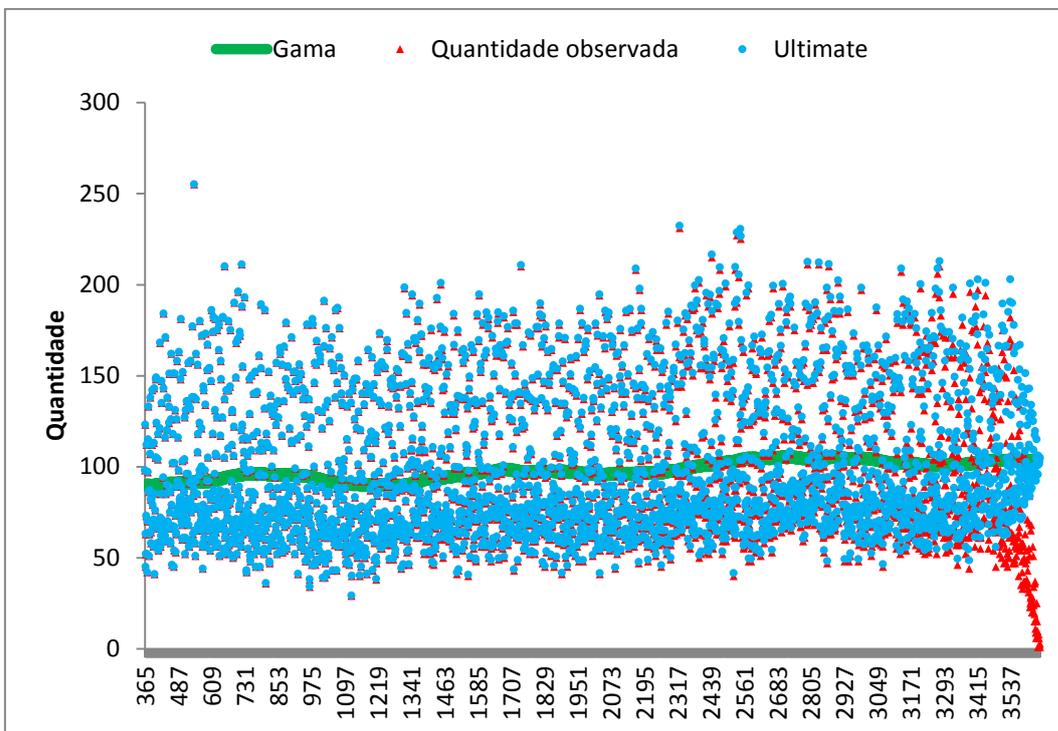


Figura 29: Quantidade observada x *Ultimate* e Gama estimado até 2010 – Modelo conjunto

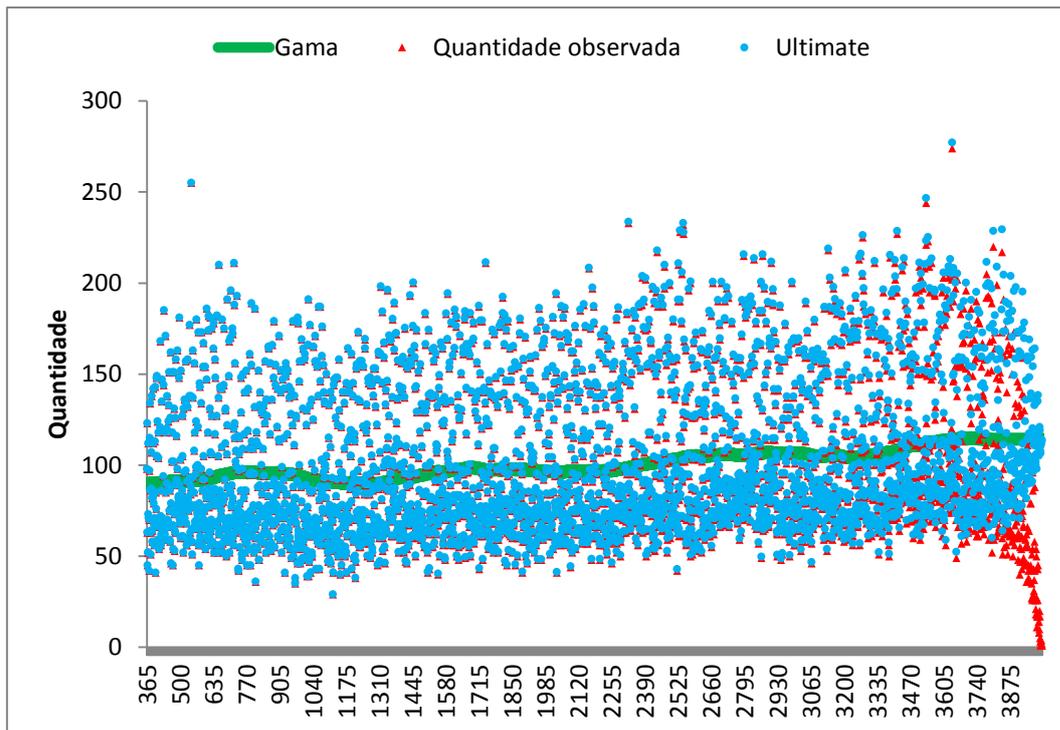


Figura 30: Quantidade observada x *Ultimate* e Gama estimado até 2011 –
Modelo conjunto

Assim como na abordagem Weissner, aqui também foi feito o gráfico de dispersão da quantidade estimada contra a observada. Muitos valores são muito próximo dos observados, o que indica que o modelo não prevê que ainda faltam quantidades a serem observadas.

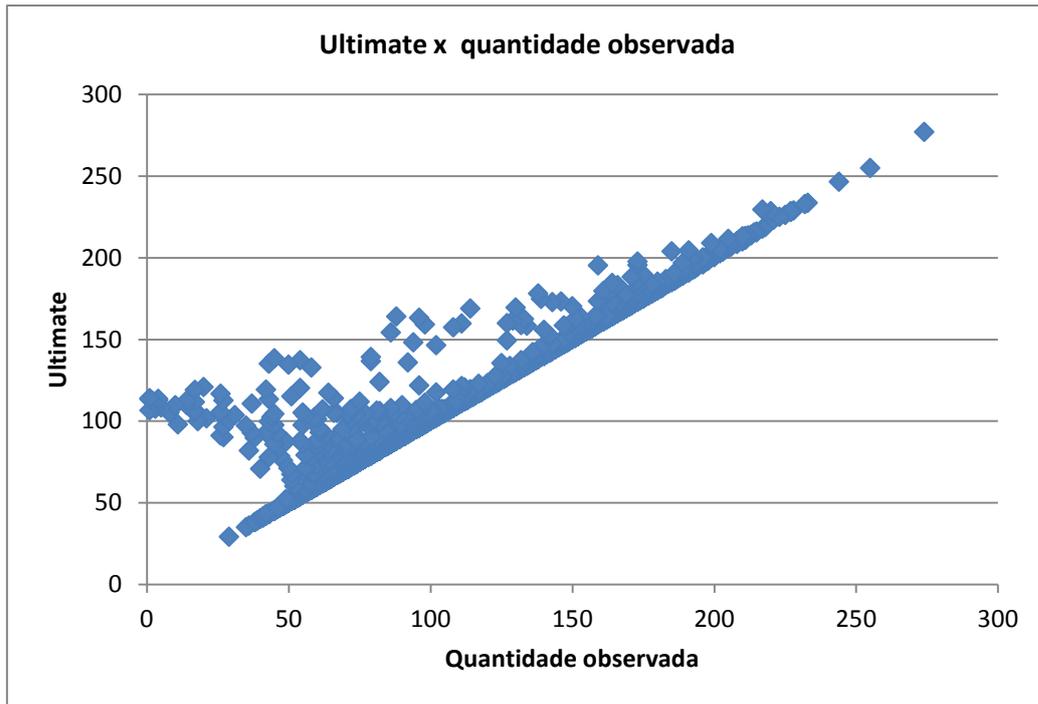


Figura 31: Gráfico de dispersão *Ultimate* x Observado (até 2011) – Modelo conjunto

E também o gráfico das quantidades médias mensais observadas e estimadas (Figura 32). Nele os valores ficam mais próximos que no Weissner, o que não é muito bom, pois indica que o modelo não espera que novos sinistros serão comunicados. Neste caso também tem-se os 3 primeiros meses com valores bem distintos dos observados, devido a uma grande queda na quantidade observada na amostra.

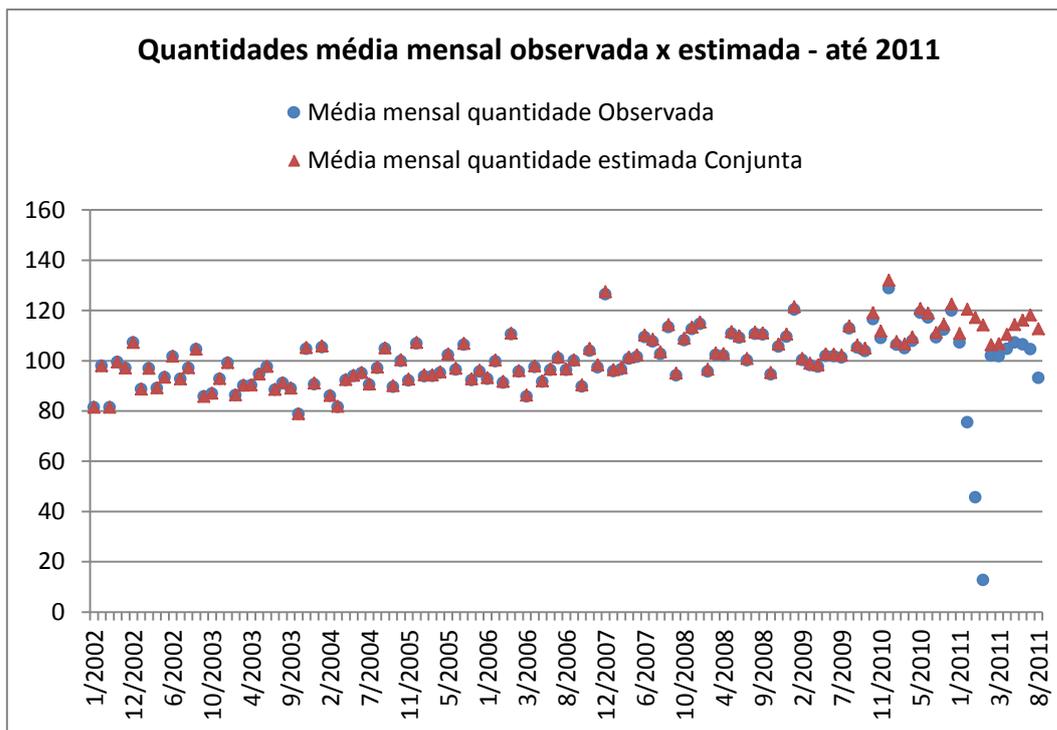


Figura 32: Gráfico de médias mensais da quantidade observada x quantidade estimada por dia (dados até 2011) – Modelo Conjunto

Na Tabela 6 a seguir estão apresentados os erros de previsão para cada um dos cortes de dados para o modelo conjunto da quantidade e atraso.

O valor estimado do IBNR para 2011 foi de 10668 sinistros que ainda não foram comunicados. Este valor está bem abaixo do valor estimado pelo método *Chain ladder* e também é um pouco inferior ao valor estimado pelo método do Weissner.

Tabela 6: Erros de previsão por cada corte e IBNR estimado para 2011

	Até Dez/2008	Até Dez/2009	Até Dez/2010	Previsão
MAPE	55%	50%	41%	IBNR 2011
MAE	2352	2725	5882	
RMSE	2779	2761	5882	

4.4

Comparação dos métodos

Neste tópico iremos comparar os resultados encontrados pelos três métodos. A Tabela 7 contém um resumo dos erros de previsão em cada corte e o valor IBNR estimado para o ano de 2011.

Tabela 7: Comparação erros de previsão e IBNR estimado

Método	Até dez 2008			Até dez 2009			Até dez 2010			Previsão
	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	IBNR
Chain Ladder	25%	930	1021	26%	1936	2232	9%	1237	1237	25164
Weissner	53%	2259	2671	46%	2421	2439	35%	5038	5038	11518
Conjunta	55%	2352	2779	50%	2725	2761	41%	5882	5882	10668

O método que utiliza a abordagem de Weissner e o modelo conjunto quantidade/atraso apresentaram erros de previsão bem superiores ao método *Chain ladder* em todos os cortes, mas o método Weissner obteve erros um pouco menores que o modelo conjunto.

A fim de aumentar as opções para comparação e dado que os mesmos dados foram utilizados e foram aplicados nas mesmas circunstâncias em ambos os trabalhos. A Tabela 8 apresenta os erros de previsão obtidos no trabalho de Souza (2013) utilizando tanto a abordagem de Weissner como o modelo conjunto quantidade/atraso no aviso, mas com as distribuições de atraso sendo a exponencial e uma mistura de exponenciais.

Tabela 8: Erros de previsão acrescentando resultados obtidos por Souza (2013)

Método	Até dez 2008			Até dez 2009			Até dez 2010			Previsão
	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	IBNR
Chain Ladder	25%	930	1021	26%	1936	2232	9%	1237	1237	25164
Weissner - Lognormal	53%	2259	2671	46%	2421	2439	35%	5038	5038	11518
Conjunta(lognormal-poisson)	55%	2352	2779	50%	2725	2761	41%	5882	5882	10668
Mistura de exponenciais	43%	1102	1142	24%	1160	1160	14%	2082	2082	19856
Mistura de exponenciais (Weissner)	46%	1299	1259	22%	856	924	20%	2878	2878	19721
Exponencial	75%	2853	3158	63%	3355	3377	40%	5832	5832	9781
Exponencial (Weissner)	75%	2844	3145	66%	3627	3685	44%	6389	6389	9057

O modelo conjunto utilizando a distribuição lognormal e o modelo com a abordagem de Weissner apresentaram erros de previsão melhores que os que usaram a distribuição exponencial, mas inferiores aos com a mistura de exponenciais. As estimativas para a quantidade IBNR do modelo conjunto e Weissner se aproximaram às obtidas pelo modelo exponencial, mas elas estão bem abaixo das estimativas dos outros modelos que apresentaram erros de previsão menores.

Também é interessante analisar se os modelos estão errando para cima ou para baixo quando fazem a previsão do valor observado. Estes valores podem ser vistos nos gráficos abaixo feitos para cada corte de dados.

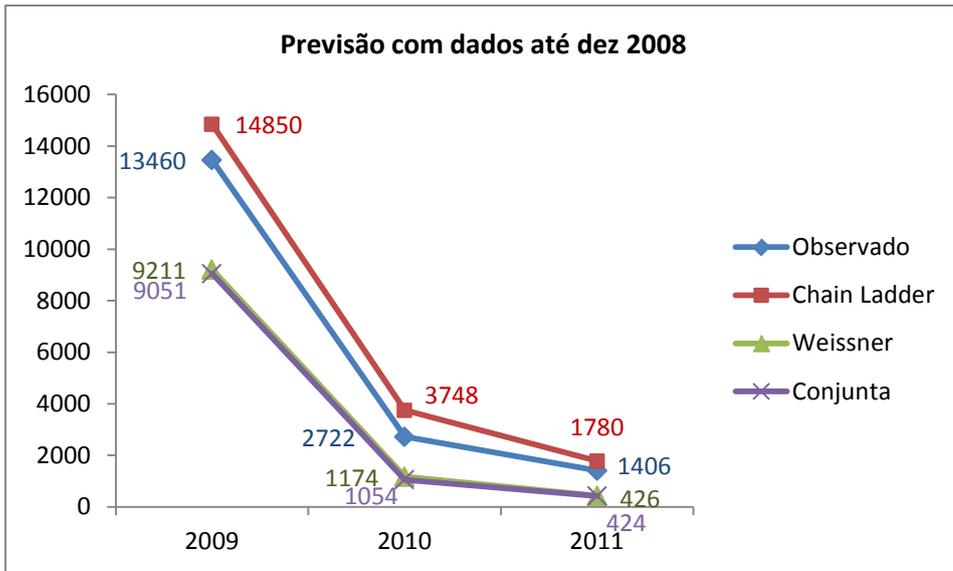


Figura 33: Gráfico de previsões 3 passos à frente com dados até 2008

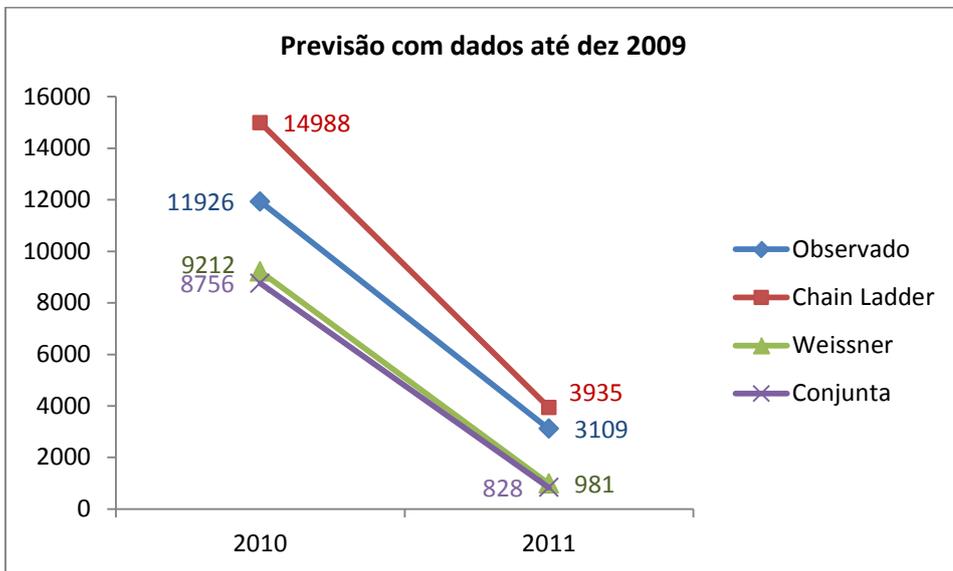


Figura 34: Gráfico de previsões 2 passos à frente com dados até 2009

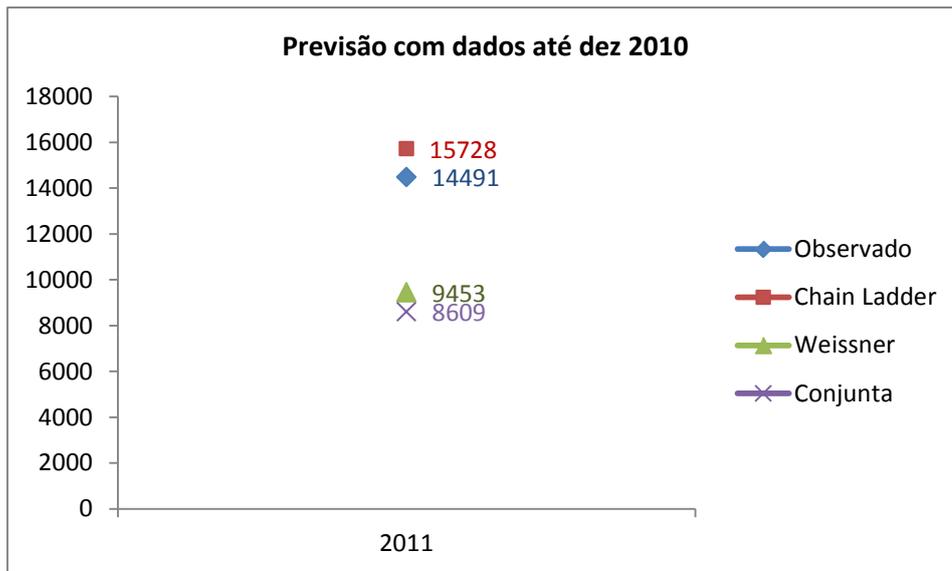


Figura 35: Gráfico de previsão 1 passo à frente com dados até 2010

Nos três gráficos nota-se que o método do *Chain ladder* sempre prevê acima do observado enquanto o método de Weissner e o modelo conjunto preveem abaixo do valor observado.

Intervalos de previsão obtidos no modelo conjunto quantidade/aviso

A quantidade IBNR para o período d (Q_d) tem distribuição $\text{Poisson}(\gamma_d F_{\tau_d}(T_d; \lambda_d))$, logo a quantidade total IBNR, que é a soma de variáveis independentes com distribuição Poisson, terá distribuição Poisson com média igual à soma das médias das variáveis Q_d . Desta maneira, o intervalo de previsão da distribuição total IBNR assim como o da quantidade IBNR para o período desejado saem diretamente pelos percentis da distribuição Poisson com média e variância correspondente de cada uma.

O intervalo de previsão de 90% para a quantidade IBNR total segundo a distribuição conjunta quantidade/aviso deste trabalho é [10499; 10839].

Os intervalos de previsão de 90% para a variável A_τ que representa a quantidade de sinistros avisados no ano τ estão na Tabela abaixo.

Tabela 9: Intervalos de previsão de 90% x quantidades observadas

Intervalo de previsão de 90% para a quantidade observada									
Horizonte	2008			2009			2010		
	Lim. Inf.	Lim. Sup.	Obs	Lim. Inf.	Lim. Sup.	Obs	Lim. Inf.	Lim. Sup.	Obs
2009	8748	9058	13460						
2010	991	1097	2722	8412	8717	11926			
2011	388	456	1406	763	856	3109	8357	8660	14491

É possível observar que nenhum dos valores observados nesses períodos está contido nos intervalos de previsão, o que nos faz questionar se o parâmetro gama da distribuição Poisson atribuída essas variáveis foi subestimado, ou ainda se esta distribuição deveria ser substituída por outra que possua maior variabilidade em relação à média. Além disso, temos poucas informações disponíveis desta variável e com isso a análise dos intervalos de previsão não é totalmente robusta. Por outro lado, aumentar essa amostra significaria perder mais dados na amostra utilizada para o ajuste dos modelos, o que não seria bom para a qualidade do ajuste.

5

Conclusão

O desenvolvimento do modelo conjunto proposto permitiu estimar uma distribuição dos atrasos no aviso da ocorrência que, ao analisar os gráficos de atraso esperado nota-se que se adequou muito bem aos dados observados, exceto para o corte 2008 que apresentou alguns picos. Além disso, foi possível observar o comportamento dos parâmetros da distribuição de atraso ao longo do tempo. No entanto, as previsões ficaram abaixo dos valores observados nas amostras removidas para a etapa de estimação. Algum dos parâmetros do modelo pode ter sido subestimado e com isso afetar nas previsões da quantidade IBNR. Este pode ser um dos fatores que fez com que os erros de previsão do modelo proposto fossem altos quando comparados aos demais. Em termos de resultados, o modelo de Weissner usando a distribuição de atrasos no aviso do sinistro usando a log-normal truncada apresentou resultados melhores que o modelo conjunto, tendo os resultados mais próximos do *Chain ladder*, que obteve os menores erros de previsão.

Uma vantagem do modelo conjunto foi ter a distribuição da quantidade IBNR bem definida e que pode ser obtida sem a necessidade de nenhuma simulação ou estimação adicional. Outro ponto positivo é que os parâmetros obtidos para a distribuição truncada são também os parâmetros da distribuição completa e assim não se faz necessário um processo adicional para estimar a cauda da distribuição de atrasos. E a estimação da distribuição de atrasos foi razoavelmente precisa, sendo obtida através da estimação com janelas deslizantes e atualização dinâmica das estimativas dos parâmetros das distribuições.

Um ponto negativo foi que o valor observado para a quantidade total de sinistros ocorridos não estava presente nos intervalos de previsão apresentados. Uma opção de correção aqui seria buscar outra distribuição com maior variabilidade.

Esta é uma das sugestões para trabalhos futuros: utilizar outra distribuição para a quantidade de sinistros ocorridos ou mesmo outra distribuição para os atrasos no aviso do sinistro. Ou ainda incluir sazonalidade na modelagem da

quantidade de sinistros ocorridos. Outra opção seria incluir informações externas, alguma medida de exposição ao sinistro, como, por exemplo, número de carros e outros veículos que estão em circulação.

Referências Bibliográficas

AL-ATHARI, M.M. **Estimation of the mean of truncated exponential distribution.** *Journal of Mathematics and Statistics*, v.4, n.4, p. 284– 288, 2008.

ANTONIO, K.; PLAT, R. **Micro-level stochastic loss reserving for general insurance.** 2012. 1.3, 3

ARJAS, E. **The claims reserving problem in non-life insurance: some structural ideas.** *Austin Bulletin*, v.19, n.2, p. 139-152, 1989. 1.3, 3, 3.3.

ATHERINO, R.S. **Um modelo em espaço de estado para estimativa de IBNR: revisitando De Jong & Zehnwirth.** Dissertação (Mestrado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2005.

ATHERINO, R.S. **Estimação de Reservas IBNR por Modelos em Espaço de Estado: Empilhamento por Linhas do Triângulo Runoff.** Tese de Doutorado (Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

COSTA, L.H. **Modelagem de sinistros IBNR com cauda: chainladder estendido, análise de regressão com heterocedasticidade e modelagem em espaço de estado linear.** Dissertação (Mestrado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2010.

DE JONG, P.; ZEHNWIRTH, B. **Claims reserving, state-space models and the Kalman filter.** *Journal of the Institute of Actuaries*, v.110, p.157–181, 1983.

DE JONG, P. **Forecasting general insurance liabilities.** *Department of Actuarial Studies Research Series*, 2004.

DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. **Maximum likelihood from incomplete data via the EM algorithm.** *Journal of the Royal Statistical Society Series B Methodological*, v.39, n.1, p. 1-38, 1977. 3.4.2.

DORAY, L.G. **Umvue of the ibnr reserve in a lognormal linearregression model.** *Insurance: Mathematics and Economics*, 18:43-57, 1996.

ENGLAND, P.D.; VERRALL, R.J. **Stochastic claims reserving in general insurance.** *Journal of the Institute of Actuaries*, 129:1-76, 2002.

FRIEDLAND, J. **Estimating unpaid claims using basic techniques.** *Casualty Actuarial Society*, 2010.

GUIAHI, F. **A probabilistic model for IBNR claims.** In: CAS Proceedings, Volume LXXIII, Number 139, 1986. 3, 3.1

HAASTRUP, S.; ARJAS, E. **Claims reserving in continuous time; a non parametric Bayesian approach.** *Insurance: Mathematics and Economics*, v.26, n.2, p. 139-164, 1997. 1.3.

HERTIG, J. **A statistical approach to ibnr-reserves in marine reinsurance.** *ASTIN Bulletin*, 15:171-183, 1985.

JEWELL, W. **Predicting IBNYR events and delays.** *Astin Bulletin*, v.20, n.11, p.93-111, 1989, 1.3, 3.

KALMAN, R.E. **A New Approach to Linear Filtering and Prediction Problems.** *Journal of Basic Engineering*, 82 (1): 35-45, 1960.

KAMINSKY, S. **Prediction of IBNR claim counts by modeling the distribution of report lags.** *Insurance: Mathematics and Economics*, v.6, p. 151-159, 1987. 3, 3.1.

KARLSSON, J.E. **The expected value of IBNR-claims.** In: CAS Proceedings, Volume LXXIII, Number 139, 1976. 3, 3.1

KUBRUSLY, J.Q. **Métodos Estatísticos para Cálculo de Reservas.** Dissertação (Mestrado em Engenharia Elétrica) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2005.

MACK, T. **Measuring the variability of Chain ladder reserve estimates.** *Casualty Actuarial Society Spring Forum*, p. 101-182, 1993.

MACK, T. **Which stochastic model is underlying the Chain ladder method?** *Insurance: Mathematics and Economics*, 15: 133-138, 1994.

NORBERG, R. **Prediction of outstanding liabilities in non-life insurance.** *Astin Bulletin*, v.23, n.I, p. 95-115, 1993. 1.3, 3, 3.3.

NORBERG, R. **Prediction of outstanding liabilities. II Model variations and extensions.** *Astin Bulletin*, v.29, n.I, p. 5-25, 1999. 1.3.

PARODI, B.P. **Triangle-free reserving: a non-traditional framework for estimating reserves and reserve uncertainty.** Número February, 2013.

SCHMIDT, K. **A bibliography on loss reserving.** 2012.

SOUZA, L. **Comparação de métodos de micro-dados e de triângulo run-off para previsão da quantidade IBNR.** Dissertação (Mestrado em Engenharia Elétrica) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2013.

TAYLOR, G.; MCGUIRE, G.; GREENFIELD, A. **Loss reserving: past, present and future.** *ASTIN Colloquium*, n.109, 2003.

WEISSNER, E.W. Estimation of the distribution of report lags by the method of maximum likelihood. *PCAS LXV*, 1978. 1.3, 3, 3.1, 3.2, 4.2.

Apêndice

A1- Máxima verossimilhança log-normal

Distribuição log-normal para atrasos para o dia de ocorrência t - Abordagem Weissner

Considere um determinado dia de ocorrência t . Seja k a quantidade de sinistros ocorridos em t e $\tau_1, \tau_2, \dots, \tau_k$ os tempos de atraso no aviso do sinistro observados de cada sinistro ocorrido no dia t . Considerando a distribuição log-normal para os tempos de atraso com parâmetros μ e σ^2 desconhecidos, temos:

$$f_{\tau}(\tau/\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma\tau}} e^{-\frac{1(\ln\tau-\mu)^2}{2\sigma^2}}$$

E a distribuição condicionada (truncada) ao atraso máximo T é dada por:

$$f_{\tau/\tau \leq T}(\tau/\mu, \sigma^2, T) = f(\tau/\mu, \sigma^2) / P(\text{atraso no aviso} \leq T)$$

$$f_{\tau/\tau \leq T}(\tau/\mu, \sigma^2, T) = \frac{1}{\sqrt{2\pi\sigma\tau}} e^{-\frac{1(\ln\tau-\mu)^2}{2\sigma^2}} / \Phi\left(\frac{\ln T - \mu}{\sigma}\right)$$

Onde $\Phi(\cdot)$ é a função acumulada da distribuição normal padrão $N(0,1)$.

A função de verossimilhança para os parâmetros μ e σ^2 , $L(\mu, \sigma^2)$, no período de ocorrência t é dada por:

$$L(\mu, \sigma^2; \tau_1, \dots, \tau_k) = \prod_{i=1}^k \left\{ \frac{1}{\sqrt{2\pi\sigma\tau_i}} e^{-\frac{1(\ln\tau_i-\mu)^2}{2\sigma^2}} / \Phi\left(\frac{\ln T - \mu}{\sigma}\right) \right\}$$

E a função de log verossimilhança, $l(\mu, \sigma^2)$, é dada por:

$$l(\mu, \sigma^2; \tau_1, \dots, \tau_k) = \sum_{i=1}^k \left\{ -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma^2) - \ln(\tau_i) - \frac{1}{2} \frac{(\ln\tau_i - \mu)^2}{\sigma^2} - \ln \left[\Phi\left(\frac{\ln T - \mu}{\sigma}\right) \right] \right\}$$

Estimador de máxima verossimilhança de μ

A derivada da função de log-verossimilhança em relação ao parâmetro μ :

$$\frac{\partial l(\mu, \sigma^2; \tau_1, \dots, \tau_k)}{\partial \mu} = \sum_{i=1}^k \left\{ \underbrace{\frac{(\ln\tau_i - \mu)}{\sigma^2}}_I - \underbrace{\frac{\partial}{\partial \mu} \ln \left[\Phi\left(\frac{\ln T - \mu}{\sigma}\right) \right]}_{II} \right\}$$

Desenvolvendo o termo II e fazendo $z = \left(\frac{\ln T - \mu}{\sigma}\right)$, temos:

$$\frac{\partial}{\partial \mu} \left\{ \ln \left[\Phi\left(\frac{\ln T - \mu}{\sigma}\right) \right] \right\} = \frac{\partial}{\partial \mu} \{ \ln[\Phi(z)] \} = \left(\frac{1}{\Phi(z)} \right) \left(\frac{\partial}{\partial z} \{ \Phi(z) \} \right) \left(\frac{\partial z}{\partial \mu} \right)$$

Como $z = \left(\frac{\ln T - \mu}{\sigma}\right)$, então $\frac{\partial z}{\partial \mu} = \frac{\partial \left(\frac{\ln T - \mu}{\sigma}\right)}{\partial \mu} = \frac{-1}{\sigma}$ e sendo $\frac{\partial}{\partial z} \{\Phi(z)\} = \phi(z)$,

obtemos:

$$\frac{\partial}{\partial \mu} \{\ln[\Phi(z)]\} = \frac{1}{\Phi(z)} \phi(z) \left(\frac{-1}{\sigma}\right) = \frac{-1}{\sigma} \frac{\phi(z)}{\Phi(z)}$$

Logo, fazendo a substituição de z , o termo II é dado por:

$$\frac{\partial}{\partial \mu} \left\{ \ln \left[\Phi \left(\frac{\ln T - \mu}{\sigma} \right) \right] \right\} = \frac{-1}{\sigma} \frac{\phi \left(\frac{\ln T - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)}$$

Desta forma, a derivada da log-verossimilhança em relação à μ é:

$$\begin{aligned} \frac{\partial l(\mu, \sigma^2; \tau_1, \dots, \tau_k)}{\partial \mu} &= \sum_{i=1}^k \left\{ \underbrace{\frac{(\ln \tau_i - \mu)}{\sigma^2}}_{\text{I}} - \underbrace{\frac{\partial}{\partial \mu} \ln \left[\Phi \left(\frac{\ln T - \mu}{\sigma} \right) \right]}_{\text{II}} \right\} \\ \frac{\partial l(\mu, \sigma^2; \tau_1, \dots, \tau_k)}{\partial \mu} &= \sum_{i=1}^k \left\{ \frac{(\ln \tau_i - \mu)}{\sigma^2} + \frac{1}{\sigma} \frac{\phi \left(\frac{\ln T - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)} \right\} \end{aligned}$$

Iguala-se a derivada a zero para obter o estimador de máxima verossimilhança para μ :

$$\begin{aligned} \frac{\partial l(\mu, \sigma^2; \tau_1, \dots, \tau_k)}{\partial \mu} &= \sum_{i=1}^k \left\{ \frac{(\ln \tau_i - \mu)}{\sigma^2} + \frac{1}{\sigma} \frac{\phi \left(\frac{\ln T - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)} \right\} = 0 \\ \sum_{i=1}^k \left\{ \frac{(\ln \tau_i - \mu)}{\sigma^2} + \frac{1}{\sigma} \frac{\phi \left(\frac{\ln T - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)} \right\} &= 0 \\ \frac{\sum_{i=1}^k (\ln \tau_i) - k\mu}{\sigma^2} &= -\frac{k}{\sigma} \frac{\phi \left(\frac{\ln T - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)} \\ \mu &= \frac{\sum_{i=1}^k (\ln \tau_i)}{k} + \sigma \frac{\phi \left(\frac{\ln T - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)} \end{aligned}$$

Estimador máxima verossimilhança de σ^2

A derivada da função de log-verossimilhança em relação ao parâmetro σ^2 :

$$\frac{\partial l(\mu, \sigma^2; \tau_1, \dots, \tau_k)}{\partial \sigma^2} = \sum_{i=1}^k \left\{ \underbrace{-\frac{1}{2\sigma^2}}_I + \underbrace{\frac{1}{2} \frac{(\ln \tau_i - \mu)^2}{\sigma^4}}_{II} - \underbrace{\frac{\partial}{\partial \sigma^2} \left\{ \ln \left[\Phi \left(\frac{\ln T - \mu}{\sigma} \right) \right] \right\}}_{III} \right\}$$

Desenvolvendo o termo III, fazendo $z = \left(\frac{\ln T - \mu}{\sigma} \right)$, temos:

$$III = \frac{\partial}{\partial \sigma^2} \left\{ \ln \left[\Phi \left(\frac{\ln T - \mu}{\sigma} \right) \right] \right\} = \frac{\partial}{\partial \sigma^2} \{ \ln[\Phi(z)] \} = \left(\frac{1}{\Phi(z)} \right) \left(\frac{\partial}{\partial z} \{ \Phi(z) \} \right) \left(\frac{\partial z}{\partial \sigma^2} \right)$$

Como $z = \frac{\ln T - \mu}{\sigma}$, então

$$\begin{aligned} \frac{\partial z}{\partial \sigma^2} &= \frac{\partial \left((\ln T - \mu)(\sigma^2)^{-1/2} \right)}{\partial \sigma^2} = (\ln T - \mu) \left(-\frac{1}{2} (\sigma^2)^{-3/2} \right) = -\frac{1}{2} \frac{(\ln T - \mu)}{\sigma^3} \\ &= -\frac{1}{2\sigma^2} \left(\frac{\ln T - \mu}{\sigma} \right) = -\frac{1}{2\sigma^2} z \end{aligned}$$

e sendo $\frac{\partial}{\partial z} \{ \Phi(z) \} = \phi(z)$, obtemos:

$$\frac{\partial}{\partial \mu} \{ \ln[\Phi(z)] \} = \frac{1}{\Phi(z)} \phi(z) \left(-\frac{1}{2\sigma^2} z \right) = -\frac{1}{2\sigma^2} z \frac{\phi(z)}{\Phi(z)}$$

Logo, fazendo a substituição de z , o termo III é dado por:

$$\frac{\partial}{\partial \sigma^2} \left\{ \ln \left[\Phi \left(\frac{\ln T - \mu}{\sigma} \right) \right] \right\} = -\frac{1}{2\sigma^2} \left(\frac{\ln T - \mu}{\sigma} \right) \frac{\phi \left(\frac{\ln T - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)}$$

Voltando a derivada em relação a σ^2 , tem-se:

$$\frac{\partial l(\tau; \mu, \sigma^2)}{\partial \sigma^2} = \sum_{i=1}^k \left\{ \frac{1}{2\sigma^2} \left[-1 + \frac{(\ln \tau_i - \mu)^2}{\sigma^2} + \left(\frac{\ln T - \mu}{\sigma} \right) \frac{\phi \left(\frac{\ln T - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)} \right] \right\}$$

Igualando a derivada a zero para obter o estimador de máxima verossimilhança para σ^2 :

$$\frac{\partial l(\mu, \sigma^2; \tau_1, \dots, \tau_k)}{\partial \sigma^2} = \sum_{i=1}^k \left\{ -\frac{1}{2\sigma^2} + \frac{1}{2} \frac{(\ln \tau_i - \mu)^2}{\sigma^4} + \frac{1}{2\sigma^2} \left(\frac{\ln T - \mu}{\sigma} \right) \frac{\phi \left(\frac{\ln T - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)} \right\} = 0$$

$$\sum_{i=1}^k \frac{1}{2\sigma^2} \left\{ -1 + \frac{(\ln \tau_i - \mu)^2}{\sigma^2} + \left(\frac{\ln T - \mu}{\sigma} \right) \frac{\phi \left(\frac{\ln T - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)} \right\} = 0$$

$$\sum_{i=1}^k (\ln \tau_i - \mu)^2 = \sigma^2 k \left(1 - \left(\frac{\ln T - \mu}{\sigma} \right) \frac{\phi \left(\frac{\ln T - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)} \right)$$

$$\sigma^2 = \frac{\sum_{i=1}^k (\ln \tau_i - \mu)^2}{k \left(1 - \left(\frac{\ln T - \mu}{\sigma} \right) \frac{\phi \left(\frac{\ln T - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T - \mu}{\sigma} \right)} \right)}$$

Em ambos os casos não é possível isolar o parâmetro na equação e obter a solução analítica para os estimadores, por isso métodos análise numéricas devem ser utilizados para obter as estimativas.

Distribuição log-normal de atrasos para todos os dias de ocorrência - Abordagem Weissner

Sejam as n variáveis k_1, k_2, \dots, k_n os totais de sinistros avisados para cada período $t=1,2,\dots,n$; seja $T_t=n-t$, as n variáveis representando o atraso máximo observável para cada período de tempo t e por fim $\tau_{t,i}$ o i -ésimo atraso observado no período de ocorrência t . As amostras de atraso $\tau_{t,i}$ seguem $f(\tau/\mu, \sigma^2, T_t)$, a distribuição log-normal de atrasos truncada do período de ocorrência t . Supondo independência entre elas, a função de verossimilhança para todos os dados será:

$$L(\mu, \sigma^2; \tau_1, \dots, \tau_k) = \prod_{i=1}^{k_1} f(\tau_{1,i} / \mu, \sigma^2, T_1) \prod_{i=1}^{k_2} f(\tau_{2,i} / \mu, \sigma^2, T_2) \dots \prod_{i=1}^{k_n} f(\tau_{n,i} / \mu, \sigma^2, T_n)$$

$$L(\mu, \sigma^2; \tau_1, \dots, \tau_k) = \prod_{t=1}^n \prod_{i=1}^{k_t} \left\{ \frac{1}{\sqrt{2\pi\sigma\tau_{t,i}}} e^{-\frac{1(\ln\tau_{t,i}-\mu)^2}{2\sigma^2}} / \Phi \left(\frac{\ln T_t - \mu}{\sigma} \right) \right\}$$

E a função de log-verossimilhança é dada por:

$$l(\mu, \sigma^2; \tau_{t,i}) = \sum_{t=1}^n \sum_{i=1}^{k_t} \left\{ -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma^2) - \ln(\tau_{t,i}) - \frac{1}{2} \frac{(\ln\tau_{t,i} - \mu)^2}{\sigma^2} - \ln \left[\Phi \left(\frac{\ln T_t - \mu}{\sigma} \right) \right] \right\}$$

E as equações dos estimadores para μ e σ^2 após derivar a função de verossimilhança em relação aos parâmetros e igualar a zero, são:

$$\mu = \frac{\sum_{t=1}^n \sum_{i=1}^{k_t} \ln \tau_{t,i}}{\sum_{t=1}^n k_t} + \sum_{t=1}^n \frac{\sigma k_t}{\sum_{t=1}^n k_t} \frac{\phi \left(\frac{\ln T_t - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T_t - \mu}{\sigma} \right)}$$

$$\sigma^2 = \frac{\sum_{t=1}^n \sum_{i=1}^{k_t} (\ln \tau_{t,i} - \mu)^2}{\sum_{t=1}^n k_t - \sum_{t=1}^n \left(\frac{\ln T_t - \mu}{\sigma} \right) \frac{\phi \left(\frac{\ln T_t - \mu}{\sigma} \right)}{\Phi \left(\frac{\ln T_t - \mu}{\sigma} \right)}}$$

Distribuição log-normal para os J dias de atrasos na janela deslizante

Na estimação feita por janelas deslizantes de J dias, os cálculos são os mesmos e a única diferença é que neste caso trabalha-se com J dados em cada uma das estimações. Sendo assim, os estimadores são dados por:

$$\mu = \frac{\sum_{t=1}^J \sum_{i=1}^{k_t} \ln \tau_{t,i}}{\sum_{t=1}^J k_t} + \sum_{t=1}^J \frac{\sigma k_t}{\sum_{t=1}^J k_t} \frac{\phi\left(\frac{\ln T_t - \mu}{\sigma}\right)}{\Phi\left(\frac{\ln T_t - \mu}{\sigma}\right)}$$

$$\sigma^2 = \frac{\sum_{t=1}^J \sum_{i=1}^{k_t} (\ln \tau_{t,i} - \mu)^2}{\sum_{t=1}^J k_t - \sum_{t=1}^J \left(\frac{\ln T_t - \mu}{\sigma}\right) \frac{\phi\left(\frac{\ln T_t - \mu}{\sigma}\right)}{\Phi\left(\frac{\ln T_t - \mu}{\sigma}\right)}}$$

A2- Quantidade IBNR por período de ocorrência

Considere um determinado dia de ocorrência t. Será avaliada a distribuição de probabilidade do número total de sinistros ocorridos em t, N, condicionado às informações disponíveis. Aplicando o teorema de Bayes à distribuição de probabilidade da variável N condicionada às informações disponíveis, quais sejam, $N \geq k_t, K = k_t, \mu, \sigma^2, \gamma$, pode-se considerar:

$$f_{N/N \geq k_t, K = k_t}(n_t; \mu, \sigma^2, \gamma)$$

$$\propto f_{K/N = n_t, N \geq k_t}(k_t; \mu, \sigma^2) f_{N/N \geq k_t}(n_t; \mu, \sigma^2, \gamma)$$

$$\propto f_{K/N = n_t}(k_t; \mu, \sigma^2) f_{N/N \geq k_t}(n_t; \gamma)$$

$$k_t = 0, 1, \dots \text{ e } n_t = k_t, k_t + 1, \dots$$

Incorporando as distribuições definidas de cada variável, tem-se:

$$f_{N/N \geq k_t, K = k_t}(n_t; \mu, \sigma^2, \gamma)$$

$$\propto \frac{n_t!}{k_t! (n_t - k_t)!} (1 - F_\tau(T, \mu, \sigma^2))^{k_t} (F_\tau(T, \mu, \sigma^2))^{n_t - k_t} \left[\frac{\gamma^{n_t} e^{-\gamma}}{n_t!} / P(N \geq k_t) \right]$$

$k=0, 1, \dots$ e $n=k, k+1, \dots$

Eliminando-se os termos que não dependem de n_t :

$$f_{N/N \geq k_t, K = k_t}(n; \mu, \sigma^2, \gamma)$$

$$\propto \frac{\gamma^{n_t}}{(n_t - k_t)!} (F_\tau(T, \mu, \sigma^2))^{n_t - k_t}$$

(multiplicando γ^{k_t}):

$$\propto \frac{[(F_\tau(T, \mu, \sigma^2)\gamma)]^{n_t - k_t}}{(n_t - k_t)!}$$

Ou seja, a distribuição condicional de N é uma distribuição assemelhada a uma Poisson para $k_t = 0, 1, \dots$ e $n_t = k_t, k_t + 1, \dots$ com parâmetro $F_\tau(T, \mu, \sigma^2)\gamma$. Isto equivale a dizer que $(N - k) \sim \text{Poisson}(F_\tau(T, \mu, \sigma^2)\gamma)$. Logo o valor esperado condicional de N é dado por:

$$\begin{aligned} & E[N/N \geq k_t, K = k_t; \mu, \sigma^2, \gamma] \\ = & E[(N - k)/N \geq k_t, K = k_t; \mu, \sigma^2, \gamma] + E[K/N \geq k_t, K = k_t; \mu, \sigma^2, \gamma] \\ & = F_\tau(T, \mu, \sigma^2)\gamma + k \end{aligned}$$

Se considerarmos a esperança de $N/N_t \geq k_t, K_t = k_t; \mu, \sigma^2, \gamma$ como a estimativa de N , temos:

$$\hat{n}_{\{N \geq k_t; \mu, \sigma^2, \gamma\}} = E[N/N \geq k_t, K = k_t; \mu, \sigma^2, \gamma] = F_\tau(T, \mu, \sigma^2)\gamma + k$$

Quantidade IBNR por período de aviso:

Seja Q a quantidade total de sinistros IBNR ocorridos em t e Q_τ a quantidade de sinistros ocorridos em t e avisados com atraso $\tau \in (\tau_1, \tau_2]$, onde τ_1 e τ_2 representam dois tempos de atraso quaisquer maiores que T_t (nas contas o índice t será omitido), Então $Q_\tau = \gamma p$, onde γ é o parâmetro da distribuição de N no período de ocorrência t e p representa a probabilidade de um sinistro ocorrido em t ser avisado com atraso $\tau \in (\tau_1, \tau_2]$.

No caso da distribuição log-normal, esta probabilidade é dada por:

$$p = P(\tau \leq \tau_2) - P(\tau < \tau_1) = \Phi\left(\frac{\ln \tau_2 - \mu}{\sigma}\right) - \Phi\left(\frac{\ln \tau_1 - \mu}{\sigma}\right)$$

ou

$$p = P(\tau > \tau_1) - P(\tau > \tau_2) = F_\tau(\tau_1, \mu, \sigma^2) - F_\tau(\tau_2, \mu, \sigma^2)$$

onde $\Phi(\cdot)$ é a distribuição acumulada da normal padrão e $F_\tau(x)$ é a probabilidade do atraso ser maior do que x .

Sabemos que $Q \sim \text{Poisson}(\gamma F_\tau(T, \mu, \sigma^2))$ e $(Q_\tau, Q) \sim f_{Q_\tau, Q}(r, q)$. Deseja-se obter a distribuição de probabilidade da quantidade de sinistros IBNR a serem avisados com atraso $\tau \in (\tau_1, \tau_2]$. É possível obter a probabilidade de um sinistro IBNR ser avisado com atraso $\tau \in (\tau_1, \tau_2]$ em função de p , conforme a seguir:

$$p_{\tau/T > T} = P(\tau_1 < \tau \leq \tau_2 | \tau > T) = \frac{P(\tau_1 < \tau \leq \tau_2)}{P(\tau > T)} = \frac{P(\tau > \tau_1) - P(\tau > \tau_2)}{P(\tau > T)}$$

$$= \frac{F(\tau_1, \mu, \sigma^2) - F(\tau_2, \mu, \sigma^2)}{F(T, \mu, \sigma^2)} = \frac{p}{F(T, \mu, \sigma^2)}$$

A partir dos resultados e definições acima, temos:

$$\begin{aligned} f_{Q_\tau} &= \sum_{q=0}^{\infty} f_{Q_\tau, Q}(r, q) \\ &= \sum_{q=0}^{\infty} f_{Q_\tau/Q=q}(r) f_Q(q) \\ &= \sum_{q=0}^{\infty} \frac{q!}{r! (q-r)!} (p_{\tau/\tau>T})^r (1 - p_{\tau/\tau>T})^{q-r} \frac{(\gamma F_\tau(T, \mu, \sigma^2))^q e^{-\gamma F_\tau(T, \mu, \sigma^2)}}{q!} \\ &= \frac{(p_{\tau/\tau>T})^r (\gamma F_\tau(T, \mu, \sigma^2))^r e^{-\gamma F_\tau(T, \mu, \sigma^2)}}{r!} \sum_{q=0}^{\infty} \frac{(1 - p_{\tau/\tau>T})^{q-r} (\gamma F_\tau(T, \mu, \sigma^2))^{q-r}}{(q-r)!} \\ &= \frac{(p_{\tau/\tau>T})^r (\gamma F_\tau(T, \mu, \sigma^2))^r e^{-\gamma F_\tau(T, \mu, \sigma^2)}}{r!} e^{(1-p_{\tau/\tau>T})(\gamma F_\tau(T, \mu, \sigma^2))} \\ &= \frac{(p_{\tau/\tau>T})^r (\gamma F_\tau(T, \mu, \sigma^2))^r e^{(1-p_{\tau/\tau>T})}}{r!} \\ &= \frac{(p_{\tau/\tau>T})^r (\gamma F_\tau(T, \mu, \sigma^2))^r e^{(1-p_{\tau/\tau>T})}}{r!} \\ &= \frac{(p\gamma)^r e^{(-p\gamma)}}{r!} \end{aligned}$$

Logo, $f_{Q_\tau} \sim \text{Poisson}(p\gamma)$.

Agora considerando τ como um período de aviso com quantidades de aviso estimáveis a partir do *triângulo run-off* e posterior ao máximo período de atraso observável, n ; isto é, $\tau = n + h$, $h=1, \dots, n-1$. Considere também $Q_{d, n-d+1+h}$ a variável que representa a quantidade de sinistros IBNR ocorridos em d e avisados com $n-d+1+h$ períodos de atraso e quantidade total de sinistros avisados no período τ , A_τ :

$$A_\tau = A_{n+h} = \sum_{d=h+1}^n Q_{d, n-d+1+h}$$

Como $Q_{d, n-d+1+h}$ tem distribuição Poisson, $A_\tau \sim \text{Poisson}(\sum_{d=h+1}^n p_{d, n-d+1+h} \gamma_d)$, onde $p_{d, n-d+1+h}$ é a probabilidade de um sinistro ocorrido em d ser avisado com atraso $n-d+1+h$.