

FUNDAÇÃO GETULIO VARGAS  
ESCOLA DE ECONOMIA DE SÃO PAULO

MARCOS HENRIQUE RIOS PEREIRA

ESTIMATIVA DE PROVISÕES DE IBNR UTILIZANDO  
ESPAÇO DE ESTADOS E FILTRO DE KALMAN  
UM CASO BRASILEIRO

SÃO PAULO  
2013

MARCOS HENRIQUE RIOS PEREIRA

ESTIMATIVA DE PROVISÕES DE IBNR UTILIZANDO  
ESPAÇO DE ESTADOS E FILTRO DE KALMAN  
UM CASO BRASILEIRO

Dissertação apresentada ao Programa de Mestrado Profissional em Economia da Fundação Getulio Vargas/EESP, como parte dos requisitos para a obtenção do título de Mestre em Economia, linha de Finanças Quantitativas.

Área de concentração: Economia

Orientador:  
Prof. Dr. Alexandre de Oliveira

SÃO PAULO  
2013

Pereira, Marcos Henrique Rios

Estimativa de provisões de IBNR utilizando Espaço de Estados e Filtro de Kalman - Um caso brasileiro / Marcos Henrique Rios Pereira - 2013.  
83 f.

Orientador: Alexandre de Oliveira.

Dissertação (MPFE) - Escola de Economia de São Paulo.

1. Modelo de Espaço de Estados. 2. Análise de séries temporais 3. Filtro de Kalman. 4. IBNR. 5. Seguros - Brasil. I. Oliveira, Alexandre. II. Dissertação (MPFE) - Escola de Economia de São Paulo. III. Título.

CDU 519.246.8

MARCOS HENRIQUE RIOS PEREIRA

ESTIMATIVA DE PROVISÕES DE IBNR UTILIZANDO  
ESPAÇO DE ESTADOS E FILTRO DE KALMAN  
UM CASO BRASILEIRO

Dissertação apresentada ao Programa de  
Mestrado Profissional em Economia da  
Fundação Getulio Vargas/EESP, como  
parte dos requisitos para a obtenção do  
título de Mestre em Economia, linha de  
Finanças Quantitativas.

**Data de aprovação:**

\_\_\_ / \_\_\_ / \_\_\_

**Banca Examinadora:**

---

Prof. Dr. Alexandre de Oliveira  
(Orientador)  
FGV - EESP

---

Prof. Dr. Afonso de Campos Pinto  
FGV - EESP

---

Prof. Dr. Oswaldo Luiz do Valle Costa  
Escola Politécnica da USP

*À Rosana pelo amor e carinho incondicionais...*

# Agradecimentos

Gostaria de agradecer ao Prof. Dr. Alexandre de Oliveira pela orientação, pelas longas conversas e pela sua paciência em repassar cada detalhe deste trabalho.

Um agradecimento especial ao Prof. Dr. Afonso de Campos Pinto por ter me aceito no seletivo grupo de alunos do Mestrado Profissional em Economia na linha de Finanças Quantitativas da Fundação Getúlio Vargas de São Paulo e por ter tido um acompanhamento tão próximo à toda a turma. Em especial, obrigado pelo apoio e incentivo pessoal para que eu pudesse terminar o curso.

Agradeço ao Prof. Dr. Alessandro Marques pela ajuda preciosa e fundamental quanto à codificação do Filtro de Kalman em MATLAB® utilizada neste trabalho, assim como a indicação de livros e artigos que foram fundamentais dentro da pesquisa.

Agradeço a todos os alunos da minha turma do Mestrado Profissional em Economia na linha de Finanças Quantitativas da Fundação Getúlio Vargas de São Paulo. Obrigado pela parceria, conversas e conselhos que foram fundamentais durante a minha jornada.

Ao meu chefe e *sponsor* Kazuyuki Okada pelo apoio e paciência durante a minha jornada.

Às minhas irmãs, sobrinhos, familiares e amigos que tiveram de entender a minha ausência nos últimos anos, sempre com a mesma desculpa de que eu tinha de estudar para o mestrado.

Aos meus pais, Oliveira Ramos Pereira e Laurita Rios Aponi Pereira, que sempre me motivaram e apoiaram na busca pelo conhecimento, incentivando-me a lutar pelos meus sonhos e por não permitirem que eu desistisse deste trabalho, mesmo quando já era certa a minha decisão.

Finalmente, agradeço de todo o meu coração à Rosana de Souza Santos, pelas repetidas revisões do texto e a toda a sua dedicação, amor e carinho incondicionais, sem os quais eu não teria forças e nem poderia ter me dedicado a este trabalho como ela possibilitou que acontecesse. ♡

Tudo é incerto e derradeiro.

Tudo é disperso, nada é inteiro.

Fernando Pessoa

# Resumo

Esta dissertação pretende discutir a provisão de sinistros do tipo IBNR, bem como qual a melhor forma de estimar estas provisões. Para tanto, serão utilizados dados reais de uma grande seguradora Brasileira para um produto de seguro de um ramo Não Vida. Serão utilizados no cálculo o clássico método *Chain Ladder* e em contrapartida um modelo de Espaço de Estados e Filtro de Kalman, discutindo as flexibilidades, vantagens e desvantagens de se utilizar tal metodologia.

**Palavras-chave:** Modelo de Espaço de Estados, Análise de séries temporais, Filtro de Kalman, IBNR, Seguros - Brasil.



# Abstract

This master thesis discusses the claims reserve of the IBNR type, as well as the best way to estimate these provisions. For this purpose will be used the real data from a large Brazilian insurer for an insurance product from a non-life business. Will be used in calculating the classic Chain Ladder method and against this a State Space model and Kalman Filter, discussing the flexibilities, advantages and disadvantages of use such methodology.

**Keywords:** State Space Models, Time Series Analysis, Kalman Filter, IBNR, Insurance - Brazil.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>15</b>
1.1	Objetivo . . . . .	15
1.2	Motivação . . . . .	15
1.3	Estrutura do trabalho . . . . .	16
<b>2</b>	<b>Revisão Bibliográfica</b>	<b>18</b>
2.1	IBNR . . . . .	18
2.2	Filtro de Kalman . . . . .	19
2.3	O modelo de Espaço de Estados no cálculo de IBNR . . . . .	20
2.4	Contribuições deste Trabalho . . . . .	21
<b>3</b>	<b>Sinistros em Seguros e a estimação de IBNR</b>	<b>22</b>
3.1	Sinistros em Seguros . . . . .	22
3.2	O Triângulo de <i>Run-off</i> . . . . .	22
3.2.1	Representação em Duplo Índice . . . . .	23
3.2.2	Forma Incremental . . . . .	23
3.2.3	Forma Acumulada . . . . .	24
3.2.4	Referência Diagonal . . . . .	24
3.3	Estimação pelo método <i>Chain Ladder</i> . . . . .	25
<b>4</b>	<b>Conceitos básicos</b>	<b>27</b>
4.1	Considerações iniciais . . . . .	27
4.1.1	Distribuição Normal . . . . .	27
4.1.2	Ruído Branco . . . . .	28

4.1.3	Passeio Aleatório . . . . .	28
4.2	Estimação por Mínimos Quadrados . . . . .	29
4.2.1	Estimador de Mínimos Quadrados Ordinários . . . . .	29
4.2.2	Estimador de Mínimos Quadrados Recursivos . . . . .	31
4.3	Modelos de Espaço de Estado . . . . .	32
4.3.1	Modelos de Espaço de Estados Lineares Gaussianos. . . . .	33
<b>5</b>	<b>O Filtro de Kalman Discreto</b>	<b>35</b>
5.1	Funcionamento do Filtro de Kalman . . . . .	36
5.1.1	Predição . . . . .	36
5.1.2	Atualização . . . . .	37
5.2	O Algoritmo do Filtro de Kalman Discreto . . . . .	38
5.3	Estimador de Máximo Verossimilhança . . . . .	39
<b>6</b>	<b>O modelo de De Jong &amp; Zenwirth</b>	<b>42</b>
6.1	Equações do Filtro de Kalman Utilizadas na Abordagem DJZ . . . . .	42
6.2	Premissas da Abordagem DJZ . . . . .	44
6.2.1	Índices de Inflação e Volume de Apólices . . . . .	45
6.2.2	A matriz de Parâmetros das Observações . . . . .	45
6.2.3	Estimativa do Vetor de Estados . . . . .	47
6.2.4	Matriz de Covariância do Ruído das Observações . . . . .	48
6.2.5	Inicialização Difusa do Filtro de Kalman . . . . .	48
6.3	Estimativa da Reserva de IBNR por EE e FK . . . . .	49
6.4	O exemplo DJZ . . . . .	50
6.4.1	Os dados DJZ . . . . .	50
6.4.2	Estimando a Provisão de IBNR pelo Método <i>Chain Ladder</i> para os Dados DJZ . . . . .	51
6.4.3	Heterocedasticidade da base DJZ . . . . .	52
6.4.4	Estimativa dos Parâmetros Calculados pelo Filtro de Kalman para os Dados DJZ . . . . .	54

6.4.5	A provisão de IBNR calculada pelo Filtro de Kalman em DJZ . . .	55
<b>7</b>	<b>Estimativa de IBNR, um Caso Brasileiro</b>	<b>58</b>
7.1	Os Dados SBR . . . . .	58
7.2	Estimativa pelo Método <i>Chain Ladder</i> para os Dados SBR . . . . .	61
7.3	Estimativa dos Parâmetros Calculados pelo Filtro de Kalman para os Dados SBR . . . . .	61
7.3.1	Normalidade dos Resíduos da Estimação . . . . .	62
7.3.2	Resultados da provisão de IBNR calculada pelo Filtro de Kalman .	64
7.4	Erros de estimação . . . . .	71
<b>8</b>	<b>Conclusões e Futuros trabalhos</b>	<b>74</b>
8.1	Conclusões . . . . .	74
8.2	Extensões . . . . .	75
	<b>Referências Bibliográficas</b>	<b>76</b>
<b>A</b>	<b>Resultados para o método <i>Chain Ladder</i> para os dados SBR</b>	<b>80</b>
<b>B</b>	<b>Resultados para o modelo de Espaço de Estados e Filtro de Kalman para os dados SBR</b>	<b>82</b>

# Lista de Figuras

4.1	Exemplo de Passeio Aleatório - <i>Randon Walk</i> . . . . .	29
4.2	Estimativa por MQO do Passeio Aleatório . . . . .	30
4.3	Estimativa por MQR para o Passeio Aleatório . . . . .	31
4.4	O valor de $\beta_t$ na estimativa por MQR . . . . .	32
5.1	Funcionamento do Filtro de Kalman . . . . .	39
5.2	Funcionamento do Filtro de Kalman com o Cálculo da Verossimilhança . .	41
6.1	Função $\Phi$ escolhida em DJZ . . . . .	46
6.2	Sinistros observados da base DJZ . . . . .	51
6.3	Momentos DJZ ao longo do tempo . . . . .	52
6.4	Erros DJZ ao longo do tempo . . . . .	52
6.5	Erros DJZ ao longo do tempo - Sem tendência . . . . .	53
6.6	Momentos DJZ ajustados ao longo do tempo . . . . .	53
6.7	Minimização dos parâmetros $\sigma_t^2$ e $\eta_t^2$ para DJZ . . . . .	55
6.8	Sinistros da base DJZ e $\hat{y}(t)$ estimado . . . . .	56
6.9	Dados DJZ - Triângulo de <i>Run-off</i> completo . . . . .	57
7.1	Sinistros já observados para os dados SBR . . . . .	60
7.2	Estatística de Jarque-Bera para os dados SBR com $5 \leq t \leq 10$ . . . . .	63
7.3	Estatística de Jarque-Bera para os dados SBR com $11 \leq t \leq 16$ . . . . .	63
7.4	Sinistros da base SBR com onze meses e $\hat{y}(t)$ estimado . . . . .	64
7.5	Sinistros da base SBR com doze meses e $\hat{y}(t)$ estimado . . . . .	65
7.6	Sinistros da base SBR com treze meses e $\hat{y}(t)$ estimado . . . . .	66

7.7	Sinistros da base SBR com quatorze meses e $\hat{y}(t)$ estimado . . . . .	67
7.8	Sinistros da base SBR com quinze meses e $\hat{y}(t)$ estimado . . . . .	68
7.9	Sinistros da base SBR com dezesseis meses e $\hat{y}(t)$ estimado . . . . .	69
7.10	Primeira diagonal estimada para a base SBR com dezesseis meses . . . . .	70
7.11	Valores da primeira diagonal estimados para os dados SBR - $R\$/Mil$ . . .	71
7.12	Erros percentuais da primeira diagonal estimada para os dados SBR . . . .	72
7.13	Valores totais estimados para os dados SBR - $R\$/Mil$ . . . . .	73

# Lista de Tabelas

3.1	Triângulo de <i>Run-off</i> Incremental . . . . .	23
3.2	Triângulo de <i>Run-off</i> acumulado . . . . .	24
3.3	Triângulo de <i>Run-off</i> Diagonal . . . . .	25
6.1	Dados DJZ . . . . .	50
6.2	Reserva de IBNR estimada pelo método <i>Chain Ladder</i> para os dados DJZ	51
6.3	Parâmetros $\sigma_t^2$ e $\eta_t^2$ para DJZ . . . . .	54
6.4	Dados DJZ em Milhares de Libras calculado por SST . . . . .	55
6.5	Resultados dos dados DJZ . . . . .	56
7.1	Dados SBR - Milhares de Reais . . . . .	59
7.2	Resultados dos parâmetros $\sigma_t^2$ e $\eta_t^2$ calculados pelo Filtro de Kalman sem critério de parada . . . . .	61
7.3	Resultados dos parâmetros $\sigma_t^2$ e $\eta_t^2$ calculados pelo Filtro de Kalman com critério de parada . . . . .	62
7.4	Resultados dos parâmetros $\sigma_t^2$ e $\eta_t^2$ calculados pelo Filtro de Kalman otimizado	62
7.5	Resultados dos dados SBR . . . . .	69
7.6	Valores da primeira diagonal estimados para os dados SBR - <i>R\$/Mil</i> . . .	72
7.7	Erros percentuais da primeira diagonal estimada para os dados SBR . . . .	72
7.8	Valores totais estimados para os dados SBR - <i>R\$/Mil</i> . . . . .	73
A.1	Dados SBR - Reserva estimada pelo método <i>Chain Ladder</i> - <i>R\$/Mil</i> . . .	81
B.1	Dados SBR - Reserva estimada pelo modelo de Espaço de Estados e Filtro de Kalman <i>R\$/Mil</i> . . . . .	83

# Capítulo 1

## Introdução

### 1.1 Objetivo

Este trabalho tem como objetivo calcular a reserva de provisão dos sinistros de seguros, ocorridos mas não avisados (IBNR<sup>1</sup>), de uma grande seguradora brasileira para um ramo Não Vida. A base de dados que contém estas informações será referenciada como "**dados SBR**".

Para o cálculo, será utilizado um modelo de Espaço de Estados e Filtro de Kalman, testando desta forma, a eficiência do modelo com dados não necessariamente bem comportados. Esta mesma provisão será calculada também pelo método *Chain Ladder*<sup>2</sup>, para efeitos de comparação. O método *Chain Ladder* foi escolhido para contrapor o modelo de Espaço de Estados e Filtro de Kalman por ser o mais conhecido e utilizado no cálculo da provisão de IBNR pelas empresas seguradoras, devido à sua simplicidade e por ser livre de distribuição estatística.

### 1.2 Motivação

O mercado segurador brasileiro tem crescido 13% em média em termos de prêmios emitidos (incluindo previdência e capitalização), desde a crise de 2008. Com perspectivas de manter seu crescimento nos próximos anos, de acordo com os dados publicados periodicamente pela Fenaseg<sup>3</sup> [17]. Tendo uma participação de cerca de 5% do PIB brasileiro atualmente.

Dessa forma, novas seguradoras têm entrado no mercado brasileiro, outras tantas vêm se fundindo com seguradoras maiores e criando novas empresas. Estes movimentos do mercado atraem grande atenção tanto por parte dos segurados quanto dos acionistas na hora de escolher em qual seguradora devem confiar, operar ou, ainda, associar-se.

---

<sup>1</sup>Sigla em inglês *Incurred But Not Reported*.

<sup>2</sup>O método *Chain Ladder* será explicado no tópico 3.3.

<sup>3</sup>Federação Nacional das Empresas de Seguros Privados e de Capitalização



A maneira como uma seguradora provisiona seu capital define de forma crucial o seu relacionamento com os agentes do mercado. Para um segurado, um capital sub-provisionado pode significar que seu sinistro, que por ventura venha ocorrer, não seja coberto. No entanto, para o acionista, um capital super-provisionado pode significar que o retorno de seu investimento será reduzido ou até mesmo consumido pelo aporte do capital.

No Brasil, o órgão responsável pela solvência das empresas seguradoras é a SUSEP<sup>4</sup>. Através dela foram publicadas diversas portarias que regulamentam a forma de provisão do capital das seguradoras brasileiras. Dentre as que se destacam estão a CNSP nº 8 de 1989 [34] e a CNSP nº 18 de 1998 [35].

Apartir de 25 de agosto de 1998 tornou-se obrigatório no Brasil, através da resolução CNSP nº 18, a constituição da provisão dos sinistros ocorridos e não avisados ou, ainda, IBNR. A SUSEP estipulou que esta provisão deverá ser estimada atuarialmente em função do montante esperado de sinistros ocorridos em riscos assumidos na carteira e não avisados até a data-base das demonstrações financeiras.

Cada Sociedade Seguradora foi autorizada a utilizar o método que considere mais adequado para o cálculo do montante desta provisão, desde que informe à SUSEP a metodologia do cálculo através de nota atuarial.

Determinou, ainda, que a estimativa deverá ser baseada na informação sobre a sinistralidade de períodos anteriores completos, de no mínimo um ano. A periodicidade do cálculo é mensal e o atuário responsável poderá ser punido no caso de não cumprimento das regras.

O que torna interessante buscar novas formas de cálculo é a possibilidade de um melhor provisionamento, sendo mais assertivo, tanto para garantir a solvência das instituições quanto para não onerar excessivamente estas. Garantindo assim, um melhor fluxo de caixa, evitando prejuízo por parte dos acionistas.

### 1.3 Estrutura do trabalho

Este trabalho partiu da ideia de compor um estudo que utilizasse as técnicas aplicadas para o cálculo de derivativos financeiros com provisões de uma companhia seguradora (assunto de interesse do autor). Nesta busca foi encontrado o artigo clássico de Piet de Jong & B. Zehnwirth [12], do qual partiu a motivação do trabalho.

Para simplificar a leitura deste trabalho, prevalecerá a notação deste artigo, precursor do uso do Filtro de Kalman em estimativas atuariais, apesar de serem utilizadas notações de outras fontes que serão descritas durante o texto.

---

<sup>4</sup>Superintendência de Seguros Privados

De modo a facilitar o aprendizado dos conceitos principais necessários para o desenvolvimento deste trabalho, o mesmo foi dividido em três partes distintas:

- 1 Discussão sobre seguros e a provisão de sinistros do tipo IBNR;
- 2 As técnicas que envolvem o modelo de Espaço de Estados e o Filtro de Kalman;
- 3 Estimação da provisão de sinistros IBNR para uma seguradora brasileira.

A primeira parte começa com o item 2.1 que apresenta uma breve história sobre o desenvolvimento do cálculo de IBNR. Em seguida, o capítulo 3 mostra as formas de organizar a informação dos sinistros ocorridos utilizados no cálculo ou, ainda, o Triângulo de *Run-off*. E por fim, o mesmo capítulo é encerrado com a explicação do método *Chain Ladder*.

A segunda parte que trata da teoria propriamente dita, inicia-se no capítulo 4 que traz os conceitos mais básicos envolvidos no algoritmo do Filtro de Kalman, bem como o modelo de Espaço de Estados e avança até o capítulo 5 dedicado ao Filtro de Kalman.

Por fim, a terceira parte concentra o capítulo 6 que contextualiza o modelo de Espaço de Estados e o Filtro de Kalman para o cálculo da provisão de IBNR utilizado no trabalho e o capítulo 7 que demonstra os resultados obtidos com os dados reais de uma seguradora brasileira.

# Capítulo 2

## Revisão Bibliográfica

### 2.1 IBNR

Os sinistros de seguros ocorridos, mas não avisados (IBNR) são parte importante da perda esperada nas operações de seguros. A provisão de IBNR compõe o valor mensal contabilizado do sinistro de seguro apurado pela empresa seguradora.

Talvez o mais aceito como o precursor no assunto sobre a importância do cálculo das reservas de IBNR foi Thomas F. Tarbel. Em seu artigo de 1934 [36], ele conceituou formalmente o IBNR como a forma correta de provisionar o valor dos sinistros ocorridos, mas não avisados de uma seguradora.

Ainda em seu artigo, Tarbel afirma que o problema da reserva de ocorrências não avisadas é essencialmente atuarial ou estatístico. Antes de Tarbel, o valor provisionado de IBNR era feito com base nos prêmios emitidos, assim como é feito hoje no Brasil para a margem de solvência, conforme a resolução CNSP nº 8 de 1989 [34].

Vale a reflexão de que Tarbel já compreendia, em 1934, a complexidade do problema de composição destas reservas e propôs uma solução matemática para o cálculo.

Em um momento em que não se pensava em computadores, ele salientou o fato de que a problemática é de natureza estatística, o que torna os cálculos complexos e algumas vezes, até mesmo impraticáveis quando feitos à mão.

Quase quarenta anos depois do artigo de Tarbel, em 1972, Robert L. Bornhuetter e Ronald E. Ferguson publicaram seu famoso artigo [6] no qual descrevem sua técnica, uma das mais utilizadas atualmente [32]. Nele, os autores também enfatizam, com mais cuidado, a importância da correta estimação de IBNR. Seu método hoje é conhecido como "Método BF". Pouco tempo depois, Finger [18] trouxe uma formalização ainda maior ao Triângulo de *Run-off*<sup>1</sup>.

Trinta anos após a publicação do Método BF, Mack [25] formalizou o método *Chain Ladder* e mais tarde [26] sugeriu a técnica para calcular o erro associado ao método no

---

<sup>1</sup>O Triângulo de *Run-off* será explicado no capítulo 3.

cálculo das observações do Triângulo de *Run-off*<sup>2</sup>. Antes de Mack, todas as tentativas de calcular tais "erros" acabavam por criar novos modelos que não o *Chain Ladder*, pois baseavam-se em alguma distribuição estatística, acabando por estimar valores que não coincidiam com os originais [3]. Um exemplo é o caso Renshaw e Verral [30] que discutiram sobre uma distribuição que fundamenta o método *Chain Ladder*, para tanto, utilizaram uma distribuição de *Poisson*.

Mais recentemente, iniciou-se a linha de pesquisa que realiza a estimativa para os sinistros do tipo IBNR e IBNER separadamente, como no trabalho de Liu & Verrall [24]. O IBNER, sigla do inglês *Incurred But Not Enough Reported*, representa as reservas referentes a sinistros IBNR que foram avisadas e mesmo após o aviso do sinistro, o valor contabilizado deste é alterado. A linha de estudo que separa as duas componentes, vem ganhando força, mesmo porque ao separá-las a precisão das estimativas encontradas aumenta. Neste trabalho, estas componentes não serão separadas.

## 2.2 Filtro de Kalman

O Filtro de Kalman<sup>3</sup>, como sugere Bianco [5], é um dos desenvolvimentos mais importantes da ciência aplicada no último século com forte destaque para as aplicações em engenharia. No entanto, as aplicações em finanças vêm ganhando cada vez mais força atualmente.

Esse destaque em finanças leva a crer que o Filtro de Kalman surgiu nos últimos anos. Ao contrário, a filtragem de sinais é um assunto que possui quase um século de pesquisas na área da Engenharia Elétrica. De acordo com Oliveira [28], após o descobrimento da eletricidade ao final do século XIX por Thomas Edison, iniciou-se a "Era da Eletrônica". Esta permeou-se de diversas descobertas e com as quais surgiu uma nova necessidade de tratar os sinais elétricos e de radiofrequência. Tais eventos culminaram, em fevereiro de 1942, na publicação de um livro o qual teve pouca circulação, segundo [23], do trabalho de Wiener que mais tarde seria conhecido como o Filtro de Wiener.

Foi em 1949 que Wiener publicou seu livro [37] que incluía seu trabalho pioneiro. Nos anos seguintes, seu trabalho foi expandido como em Zadeh & Ragazzini [38] que generalizaram o trabalho de Wiener nos problemas de filtragem.

Estas pesquisas culminaram, em 1960, no surgimento do trabalho de Rudolf E. Kalman [21] expandido em 1961 por este em parceria com Richard S. Bucy que contribuiu com a teoria, levando o novo trabalho a ser conhecido como o Filtro de Kalman-Bucy [22].

Stanley Schmidt, geralmente creditado como o primeiro a implementar o Filtro de Kalman, uniu-se ao centro de pesquisa AMES da *National Aeronautics and Space Administration* (NASA) em 1946, no qual passou a trabalhar em instrumentação, computação

<sup>2</sup>O Triângulo de *Run-off* será detalhado no item 3.2.

<sup>3</sup>O Filtro de Kalman será mais detalhado no item 2.3.

analógica e teoria de perturbação linear.

Mais tarde, quando a NASA estava explorando o problema de navegar até a lua no programa Apollo, Schmidt viu o potencial de estender o Filtro de Kalman linear para resolver o problema da estimativa da trajetória.

O resultado foi chamado de Filtro de Kalman-Schmidt [33] (agora conhecido como o Filtro de Kalman Estendido). Em 1961, Schmidt e John White haviam demonstrado que esse filtro, combinado com medições ópticas das estrelas e dos dados sobre o movimento da nave espacial, poderia fornecer a precisão necessária para uma inserção bem sucedida na órbita ao redor da lua. O Filtro de Kalman-Schmidt foi incorporado ao computador de navegação da cápsula Apollo e, finalmente, em todos os sistemas de navegação aérea [14].

## 2.3 O modelo de Espaço de Estados no cálculo de IBNR

Pouco mais de vinte anos após a publicação do Filtro de Kalman, em 1983, De Jong e Zehnwirth publicaram seu trabalho *Claims reserving state space models and the Kalman filter* [12]. Nele foi apresentado um método detalhado de como calcular as reservas de IBNR utilizando o Triângulo de *Run-off*, ordenado pela diagonal<sup>4</sup>, para sugerir um equacionamento do modelo de Espaço de Estados e daí utilizar o Filtro de Kalman na estimação. Apesar de utilizar conceitos ainda pouco explorados na área de Finanças, como o Filtro de Kalman, o método se mostrou muito adequado à aplicação.

Em 2002, England & Verrall em seu artigo *Stochastic Claims Reserving in General Insurance* [16] levantaram a literatura atuarial que utilizava métodos estocásticos para o cálculo de IBNR, debatendo o conceito de "o melhor estimador" nessas técnicas.

O Filtro de Kalman pressupõe encontrar o melhor estimador desde que a distribuição dos erros de observação siga uma distribuição Normal<sup>5</sup>. Essa é sem dúvida a premissa forte no modelo de Espaço de Estados utilizado no Filtro de Kalman, pois atualmente já é notório que distribuições baseadas em séries financeiras, via de regra, não possuem distribuição Normal. Para mais detalhes sobre esta discussão, vale a leitura do artigo de Cont: *Empirical properties of asset returns* [11].

Em 2006, De Jong volta a falar de modelos estocásticos em seu novo artigo [13]. E em 2007 Atherino & Fernandes [2] revisitam o modelo de De Jong & Zenwirth retirando das estimativas os índices de inflação e de volumes, citados pelos autores do artigo original de 1983 [12].

---

<sup>4</sup>A ordenação pela diagonal do Triângulo de *Run-off* será detalhada no item 3.2.4.

<sup>5</sup>A distribuição Normal será explicada no item 4.1.1.

## 2.4 Contribuições deste Trabalho

Este trabalho tem a pretensão de calcular a reserva de provisão dos sinistros de seguros do tipo IBNR utilizando um modelo de Espaço de Estados e Filtro de Kalman, com dados observados de uma grande seguradora brasileira. É notório na bibliografia que trata do assunto, o uso de bases de dados europeias, americanas e da Oceania, mas nunca dados brasileiros. Sendo assim, verificar a eficácia do método de De Jong & Zenwirth em dados nacionais traz uma maior garantia no uso deste, por parte das áreas atuariais brasileiras.

De forma a atingir este objetivo, na parte prática será utilizado o software MATLAB<sup>®</sup>, já na parte teórica será utilizado, principalmente, o artigo de De Jong & Zenwirth [12], além das outras fontes citadas ao longo do texto.

Para verificar a eficácia da estimação, a provisão será calculada também pelo método *Chain Ladder*. Este método é bem conhecido e aceito pelas áreas atuariais por ser simples e livre de distribuição estatística.

Por fim, os valores estimados serão comparados com os valores reais obtidos no período posterior ao utilizado na estimação, testando a assertividade de ambos os métodos.

## Capítulo 3

# Sinistros em Seguros e a estimação de IBNR

As empresas seguradoras incorrem em custos nas suas operações que são exclusivos do mercado segurador, dentre eles, os mais notórios, certamente, são os sinistros de seguros e suas provisões.

### 3.1 Sinistros em Seguros

A palavra "sinistro" é muito utilizada no mercado financeiro. No mercado de crédito, ela tem sentido de calote, ou seja, ocorre quando o cliente deixa de cumprir sua obrigação de pagamento com o seu credor, tornando-se inadimplente. O credor por sua vez, toma ciência do fato no ato da ocorrência deste. Já no mercado segurador, o sinistro ocorre quando existe perda ou dano do bem segurado, não necessariamente ocorrendo "dolo" por parte do cliente. Este por outro lado, ao envolver-se em um sinistro, por razões diversas, nem sempre notifica o mesmo no ato da ocorrência. Pior ainda, pode ocorrer com os sinistros do ramo Vida, em que os beneficiários da apólice muitas vezes desconhecem a mesma ou ainda demoram muito para notificar a seguradora devido ao seu período de luto ou mesmo a demora no processo de inventário do segurado.

### 3.2 O Triângulo de *Run-off*

O triângulo de desenvolvimento de sinistros ou como é mais conhecido, Triângulo de *Run-off*, é a forma mais utilizada para organizar as ocorrências de sinistros. A razão disto é a facilidade de visualizar as informações dos sinistros ocorridos com ambas as datas de ocorrência e de tempo de atraso no aviso deste sinistro em relação ao acontecimento do sinistro em si. Caso fosse utilizado um único índice para representação destes sinistros, a cada novo período de tempo (meses, trimestres, anos, etc) o valor do total dos sinistros

ocorridos em períodos anteriores seria alterado, pois novos valores de sinistros "antigos" seriam informados.

A escolha do Duplo Índice no Triângulo de *Run-off*, portanto, elimina este problema na hora de organizar os sinistros ocorridos, já conhecidos ou pagos. Também facilita no que tange à intuição do que deve ser estimado no cálculo da provisão, já que os dados não preenchidos são os que devem ser estimados.

### 3.2.1 Representação em Duplo Índice

### 3.2.2 Forma Incremental

A forma Incremental do Triângulo de *Run-off* é representada por uma matriz diagonal superior, na qual os elementos desta são **os valores pagos ou ocorridos de sinistros** de uma dada seguradora. Ou seja, a variável  $y$  representa o valor financeiro ou de ocorrências (quantidade) no determinado período de tempo estipulado para medir os valores, sendo estes meses, trimestres, anos, etc.

Tabela 3.1: Triângulo de *Run-off* Incremental

w/d	0	1	...	j	...	d-1	d
1	$y_{1,0}$	$y_{1,1}$	...	$y_{1,j}$	...	$y_{1,d-1}$	$y_{1,d}$
2	$y_{2,0}$	$y_{2,1}$	...	$y_{2,j}$	...	$y_{2,d-1}$	
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\cdot$		
i	$y_{i,0}$	$y_{i,1}$	...	$y_{i,j}$			
$\vdots$	$\vdots$	$\vdots$	$\cdot$				
w-1	$y_{w-1,0}$	$y_{w-1,1}$					
w	$y_{w,0}$						

A tabela 3.1 evidencia a organização na Forma Incremental. Nela, os valores  $y_{i,j}$ , com  $i = 1, \dots, w$  e  $j = 0, \dots, d$ , podem representar o montante total ou a média das indenizações pagas, o valor de sinistros declarados, o valor de sinistros pagos, o montante dos prêmios ou, ainda, a quantidade de apólices.

Dois exemplos da forma Incremental são mostrados neste trabalho, um na tabela 6.1 da página 50 e o outro na tabela 7.1 da página 59.

Cada linha da matriz representa um período de ocorrência de sinistros e as colunas correspondem aos períodos de liquidação ou pagamento das indenizações de sinistros.

As diagonais da matriz são os "períodos calendários", ou seja, períodos em que são feitos os pagamentos.



### 3.2.3 Forma Acumulada

A forma acumulada do Triângulo de *Run-off* é utilizada para o cálculo da provisão de IBNR pelo método *Chain Ladder*, que será detalhado no item 3.3. A diferença entre a forma Incremental e a acumulada consiste em **somar as colunas do triângulo no sentido da passagem do tempo de atraso  $d$** . Desta forma, a última coluna preenchida de cada linha representará o valor pago do sinistro daquele período  $i$ , com  $i$  compreendido entre um e  $w$ , do início deste até o último valor conhecido no instante  $t$  como mostrada na tabela 3.2. Para ilustrar cada termo  $C_{i,j}$  respeitará a equação 3.1 abaixo:

$$C_{i,j} = \sum_{k=0}^j y_{i,k} \quad (3.1)$$

Tabela 3.2: Triângulo de *Run-off* acumulado

w/d	0	1	...	j	...	d-1	d
1	$C_{1,0}$	$C_{1,1}$	...	$C_{1,j}$	...	$C_{1,d-1}$	$C_{1,d}$
2	$C_{2,0}$	$C_{2,1}$	...	$C_{2,j}$	...	$C_{2,d-1}$	
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	.		
i	$C_{i,0}$	$C_{i,1}$	...	$C_{i,j}$			
$\vdots$	$\vdots$	$\vdots$	.				
w-1	$C_{w-1,0}$	$C_{w-1,1}$					
w	$C_{w,0}$						

As letras  $y$  (sinistro no período) da forma incremental e  $C$  (sinistro acumulado pela data de ocorrência) da forma acumulada foram escolhidas de forma a facilitar o estudo de referência como, por exemplo, Mack [25] e Renshaw & Verral [30].

### 3.2.4 Referência Diagonal

Referenciar, ou ainda ordenar, o Triângulo de *Run-off* pela sua diagonal nada mais é que reorganizar os índices de linhas e colunas de forma que a última diagonal preenchida do triângulo corresponda ao vetor  $Y_t$ . Este contendo todos os lançamentos daquele período calendário (meses, anos, etc) independente de qual o seu ano de origem ou de desenvolvimento. Os valores do Triângulo de *Run-off* permanecem os mesmos que no triângulo incremental, o que muda de fato é a forma na qual os índices são organizados. A cada novo período, o vetor  $Y_t$  aumenta sua dimensão, ou seja, no período  $t - 1$  ele terá  $t - 1$  elementos e no período  $t$  terá  $t$  elementos e assim por diante. Nesta abordagem,

para um dado  $y_{i,j}$  haver um  $y_j(t)$  para  $j = 0, 1, \dots, t-2, t-1$ . Sendo que  $t$   agora a representao do tempo de calendrio e no de ocorrncia ou desenvolvimento do sinistro, como no caso das ordenaes por Duplo ndice.

No entanto, o ndice  $j$  remete ao tempo de desenvolvimento do sinistro. Isto ser importante para identificar qual perodo de ocorrncia  $w$  pertence a componente do vetor com os sinistros  $Y_t$  no tempo  $t$ . Em outras palavras, no perodo calendrio  $t$ , para  $j = 0$  resultar o perodo de ocorrncia do sinistro  $w = t$ .

Tabela 3.3: Tringulo de *Run-off* Diagonal

w/d	0	1	...	j	...	t-2	t-1
1	$y_0(1)$	$y_1(2)$	...	$y_j(j+1)$	...	$y_{t-2}(t-1)$	$y_{t-1}(t)$
2	$y_0(2)$	$y_1(3)$	...	$y_j(j+2)$	...	$y_{t-2}(t)$	
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\cdot$		
i	$y_0(i)$	$y_1(i+1)$	...	$y_j(t)$			
$\vdots$	$\vdots$	$\vdots$	$\cdot$				
t-1	$y_0(t-1)$	$y_1(t)$					
t	$y_0(t)$						

Assim  $y_j(t)$   o valor do Tringulo de *Run-off* observado no desenvolvimento  $j$  e ano de calendrio  $t$ , desta forma, o ano de origem do sinistro  encontrado pela relao  $w = t - j$ .

A diagonal utilizada como entrada nas estimaes de IBNR utilizando o modelo de Espaço de Estados  dada pela diagonal em vermelho da tabela 3.3. Sua forma vetorial  mostrada na equao 3.2 abaixo:

$$y(t) = \begin{bmatrix} y_0(t) \\ y_1(t) \\ \vdots \\ y_j(t) \\ \vdots \\ y_{t-1}(t) \end{bmatrix} \quad (3.2)$$

O vetor  $y(t)$  possui dimenso  $t \times 1$  no instante  $t$ . Nota-se que o primeiro valor do vetor  o valor de menor ndice de atraso na tabela 3.3 para o instante  $t$ .

### 3.3 Estimaco pelo mtodo *Chain Ladder*

Segundo Mack [25], o mtodo mais popular para estimaes de reservas de IBNR  o *Chain Ladder*, e a razo principal disso  o fato do mtodo ser simples e livre de

distribuição estatística.

Como explicado no item 3.2.3,  $C_{ij}$  foi definido como o total de reservas acumuladas do período de acidente ou sinistro  $i$ , do início deste até os períodos de liquidação ou pagamento das indenizações de sinistros ou, ainda, a coluna do Triângulo de *Run-off*  $j$ .

Considerando-se as reservas acumuladas  $C_{i,j}$  como variáveis aleatórias, estas possuem valores observáveis para  $1 \leq i \leq t$  e  $0 \leq j \leq t - i$ , com  $t > 0$ . Para os outros casos, os valores das reservas devem ser estimados por algum método que complete as lacunas do Triângulo de *Run-off*. O método *Chain Ladder* consiste em calcular tais sinistros com base nos valores já observados de acordo com a seguinte regra de estimação mostrada a seguir:

$$E[C_{i,j+1}|C_{i,0}, \dots, C_{i,j}] = \hat{C}_{i,j+1} = \hat{C}_{i,j} \cdot \hat{f}_{i,j} \quad (3.3)$$

Aqui, o fator  $\hat{f}_{i,j}$  definido para cada período  $i$  é dado por:

$$\hat{f}_{i,j} = \frac{\sum_{i=1}^{t-j-1} C_{i,j+1}}{\sum_{i=1}^{t-j-1} C_{i,j}} \quad t - i \leq j \leq t - 1 \quad (3.4)$$

Sendo que o último termo a ser calculado no Triângulo de *Run-off* será dado por:

$$\hat{C}_{i,t-1} = C_{i,t-i} \cdot \prod_{j=0}^{t-2} \hat{f}_{i,j} \quad 2 \leq i \leq t \quad (3.5)$$

A razão de se utilizar  $\hat{f}_{i,j}$  e não simplesmente  $f_{i,j}$  é o fato de ser possível calcular um desvio padrão para o método *Chain Ladder*, segundo Mack [26].

Lembrando-se do fato de que  $C_{i,j}$  é dada pela equação 3.1, a equação 3.3 terá estimado, então, o valor acumulado para o Triângulo de *Run-off*. Como próximo passo, deverão ser obtidos os respectivos  $y_{i,j}$  da forma Incremental do Triângulo de *Run-off*.

Uma forma de realizar tal cálculo para se obter os sinistros que completam o Triângulo de *Run-off* Incremental é dado pela equação 3.6 abaixo:

$$y_{i,j} = C_{i,j} - C_{i,j-1} \quad 1 \leq i \leq t, \quad 1 \leq j \leq t - 1, \quad t > 0 \quad (3.6)$$

De forma prática, o método *Chain Ladder* nada mais é que uma razão entre as colunas do Triângulo de *Run-off* que devolve uma proporção de decaimento dos sinistros apurados ao longo do tempo de atraso  $d$ .

# Capítulo 4

## Conceitos básicos

### 4.1 Considerações iniciais

Antes de iniciar o processo de estimação propriamente dito, é interessante rever alguns conceitos que serão amplamente comentados ao longo do trabalho.

#### 4.1.1 Distribuição Normal

Neste ponto do trabalho, é importante uma breve discussão sobre a distribuição Normal. Se uma variável aleatória segue esta distribuição escreve-se:  $X \sim N(\mu, \sigma)$ .

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}, \quad -\infty < x < \infty, \sigma > 0 \quad (4.1)$$

Na equação acima  $x$  é a variável aleatória, ou seja, é a variável de interesse a ser estudada ou estimada, já  $\mu$  é a média e  $\sigma^2$  é a variância. Estes dois últimos são os parâmetros da distribuição. E conhecendo-os é possível simular quaisquer dados que sigam uma distribuição Normal.

Quando  $\mu = 0$  e  $\sigma^2 = 1$ , a distribuição é dita ser uma "Normal Padrão". A Normal Padrão possui um equacionamento mais simples que o da Normal, por se tratar de um caso específico, e é dada a seguir:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x^2}{2}\right)}, \quad -\infty < x < \infty, \quad (4.2)$$

A distribuição Normal é a mais conhecida e utilizada na Estatística. Segundo Ross [31], sua popularidade se deu com a lei dos "grandes números". Em linhas gerais, esta afirma que a soma de um grande número de variáveis aleatórias independentes e identicamente distribuídas ou *i.i.d.* têm uma distribuição que é aproximadamente uma Normal.

Curva de Gaus e "curva em forma de sino" são outros nomes dados à distribuição Normal. Gaus foi um dos pioneiros no uso desta distribuição, popularizando-a em seus trabalhos na Engenharia Elétrica e por isso esta ligação ao seu nome.

### 4.1.2 Ruído Branco

O Ruído Branco (RB) é descrito como uma série temporal  $\epsilon_1, \dots, \epsilon_t$ . Ele recebe este nome devido à luz branca. Como explicado em Haykin [20], essa ligação se faz por conta de que a luz branca possui intensidades iguais de todas as frequências dentro da banda visível de radiação eletromagnética.

Para ser classificado como Ruído Branco é necessário que a série temporal possua as seguintes propriedades:

- 1  $E(\epsilon_t) = E(\epsilon_{t-1}) = 0, \quad \forall t = 1, 2, \dots, \infty ;$
- 2  $Var(\epsilon_t) = Var(\epsilon_{t-1}) = \sigma^2, 0 < \sigma^2 < \infty ;$
- 3  $Cov(\epsilon_t, \epsilon_{t-1}) = 0 .$

A propriedade um equivale a dizer que o primeiro momento ou a esperança matemática de cada um dos termos da série do Ruído Branco é zero. A segunda propriedade afirma que a variância da série é constante e dada por  $\sigma^2$ . A terceira e última propriedade equivale a dizer que todos os elementos da distribuição são independentes entre si um a um, ou seja, são não - correlacionados.

Nota-se que em decorrência das propriedades, tem-se que o Ruído Branco é estacionário de segunda ordem ou ainda possui estacionariedade fraca. A estacionariedade é uma característica importante em séries temporais, pois pode-se estimar os momentos da série, caso ela ocorra. Uma vez que os momentos são estimados, fica fácil obter previsões das informações.

Em Estatística, um Ruído Branco é um conceito econométrico muito presente no estudo das séries temporais, especialmente as estocásticas discretas. Sua representação é RB.

Quando o Ruído Branco segue uma distribuição Normal, ou seja, a sua função de distribuição é dada pela equação 4.1, este é chamado de Ruído Branco Gaussiano (RBG). O RBG é importante, pois o tratamento deste nas equações torna-se simples, uma vez que obedece às propriedades de uma Normal.

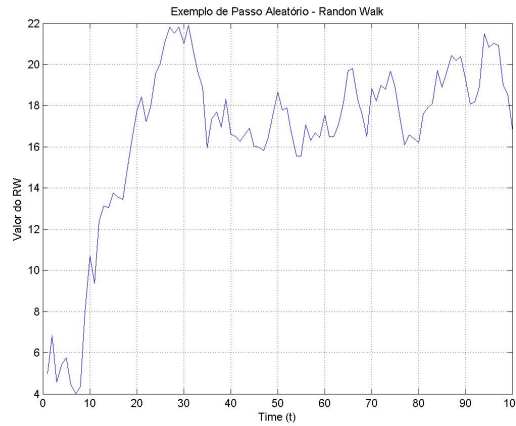
### 4.1.3 Passeio Aleatório

O Passo ou Passeio Aleatório, também conhecido como *Random Walk*, é o nome dado a uma série temporal que descreve uma tendência estocástica pura. Sua equação é descrita por um  $\beta_t$  definido pelo termo anterior  $\beta_{t-1}$  da série mais um Ruído Branco Gaussiano  $\epsilon_t$ , como descrito no item 4.1.2. Um exemplo de Passeio Aleatório é dado pela equação 4.3 abaixo:

$$\beta_t = \beta_{t-1} + \epsilon_t \quad (4.3)$$

Uma simulação de um Passeio Aleatório é mostrado na figura 4.1. Nesta simulação foi utilizado um  $\beta_0 = 5$  e um passo de tempo  $T = 100$ .

Figura 4.1: Exemplo de Passeio Aleatório - *Randon Walk*



Para o termo aleatório  $\epsilon_t$  foi utilizado um Ruído Branco que segue uma distribuição Normal Padrão, ou seja, possui variância unitária. Em outras palavras, foi utilizado um um RBG com  $\sigma^2 = 1$ .

## 4.2 Estimação por Mínimos Quadrados

A estimação por Mínimos Quadrados é a mais utilizada para obter estimadores, segundo Moretin & Bussab [9]. Ela consiste em um método de estimação *off-line*. Colocado de outra forma, necessita-se possuir, *a priori*, a maior quantidade possível de dados para se obter o estimador ótimo amostral.

### 4.2.1 Estimador de Mínimos Quadrados Ordinários

O estimador de Mínimos Quadrados Ordinários (MQO) ou OLS do inglês *Ordinary Least Squares*, é o estimador  $\beta$  encontrado ao se utilizar o método de estimação de MQO que consiste em minimizar a soma dos quadrados dos resíduos entre o valor estimado e os dados observados. Nesta abordagem, a variável que se pretende estimar é chamada de Variável Dependente e as variáveis utilizadas na estimação são chamadas de Variáveis Explicativas.

Em um modelo univariado, ou seja, só existe uma Variável Explicativa, o que se pretende estimar é a curva  $\hat{y}_t$  (Variável Dependente) que melhor descreve os dados da população de interesse. Esta curva pode ser descrita pela equação 4.4.

$$y_t = b_0 + b_1 X_t + e_t \quad (4.4)$$

Sendo que  $e_t$  é o vetor de erros de estimação. O método de MQO exige certas premissas sobre  $e_t$ . Este deve possuir média zero e variância fixa. Além disso, os termos devem ser independentes um a um ou, ainda, devem fazer parte de uma distribuição de termos independentes e identicamente distribuídos (*i.i.d.*). A notação é dada por:  $e_t \sim i.i.d.(0, \sigma)$ .

O estimador de Mínimos Quadrados Ordinários dado por:

$$\beta = (X_t' X_t)^{-1} X_t' y_t \quad (4.5)$$

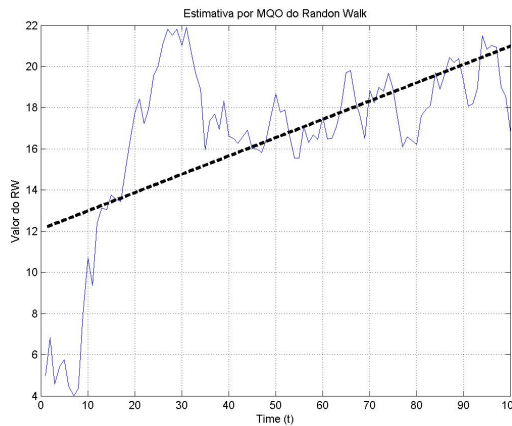
Sendo  $\beta$  um vetor de parâmetros composto por  $b_0$  e  $b_1$  e  $X_t$  uma matriz composta do vetor  $x_1, \dots, x_t$  e um vetor de "uns" de dimensão  $t \times 1$ .

Para o caso da regressão univariada, a representação matricial da equação 4.5 é mostrada na equação 4.6 abaixo:

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \left[ \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_t \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_t \end{bmatrix} \right]^{-1} \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_t \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_t \end{bmatrix} \quad (4.6)$$

Na figura 4.2 é mostrada uma estimação por MQO do Passo Aleatório exposto na figura 4.1, tomando-se o tempo  $t$  como Variável Explicativa.

Figura 4.2: Estimativa por MQO do Passeio Aleatório



Fica evidente na figura 4.2 que tomar o tempo como Variável Explicativa na estimação de MQO não é a melhor escolha para se estimar a função estocástica de um Passeio Aleatório. De fato, no entanto, a estimação por MQO é uma etapa para tal estimação, que consiste na estimação de modelos ARMA( $p, q$ ) em que  $p$  e  $q$  são os parâmetros deste modelo. Para mais detalhes sobre os modelos ARMA( $p, q$ ), consulte o trabalho de Bueno [8, pag. 114].

Mais informações sobre o método MQO e suas propriedades, bem como as suas aplicações, podem ser encontradas no livro de Mendenhall [27, cap 11].

### 4.2.2 Estimador de Mínimos Quadrados Recursivos

A estimação por MQO, como foi mostrado, apresenta o problema comum às estimações *off-line*. É necessário obter todos os dados possíveis para obter o melhor estimador. Acontece que na estimação de séries temporais, o maior problema é a quantidade de dados disponíveis. Até porque o objetivo na estimação de séries temporais é encontrar o valor futuro desta no instante seguinte, portanto, nunca será possível obter todos os dados da mesma.

Quando a série é não estacionária, a tarefa de estimá-la torna-se árdua e não trivial, pois é impossível estimar todos os momentos desta, como afirma [8].

Já na série estacionária, mesmo que fraca, sabe-se que ao menos o primeiro momento desta é constante e, portanto, pode ser estimado.

A mesma formulação utilizada para encontrar o estimador de MQO é aplicada para encontrar o estimador de Mínimos Quadrados Recursivos (MQR) só que neste caso, de maneira recursiva. Abaixo é mostrada a equação que descreve a série no instante  $t$ .

$$y_t = \beta_t x_t + e_t \quad (4.7)$$

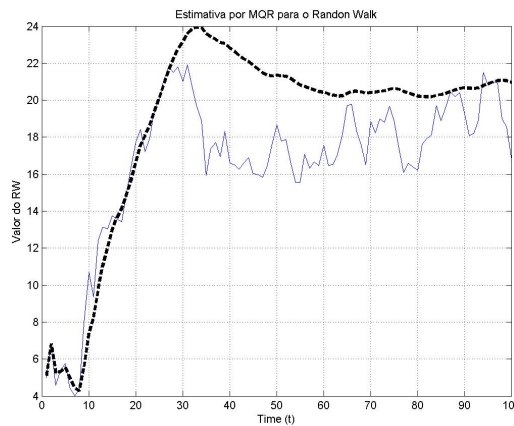
Agora,  $e_t$  é o erro de estimação no passo  $t$ . Deseja-se saber portanto, qual o valor seguinte  $\beta_{t+1}$  de uma série no instante  $t + 1$  que melhor descreva a função.

$$y_{t+1} = \beta_{t+1} x_{t+1} + e_{t+1} \quad (4.8)$$

Para resolver o problema, pode-se utilizar a equação 4.5 recursivamente, de forma que para cada novo  $x_t$ , uma nova estimativa de  $\beta_t$  seja refeita.

Na estimação por Mínimos Quadrados Recursivos (MQR), este novo dado é adicionado à estimação, com o intuito de corrigir o  $\beta$  encontrado.

Figura 4.3: Estimativa por MQR para o Passeio Aleatório



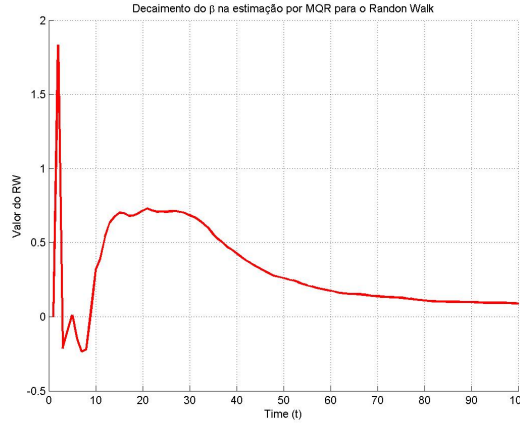
A figura 4.3 refere-se a estimação por MQR para o Passeio Aleatório. A estimação,



diferentemente da *off-line*, acompanha a evolução da série, agindo como um atenuador ou "esfumaçador".

O gráfico do decaimento de  $\beta_t$  é dado a seguir:

Figura 4.4: O valor de  $\beta_t$  na estimativa por MQR



Quanto maior a quantidade de informação, mais perto dos parâmetros populacionais estará a amostra e assim, o  $\beta_t$  tende para o estimador não - viesado. No caso da figura 4.4,  $\beta_t$  converge para os valores encontrados no método de MQO explicado no item 4.2.1.

### 4.3 Modelos de Espaço de Estado

Os modelos de Espaço de Estados são largamente empregados em Teoria de Sistemas, na Física e Engenharia e seu nome deriva destes campos de conhecimento. A ideia geral utilizada por trás destes é a de que uma série temporal observável  $y_1, \dots, y_t$  depende de possíveis "estados"  $\beta_1, \dots, \beta_t$ , não observáveis do sistema. Os quais são definidos por um processo estocástico.

De acordo com Harvey [19], um processo estocástico é aquele cujo estado atual  $y_t$  depende de um estado anterior mais um distúrbio  $\eta_t$  causado por um Ruído Branco e pode ser descrito como:

$$\beta_t = h_t \beta_{t-1} + \eta_t, \quad t = 1, \dots, T \quad (4.9)$$

Na equação 4.9,  $h_t$  é apenas um parâmetro, o valor que ela toma determina o comportamento futuro das observações. Se  $h_t < 1$ , o processo ao longo do tempo retornaria para uma média em torno de zero e se tornaria o próprio RB que o define, portanto estacionário, não restando nada a prever. Quando  $h_t > 1$  por outro lado, a influência dos valores passados aumenta ao longo do tempo fazendo com que a função seja divergente

e praticamente imprevisível. Sendo assim, o único caso de interesse é quando  $h_t = 1$ , ou seja, a equação 4.9 torna-se um Passeio Aleatório como descrito no item 4.1.2.

#### 4.3.1 Modelos de Espaço de Estados Lineares Gaussianos.

Os modelos de Espaço de Estados Lineares Gaussianos são um caso específico dos modelos de Espaço de Estados Lineares, como explicado por Durbin & Koopman [15].

Eles consistem em duas equações que são mostrados abaixo:

$$\begin{aligned} y_t &= Z_t \alpha_t + \epsilon_t, & \epsilon_t &\sim N(0, H_t); \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t, & \eta_t &\sim N(0, Q_t), & t = 1, \dots, n. \end{aligned} \quad (4.10)$$

Sendo  $\epsilon_t$  e  $\eta_t$  considerados Ruídos Brancos Gaussianos como descrito no item 4.1.2 da página 28, o que torna o modelo baseado em uma distribuição Normal ou Gaussiana. O termo linear deste, refere-se a relação linear dos parâmetros da equação 4.10.

As matrizes  $Z_t$ ,  $T_t$  e  $R_t$  são supostamente conhecidas e determinísticas e descrevem o comportamento do sistema definido pelo modelo de Espaço de Estados.

Por fim, as matrizes  $H_t$  e  $Q_t$  são respectivamente as matrizes de covariância dos distúrbios  $\epsilon_t$  e  $\eta_t$  e também devem ser especificadas.

As equações descritas em 4.10 utilizam uma notação encontrada em Durbin & Koopman [15]. No entanto, como comentado em 1.3, neste trabalho será utilizada uma outra notação o mais próximo possível da encontrada em De Jong & Zenwirth [12]. As novas equações 4.12 e 4.12 utilizando esta notação serão mostradas na página 33.

$$y_t = X_t \beta_t + \epsilon_t, \quad \epsilon_t \sim N(0, U_t); \quad (4.11)$$

$$\beta_t = H_t \beta_{t-1} + G_t \eta_t, \quad \eta_t \sim N(0, V_t). \quad (4.12)$$

A equação 4.11 é conhecida como a equação do sistema ou a equação das observações. Em fenômenos físicos, ela poderia representar a medida de um sensor de velocidade ou de temperatura. Em [12],  $y_t$  representa o vetor de sinistros já ocorridos e avisados no período  $t$ .

A matriz  $X_t$  representa uma matriz de parâmetros e é determinística *i.e.*, depende de uma função definida (no caso depende do tempo). O termo  $\epsilon_t$  representa a perturbação do sensor ou, o "ruído de medição" do sistema. Em [12] o termo  $\epsilon_t$  é representado por  $u_t$ , no entanto, aqui  $\epsilon_t$  será utilizada para não causar confusões na notação do Filtro de Kalman. Este distúrbio deve ser *i.i.d.* com média zero e matriz de covariância  $U_t$ .

Os estados  $\beta_t$  do sistema são definidos pela equação 4.12 e representam o processo estocástico do problema. Este não é observado diretamente e deve ser obtido implicitamente através do Filtro de Kalman.

A perturbação do estado ou o ruído estocástico é representado por  $\eta_t$ . Do mesmo modo como  $u_t$  foi substituído por  $\epsilon_t$ , optou-se por substituir  $v_t$  por  $\eta_t$ . O uso de  $v_t$  poderia causar confusões de notação com o vetor de resíduos do Filtro de Kalman, como será explicado no capítulo 5. O distúrbio estocástico  $\eta_t$  também deve ser *i.i.d.* de média zero e matriz de covariância  $V_t$  a ser definida.

Assim como  $X_t$ , a matrix  $H_t$  também representa uma matriz de parâmetros, tanto  $H_t$  quanto  $G_t$  são determinísticas. As matrizes  $X_t$ ,  $H_t$ ,  $G_t$  devem ser conhecidas previamente ou então especificadas antes de utilizar o Filtro de Kalman.

A última premissa para o uso do Filtro de Kalman é a de que os distúrbios ou ruídos  $\epsilon_t$  e  $\eta_t$  são presumidamente não - correlacionados entre si e não - correlacionados termo a termo, *i.e.* são *i.i.d.*. Devem ainda possuir média zero e suas variâncias devem ser conhecidas ou especificadas.

# Capítulo 5

## O Filtro de Kalman Discreto

Nesta seção será descrito o método, bem como serão expostos e discutidos os pontos relevantes sobre o Filtro de Kalman.

O Filtro de Kalman geralmente é descrito como um cálculo recursivo ótimo do algoritmo de Mínimos Quadrados Recursivos, como o visto na seção 4.2.2.

Porém, ele é mais do que isto. Pois, ele pode ser considerado como um estimador recursivo. Isto significa que apenas a estimativa do estado no passo anterior e a medição atual são necessários para computar a estimativa do estado atual. Ao contrário de outras técnicas de estimação, nem o histórico das observações nem o histórico das estimativas são necessários.

O Filtro de Kalman é empregado com sucesso em sistemas dinâmicos bem definidos. Em fenômenos físicos como situações em que a trajetória e velocidade de um corpo devem ser determinadas por regras bem definidas como as equações do movimento, por exemplo. Além disso, as entradas de controle devem ser bem especificadas por serem baseadas nos parâmetros do sensor que mede a velocidade deste corpo.

Imaginando que o sensor mede a informação real corrompida por um Ruído Branco e esta medida é relacionada com a anterior, pois trata-se da velocidade instantânea do corpo. O Filtro de Kalman irá funcionar no sentido de descartar as medidas muito dispersas das estipuladas pelo modelo físico, por não fazerem sentido.

No entanto, mesmo descartando o resultado disperso do sensor, a informação não é totalmente descartada. O ganho de Kalman captura a dispersão e a reutilizada comparando-a com a nova medida identificando mudanças de tendência, caso existam. A estimativa obtida desta forma é melhor que a estimativa obtida utilizando-se qualquer resultado obtido pelo sensor unicamente.

## 5.1 Funcionamento do Filtro de Kalman

Além do algoritmo de MQR, o filtro corrige tanto a estimativa de  $\beta_{t|t-1}$  quanto sua matriz de covariância  $C_{t|t-1}$  obtidas *a priori* com o valor medido da variável  $y_t$ . A covariância do erro  $v_t$  dado pela estimativa  $\hat{y}_t$  e o valor medido  $y_t$  é então utilizada para se obter o Ganho de Kalman  $K_t$  que por sua vez multiplica o mesmo erro  $v_t$  para se obter a nova estimativa de  $\beta_{t+1}$ .

Em outras palavras, o Filtro de Kalman é, essencialmente, um conjunto de equações matemáticas que implementam um estimador de  $\hat{\beta}_{t+1|t}$  que prediz  $\hat{y}_{t+1|t}$  e a cada nova interação corrige  $\hat{y}_{t+1|t}$  com  $y_{t+1|t+1}$ . Fornecendo assim, a solução do Modelo de Espaço de Estados Linear Gaussiano colocado em 4.3.1.

No entanto, o Filtro de Kalman não realiza o trabalho todo sozinho. O Modelo de Espaço de Estados deve estar bem especificado, bem como todas as premissas adotadas para garantir o funcionamento correto deste.

Dentre as premissas mais fortes adotadas estão a de que o estado atual é linearmente dependente do estado anterior e a de que os ruídos, tanto da equação das observações quanto da equação de estados, possuem distribuições Normais ou Gaussianas como discutido no item 4.1.1.

As premissas não param por aí. O Filtro não fornece as matrizes de parâmetros  $X_t$ ,  $H_t$ ,  $G_t$  colocadas no final do item 4.3.1. Estas devem ser fornecidas ou especificadas por quem estiver construindo o modelo.

Considerando que tudo foi especificado corretamente, daí sim, o Filtro de Kalman fornece uma solução considerada ótima, no sentido em que minimiza o erro de estimativa da covariância quando o Espaço de Estados Linear Gaussiano for seguido conforme as especificações.

O estado do filtro é representado por duas variáveis:  $\hat{\beta}_{t|t}$ , a estimativa *a posteriori* do estado no tempo  $t$ , dadas as observações até o tempo  $t$ , inclusive;  $C_{t|t}$ , a matriz de covariância do erro *a posteriori* (uma medida da acurácia estimada da estimativa do estado).

O Filtro de Kalman pode ser escrito por um conjunto de equações, porém ele é mais comumente descrito em duas fases distintas: predição e atualização.

### 5.1.1 Predição

A fase de predição utiliza a estimativa do estado no passo anterior  $\hat{\beta}_{t-1|t-1}$  para obter uma estimativa do estado no tempo atual  $\hat{\beta}_{t|t-1}$ . Esta predição é chamada de estimativa *a priori*, pois não inclui a informação vinda da observação do estado atual.

**Predição do estado (estimativa *a priori*)**

A estimativa *a priori* do estado atual é mostrado na equação 5.1.

$$\hat{\beta}_{t|t-1} = H_t \hat{\beta}_{t-1|t-1} + G_t \eta_t \quad (5.1)$$

**Predição da covariância (estimativa *a priori*)**

$$C_{t|t-1} = H_t C_{t-1|t-1} H_t' + G_t V_t G_t' \quad (5.2)$$

**5.1.2 Atualização**

Na fase de atualização, a predição *a priori* é combinada com a observação atual para refinar a estimativa do estado. A estimativa refinada  $\hat{\beta}_{t|t}$  é chamada de estimativa *a posteriori*.

**Atualização do resíduo da medição**

No tempo  $t$ , uma observação (ou medição)  $y_t$  do estado real  $\beta_t$  é realizada. O erro desta medida em relação à sua estimativa é chamada de erro da medida  $v_t$  e é dado pela equação 5.3.

$$v_t = y_t - X_t \hat{\beta}_{t|t-1} \quad (5.3)$$

**Covariância dos resíduos da medição**

A covariância dos resíduos da medição  $F_t$ , dada pela equação 5.4, é a equação do Filtro de Kalman que inclui a informação da matriz de covariância do distúrbio da medida  $U_t$  no algoritmo.

$$F_t = X_t C_{t|t-1} X_t' + U_t \quad (5.4)$$

**Ganho ótimo de Kalman**

O ganho de Kalman  $K$  é dado pela equação 5.5.

$$K_t = C_{t|t-1} X_t' F_t^{-1} \quad (5.5)$$

O ganho ótimo de Kalman só pode ser garantido se as hipóteses de normalidade forem mantidas. Caso isto não ocorra, o estimador não será mais ótimo, e sim um estimador linear ótimo.

### Estado atualizado (estimativa *a posteriori*)

O estado atualizado é formado por duas componentes. A primeira tem a ver com o estado estimado no momento atual, condicionado à observação no estado anterior. E a segunda com a informação proveniente da observação atual. O que irá determinar o peso de cada uma das componentes e interferir no valor do estado atual é o ganho de Kalman dado pela equação 5.5.

A atualização do estado estimado é mostrada a seguir na equação 5.6.

$$\hat{\beta}_{t|t} = \hat{\beta}_{t|t-1} + K_t v_t \quad (5.6)$$

### Covariância estimada (estimativa *a posteriori*)

A atualização da covariância estimada é mostrada na equação 5.7, também conhecida como *Equação de Riccati*.

$$C_{t|t} = (I - K_t X_t) C_{t|t-1} \quad (5.7)$$

Tipicamente, as fases de predição e de atualização se alternam em um algoritmo recursivo.

Com a predição prevendo o estado até o instante da próxima observação e a atualização incorporando a informação da observação.

A notação encontrada na literatura citada neste trabalho como [15] e [19] pode variar, mas o mais importante é que o conceito de tempo *a priori* e tempo *a posteriori* esteja bem claro. Com o primeiro se referindo ao instante em que a estimativa é feita sem que a medida atual de  $y_t$  tenha sido realizada e o segundo após a medição deste.

De forma geral, pode-se encontrar notações como  $\beta^-$  e  $\beta^+$ . O primeiro, via de regra, representando o estado *a priori* e o segundo o estado *a posteriori*.

## 5.2 O Algoritmo do Filtro de Kalman Discreto

Esta seção descreve o Filtro de Kalman, em sua fórmula original [21], na qual a mensuração da medida do estado estimado é discretizada no tempo.

Sistemas dinâmicos <sup>1</sup> são frequentemente representados em um modelo de Espaço de Estados.

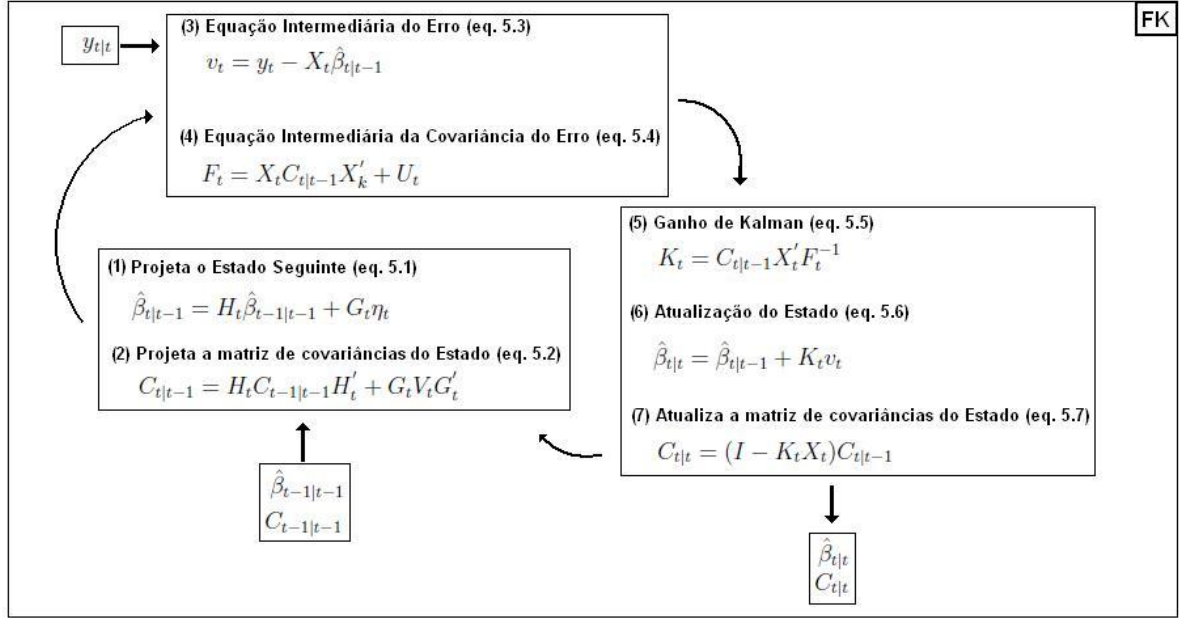
O Filtro de Kalman soluciona o problema geral de tentar estimar o estado de um processo controlado em tempo discreto, que é regido pelas equações de 5.1 a 5.7.

A cada passo, o algoritmo propaga simultaneamente uma estimativa do estado  $\hat{\beta}_{t|t}$  e uma estimativa para sua matriz de covariâncias o  $C_{t|t}$ .

---

<sup>1</sup>Os sistemas dinâmicos são aqueles em que os estados do sistema variam ao longo do tempo.

Figura 5.1: Funcionamento do Filtro de Kalman



O algoritmo mostrado na figura 5.1 descreve o funcionamento do Filtro de Kalman Discreto. Os passos de um até dois compõem a etapa de previsão. Já os passos de três a quatro são equações intermediárias do Filtro de Kalman. Por fim, os passos de cinco até sete compõem a etapa de atualização.

### 5.3 Estimador de Máximo Verossimilhança

O Filtro de Kalman não estima nenhum parâmetro do modelo de Espaço de Estados sozinho. As variâncias  $\sigma^2$  e  $\eta^2$  devem ser fornecidas ao filtro, ou então calculadas por algum outro método. A Maximização da Verossimilhança fornece de forma simples o valor destes estimadores.

A verossimilhança é dada pelo produtório dos termos da distribuição como mostrado a seguir:

$$L(y; \psi) = \prod_0^T p(y_t) \quad (5.8)$$

Uma vez que em uma série temporal, como as descritas neste trabalho, os termos não são independentes, de acordo com Harvey [19] a verossimilhança não pode ser escrita pela equação 5.8 e sim como em 5.9, ou seja:

$$L(y; \psi) = \prod_0^T p(y_t | Y_{t-1}) \quad (5.9)$$



Sabe-se que<sup>2</sup> o vetor de observação  $y_t|Y_{t-1}$  segue uma distribuição Normal mostrada na equação 4.1.

Relembrando as equações intermediárias do Filtro de Kalman 5.3 e 5.4. Tem-se que  $v_t$  é o vetor de erros entre o vetor de observações  $y_t$  e o vetor estimado  $\hat{y}_t$ . E ainda, que  $F_t$  é a matriz de covariâncias de  $v_t$ .

Sendo assim, como descrito por Harvey [19], a expressão 5.10 pode ser maximizada.

$$\log(L(\psi)) = -\frac{T}{2}\log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log|F_t| - \frac{1}{2} \sum_{t=1}^T v_t' F_t^{-1} v_t \quad (5.10)$$

Da mesma forma, Harvey afirma que a expressão 5.11 pode ser minimizada.

$$\log(L(\psi)) = \sum_{t=1}^T \log|F_t| + \sum_{t=1}^T v_t' F_t^{-1} v_t \quad (5.11)$$

Lembrando-se do fato de que tanto o vetor de erros das observações  $v_t$  quanto a matriz de covariâncias  $F_t$  são fornecidas pelo cálculo recursivo do Filtro de Kalman.

O resultado destas maximizações ou minimizações da verossimilhança fornecem as variâncias  $\sigma_t^2$  e  $\eta_t^2$  utilizadas no Filtro de Kalman.

Uma vez encontrada a equação para a verossimilhança, é necessário utilizar de algum processo numérico de otimização que obtenha o valor dos parâmetros desejados.

O algoritmo *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) é citado por Durbin & Koopman [15] mais especificamente no capítulo 7 que trata do assunto, para calcular os parâmetros desejados através da maximização da verossimilhança.

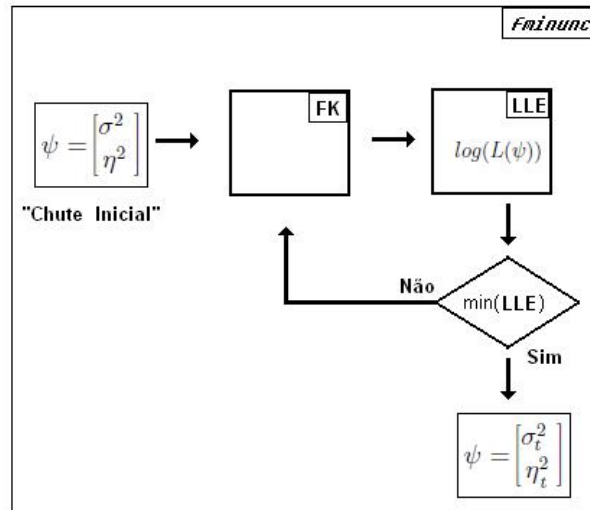
Como o software MATLAB<sup>®</sup>, utilizado neste trabalho, não possui funções de maximização a expressão 5.10, que maximiza a verossimilhança, não foi utilizada na otimização e sim a expressão 5.11 que pode ser utilizada na minimização da verossimilhança. Na implementação foi utilizada a função *fminunc* que possui o algoritmo BFGS implementado.

---

<sup>2</sup>Este ponto foi explicado no item 4.3.1.

A figura 5.2 mostra o fluxograma contendo o funcionamento da função de minimização *fminunc*.

Figura 5.2: Funcionamento do Filtro de Kalman com o Cálculo da Verossimilhança



Após serem obtidos os parâmetros que minimizam a log-verossimilhança. O Filtro de Kalman mostrado na figura 5.1 pode ser utilizado para obter-se a estimativa desejada.

O MATLAB<sup>®</sup> possui outras funções de minimização, como a *fmincon* e a *fminsearch*, mas não serão foco neste trabalho.

# Capítulo 6

## O modelo de De Jong & Zenwirth

O modelo de Espaço de Estados de De Jong & Zenwirth, publicado em 1983 [12], foi o precursor do uso do Filtro de Kalman para o cálculo das provisões de IBNR. Neste trabalho, este modelo, bem como a metodologia utilizada para a estimativa da reserva de IBNR serão tratados como "Abordagem DJZ".

A Abordagem DJZ modela as provisões de sinistros IBNR, utilizando os sinistros já ocorridos e organizados pela diagonal  $y(t)$  do Triângulo de *Run-off*, conforme explicado no item 3.2.4. Para isto, coloca o problema da estimativa da reserva de IBNR na forma de um modelo de Espaço de Estados e se utiliza do Filtro de Kalman para identificar os parâmetros estocásticos  $\beta_t$  do modelo.

A principal contribuição dos autores foi a de "visualizar" esta diagonal  $y(t)$  como o resultado da medida de um sensor com ruídos como tratado por Kalman [21]. Podendo então utilizar-se de um modelo de Espaço de Estados para estimar as provisões futuras destes sinistros.

O Filtro de Kalman foi escolhido para estimar os betas  $\beta_t$  do modelo, criando assim a possibilidade de calcular dados futuros da série observada  $y_t$  obtendo a provisão de IBNR desejada.

Isto equivale, assim como no modelo *Chain Ladder*, a calcular a estimativa futura do vetor de entrada  $y_t$  com o objetivo de estimar  $\hat{y}_{t+1}$  de forma a completar os dados faltantes do Triângulo de *Run-off* os quais compõem a provisão total de IBNR.

### 6.1 Equações do Filtro de Kalman Utilizadas na Abordagem DJZ

As equações do Filtro de Kalman utilizadas na Abordagem DJZ são descritas na forma matricial, isto ocorre, pois existe uma peculiaridade no algoritmo que o difere da abordagem convencional no uso do filtro.

Comumente, o Filtro de Kalman é utilizado para estimar o processo estocástico  $\beta_t$  que

não pode ser visualizado diretamente em um sistema definido. Já na Abordagem DJZ o objetivo é estimar  $\hat{y}(t+1)$ . Ou seja, estimar o vetor de parâmetros  $\hat{\beta}_t$  é o meio para um fim.

Mas por que não utilizar uma modelagem direta para  $\hat{y}(t+1)$  e sim o modelo de Espaço de Estados? Esta é a segunda colaboração do artigo de De Jong & Zenwirth, pois cada diagonal é considerada como sendo independente da diagonal anterior, mas o vetor de estados não. A cada nova passagem de tempo no calendário, a diagonal do Triângulo de *Run-off* aumenta de tamanho (dimensão do vetor) e um novo vetor  $y_t$  é definido, este sendo independente do anterior. Porém, ambos são ligados pelo mesmo vetor de estados  $\beta(t)$ .

Sendo assim, De Jong & Zenwirth [12] utilizam o índice  $t$  para definir qual a ordem de grandeza de  $y(t)$ , observado. Lembrando que existe mais um sub-índice no sentido do tempo de atraso  $d$ , como mostrado pela equação 3.2, que determina a ordem dos elements dentro do vetor.

Considerando  $C(t)$ , e não  $P(t)$  como é convencionalmente chamada, a matriz de covariância de  $\hat{\beta}(t) - \beta(t)$  então o resultado dos estados calculados pelo Filtro de Kalman que estimam a matriz de covariância  $C(t)$  satisfaz as seguintes relações:

$$\hat{y}(t+1) = X(t+1)H(t+1)\hat{\beta}(t) \quad (6.1)$$

$$\hat{\beta}(t) = H(t)\hat{\beta}(t-1) + K(t)\{y(t) - \hat{y}(t)\} \quad (6.2)$$

$$R(t) = H(t)C(t-1)H'(t) + G(t)V(t)G'(t) \quad (6.3)$$

$$K(t) = R(t)X'(t)\{X(t)R(t)X'(t) + U(t)\}^{-1} \quad (6.4)$$

$$C(t) = R(t) - K(t)X(t)R(t) \quad (6.5)$$

Os autores descrevem em seu trabalho as equações de 6.1 a 6.5 como as equações do Filtro de Kalman. Eles também apresentam de forma alternativa as equações 6.4 e 6.5 dadas a seguir:

$$C(t) = \{X'(t)U^{-1}(t)X(t) + R^{-1}(t)\}^{-1} \quad (6.6)$$

$$K(t) = C(t)X(t)U^{-1}(t) \quad (6.7)$$

Na prática só são necessárias as equações de 6.1 a 6.5, pois a partir delas é possível obter a forma alternativa das equações de 6.6 e 6.7.

As equações de 6.1 a 6.5 podem causar confusão quando comparadas às equações de 5.1 a 5.7 do capítulo 5.

A principal delas se refere às matrizes  $C(t)$  e  $R(t)$ . Ambas representam a matriz de covariância de  $\hat{\beta}(t) - \beta(t)$ . No entanto,  $R(t)$  é a covariância *a priori* e  $C(t)$  *a posteriori*. De Jong & Zenwirth devem ter optado por utilizar variáveis diferentes para não causar confusões.

Para um leitor que escolha uma leitura mais atual como Brown [7], esta confusão será inevitável. Pois,  $C(t)$  é comumente chamada de  $P^+(t)$  ou ainda  $P(t)$  e  $R(t)$  é chamada de  $P^-(t)$ .

## 6.2 Premissas da Abordagem DJZ

Até este momento do trabalho, houve a preocupação em criar uma base sólida e consistente<sup>1</sup> sobre os conceitos que envolvem a modelagem do Filtro de Kalman, bem como a provisão de IBNR.

De forma resumida, a abordagem DJZ busca calcular os estados não observáveis  $\hat{\beta}(t)$  utilizando os sinistros observados  $y(t)$ , de forma a poder estimar  $\hat{\beta}(t+1)$  e consecutivamente  $\hat{y}(t+1)$ .

No capítulo 4.3.1 foram colocadas as equações 4.11 e 4.12 que definem o modelo de Espaço de Estados. Para realizar o cálculo da provisão de IBNR basta entender que  $y_t$  descrito pela equação 3.2 é o sinistro já ocorrido e observado e  $\hat{y}_{t+1}$  é o sinistro a ocorrer ou provisão.

Ao se definir o modelo de Espaço de Estados, há a necessidade de se possuir algum conhecimento sobre os dados de interesse e como estes se relacionam para se definir as matrizes de parâmetros do modelo.

Sendo assim, é preciso definir as matrizes de parâmetros  $X_t$  da equação das observações 4.11, bem como as matrizes  $H_t$  e  $G_t$  da equação de estados 4.12.

Estas matrizes irão descrever a forma como a equação de Espaço de Estados irá se relacionar com a equação das observações.

Em seguida serão discutidas as premissas adotadas para definir essas matrizes, da mesma forma como as que foram utilizadas na inicialização do Filtro de Kalman.

---

<sup>1</sup>Espera-se que o trabalho tenha sido todo lido e que as citações tenham sido ao menos consultadas para tanto.

### 6.2.1 Índices de Inflação e Volume de Apólices

O modelo proposto na Abordagem DJZ é bem genérico e permite, por exemplo, que índices de volume de novos contratos e índices de inflação estejam incluídos na modelagem. Os autores podem ter sido influenciados por [6] no qual foi proposto um modelo habilitado a tais inclusões. Dado que este tipo de modelagem gera novos erros na estimação para passos futuros, a escolha, neste trabalho, foi a de não considerar tais índices. Além disso, a ideia é testar o funcionamento do Filtro de Kalman como preditor e sua eficiência em contrapartida do método *Chain Ladder*. Sendo assim, "erros" de outras estimativas, como os valores futuros do índice de inflação, utilizados na modelagem poderiam interferir na decisão.

### 6.2.2 A matriz de Parâmetros das Observações

Na equação das observações 4.11,  $y_t$  é descrito por um nível mais um ruído de média zero. A matriz de parâmetros  $X_t$  é que determina a forma como o vetor de estados  $\beta(t)$  descreve as observações dos sinistros medidos. Esta matriz deve ser parametrizada com valores constantes, ou ainda com uma função determinística que depende do tempo. Em [12] os autores optam pela segunda opção.

Essa função determinística, que será definida como  $\phi(d)$ , representa o conhecimento *a priori* do sistema físico a ser estimado pelo modelo de Espaço de Estados. Por se tratar de uma função arbitrária, deve ser informada pelo atuário ou usuário do modelo no caso da estimativa da provisão de IBNR.

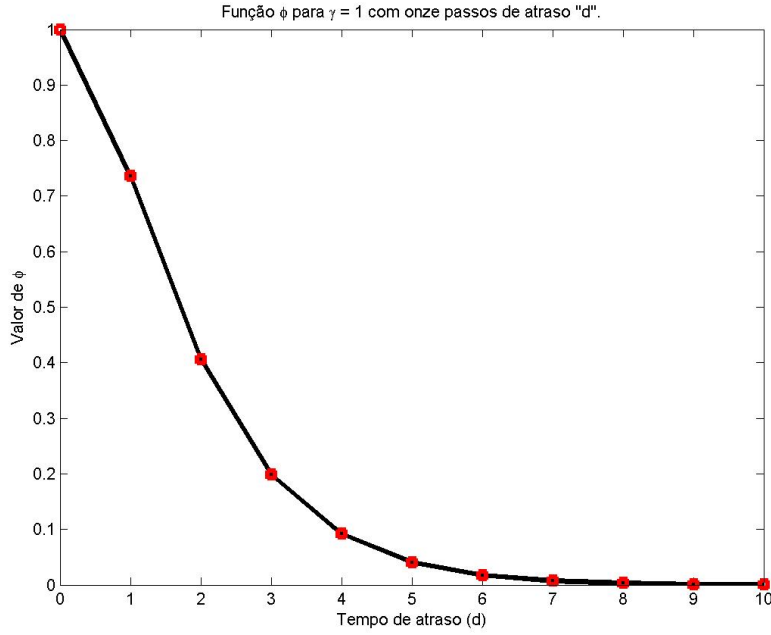
De Jong & Zenwirth optaram por utilizar uma função da família exponencial negativa.

$$\phi(d) = (d + 1)e^{(-d\gamma)} \quad (6.8)$$

Nota-se que na equação 6.8,  $d$  representa o "passo" de tempo de atraso no aviso do sinistro em relação a abertura do mesmo. Já  $\gamma$  é um parâmetro da função a ser definido. Na abordagem DJZ [12], os autores utilizam  $\gamma = 1$ . Sendo assim a equação torna-se:

$$\phi(d) = (d + 1)e^{(-d)} \quad (6.9)$$

O comportamento da equação 6.9 é mostrada na figura 6.1.

Figura 6.1: Função  $\Phi$  escolhida em DJZ

Segundo os autores, o uso da função  $\phi$  se justifica, pois para cada ano de origem  $w$ , os valores de ocorrência de aviso dos sinistros decaem monotonicamente ou no caso exponencialmente, conforme o desenvolvimento  $d$  do sinistro aumenta. No item 6.4.3 uma discussão mais aprofundada sobre essa função é realizada, justificando a sua utilização.

A Abordagem DJZ referencia a função  $\phi(d)$  como  $\phi_d(t)$ . Porém,  $\phi_d(t)$  não depende do tempo  $t$  e sim do tempo de atraso  $d$ . Ao tratar a função como  $\phi_d(t)$ , os autores desejaram indicar a ordem do vetor  $\phi$  que depende de  $t$ . A forma matricial da função é dada a seguir:

$$\phi(t) = \begin{bmatrix} \phi_0(t) \\ \phi_1(t) \\ \vdots \\ \phi_j(t) \\ \vdots \\ \phi_d(t) \end{bmatrix}$$

Salientando-se neste ponto que esta função  $\phi_d(t)$  determinará se o filtro estará calibrado em relação aos sinistros observados, *i.e.*, o quanto os sinistros estimados irão "seguir" o modelo. Mais detalhes serão discutidos no capítulo 7.

Por fim, a matriz de parâmetros  $X_t$ , que determina a forma como o vetor de estados descreve as observações dos sinistros medidos, será uma matriz diagonal, tendo sua diagonal principal formada pelo vetor  $\phi(t)$ .

Por fim **a equação das observações estimadas** é dada pela forma matricial como segue:

$$\begin{bmatrix} y_t(t) \\ y_{t-1}(t) \\ \vdots \\ y_1(t) \end{bmatrix} = \begin{bmatrix} \phi_0(t) & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \phi_1(t) & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_j(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \phi_{t-1}(t) \end{bmatrix} \cdot \begin{bmatrix} b_t(t) \\ b_{t-1}(t) \\ \vdots \\ b_1(t) \end{bmatrix}$$

Vale lembrar que na notação acima,  $\phi_j(t)$  depende de  $t$  somente pela ordem do vetor  $\phi(t)$ , no entanto, substituindo  $d = t - 1$ , o tempo de atraso  $d$  pode ser eliminado da expressão, simplificando a notação.

### 6.2.3 Estimativa do Vetor de Estados

O Filtro de Kalman é capaz de estimar o vetor de estados  $\beta(t)$  e a sua matriz de covariâncias  $C$ . No entanto, é necessário definir antes as matrizes  $H_t$  e  $G_t$  da equação 4.12 que definem a transição entre estados.

A Abordagem DJZ também é genérica quanto a este ponto. *Lags*, ou ainda atrasos de tempo  $q$  são utilizados para definir o quão dependente dos estados anteriores é o estado atual  $\beta_t$ . Considerando  $\beta_t(t) = [b_t(t), b_{t-1}(t), \dots, b_1(t)]'$  de dimensões  $t \times 1$ , cada  $b_t(t)$  da equação 4.12 é dado por:

$$b_t(t) = \sum_{j=1}^q a_j b_{t-j}(t) + \psi \eta_t \quad (6.10)$$

Novamente neste trabalho, uma simplificação é feita na equação 6.10. Tomando apenas um *lag*, ou seja, para  $q = 1$  e  $a_j = 1$ . E definindo ainda que não há nenhum multiplicador no distúrbio  $\eta_t$ , ou seja,  $\psi = 1$ . A equação 6.10 se torna um Passeio Aleatório mostrado na equação 6.11 a seguir:

$$b_t(t) = b_{t-1}(t) + \eta_t \quad (6.11)$$

Seguindo o raciocínio anterior, a equação 4.12, ou ainda **a equação dos estados** pode ser escrita na forma matricial como:



$$\begin{bmatrix} b_t(t) \\ b_{t-1}(t) \\ \cdot \\ \cdot \\ \cdot \\ b_1(t) \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 \\ 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} b_{t-1}(t) \\ b_{t-2}(t) \\ \cdot \\ \cdot \\ b_1(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \eta_t$$

#### 6.2.4 Matriz de Covariância do Ruído das Observações

Ao construir o vetor de observações  $y_t$  a partir do Triângulo de *Run-off* fica fácil notar<sup>2</sup> que os valores de  $y_t$  decrescem com o tempo. Desta forma, seus momentos dependem do tempo de atraso  $d$ , portanto, o ruído das observações de  $y_t$  não pode ser considerado Homocedástico (variância constante).

Sendo assim, alguma função dependente do tempo deve ser estabelecida para que a tendência Heterocedástica seja retirada da matriz de covariância do ruído das observações.

#### 6.2.5 Inicialização Difusa do Filtro de Kalman

Ao iniciar o algoritmo do Filtro de Kalman é necessário especificar qual o valor inicial do estado estocástico  $\beta_0(0)$  e de sua matriz de covariância  $C_0(0)$ . Uma vez que estes valores teriam de ser definidos para um momento inicial anterior ao que os dados observados de  $y_1(1)$  se referem, ou ainda no ano de ocorrência  $w = 0$  não observado, o que na prática seria inviável.

Na Abordagem DJZ, os autores citam no apêndice C de seu trabalho um algoritmo para a Inicialização Difusa Exata (IDE) do Filtro de Kalman.

No entanto, segundo Pizzinga [29], a Inicialização Difusa Aproximada (IDA) é uma prática bastante utilizada na inicialização do Filtro de Kalman devido à sua facilidade de implementação e por obter resultados muito similares aos encontrados na Inicialização Difusa Exata do Filtro de Kalman. Optou-se, portanto, pela Inicialização Difusa Aproximada do Filtro de Kalman para assim simplificar o método.

Na Inicialização Difusa Aproximada, a função estocástica  $\beta(0)$  é definida como zero e a sua matriz de covariância  $C_0(0)$  é inicializada com um valor tendendo ao infinito. No caso, esta foi inicializada com  $C_0(0) = 10^9$ .

Com o valor da matriz de covariância alto, o Filtro "aprenda rápido", *i.e.* as estimativas caminham no sentido das medidas e não do modelo preconcebido.

---

<sup>2</sup>Esta discussão será aprofundada no item 6.4.3.

### 6.3 Estimativa da Reserva de IBNR por EE e FK

Diferentemente da aplicação clássica do Filtro de Kalman, que estima o estado  $\beta(t)$  com base na observação atual  $y(t)$ , sem carregar toda a informação histórica, a Abordagem DJZ estima a cada passo, todo o vetor  $\beta(t)$ , que aumenta de tamanho conforme cada novo período de calendário ocorre. No passo seguinte, o algoritmo leva todo o novo vetor para a estimação do próximo período. Ou seja, nenhuma informação é descartada, pois o vetor de parâmetros  $\beta(t)$  que possui dimensão  $t \times 1$  aumentará de tamanho e terá dimensão  $t + 1 \times 1$  para  $\beta(t + 1)$ . Isto torna a estimativa do Filtro de Kalman eficiente no sentido de estimar todos os estados para o cálculo da diagonal seguinte  $\hat{y}(t + 1)$ .

Seguindo este raciocínio, para completar o Triângulo de *Run-off* será preciso "avançar" no tempo, ou seja, o algoritmo irá além do tempo  $t$ . No entanto, não existem mais dados observados para um tempo  $s > t$ . Sendo assim, o filtro não terá mais "etapas de atualização".

Como o objetivo é encontrar os valores que completem o Triângulo de *Run-off* e não valores de sinistros futuros que ainda não ocorreram, o valor inicial estimado  $\hat{y}_0(s)$  de  $\hat{y}(s)$  será descartado a cada passo.

Sendo assim,  $\hat{y}(s)$  terá uma dimensão menor do que  $\hat{y}(t)$ . Considerando  $s = t + k$ ,  $\hat{y}(s)$  terá dimensões  $t - k \times 1$ .

$$\hat{y}(t + k) = \hat{y}(s) = \begin{bmatrix} \hat{y}_k(s) \\ \vdots \\ \hat{y}_t(s) \\ \vdots \\ \hat{y}_{s-1}(s) \end{bmatrix} \quad (6.12)$$

Como não ocorrerão novos valores de  $y(t)$ , os estados  $\beta(t)$  permanecerão fixos tornando a previsão determinística e não mais estocástica.

Em seguida, utilizando o vetor de estados  $\hat{\beta}(s)$  e sua matriz de covariâncias  $C(s)$ , ambos calculados pelo Filtro de Kalman para o instante  $t$ , é possível estimar o valor da provisão de IBNR  $\hat{y}(s)$  em cada instante de tempo  $s$ , sendo  $t < s < 2t - 1$ .

## 6.4 O exemplo DJZ

### 6.4.1 Os dados DJZ

Como forma de ilustrar o modelo proposto em [12], os autores utilizam os dados da tabela 6.1 que serão chamados a partir deste ponto de "dados DJZ". Estes por sua vez foram obtidos de Benjamin (1977) [4] e se referem aos sinistros de uma seguradora do Reino Unido referentes aos anos de 1970 até 1974. Os mesmos são descritos em milhares de Libras Esterlinas<sup>3</sup>.

Tabela 6.1: Dados DJZ

w/d	0	1	2	3	4
1	753,5	648,9	311,7	173,5	71,3
2	642,3	648,4	249,7	206,5	
3	715,8	661,1	309,4		
4	841,6	862,6			
5	968,8				

Na tabela 6.1 foi propositalmente omitido o índice de inflação, bem como o índice de volume, pois estes dados não serão utilizados neste trabalho, como mencionado no item 6.2.1.

A Abordagem DJZ utiliza a ordenação pela diagonal explicada no item 3.2.4 da página 24. Sendo assim, a ordem dos dados DJZ sofrem uma alteração ao serem guardados nos vetores  $y_t$ . Esta ordem é mostrada abaixo, de forma a facilitar o entendimento, este mesmo arranjo pode ser encontrado em [12] e [1].

$$\begin{aligned}
 y(1) &= (753,5) \\
 y(2) &= (642,3; 648,9)' \\
 y(3) &= (715,8; 648,4; 311,7)' \\
 y(4) &= (841,6; 661,1; 249,6; 173,5)' \\
 y(5) &= (968,8; 862,6; 309,4; 206,5; 71,3)'
 \end{aligned}$$

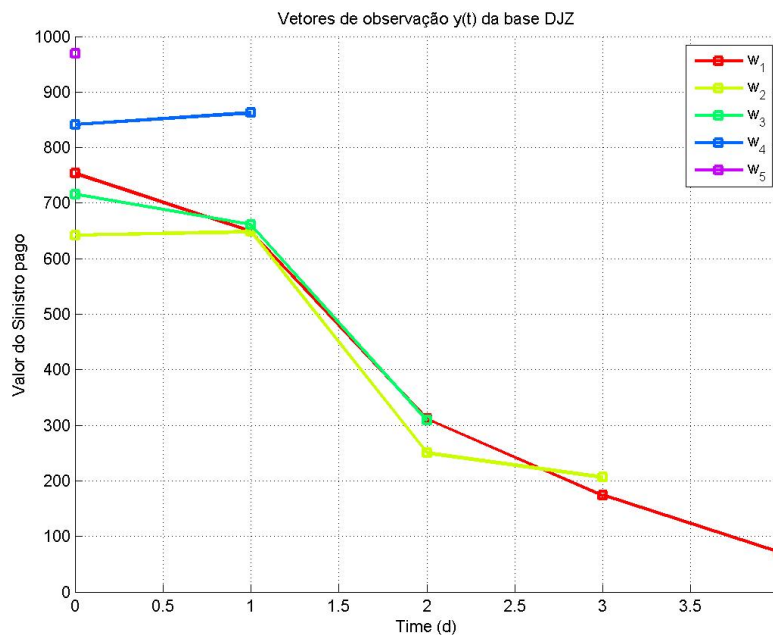
Os vetores descritos possuem dimensões  $t \times 1$ , ou seja, o último vetor terá dimensão  $5 \times 1$ , com cinco linhas e uma coluna.

A figura 6.8 mostra os estados medidos, ou ainda, os sinistros já ocorridos e separados pelo período de ocorrência.

---

<sup>3</sup>Moeda oficial do Reino Unido

Figura 6.2: Sinistros observados da base DJZ



### 6.4.2 Estimando a Provisão de IBNR pelo Método *Chain Ladder* para os Dados DJZ

Utilizando o modelo de *Chain Ladder*, explicado no item 3.3 da página 25, foram calculadas as estimativas das provisões de IBNR para os dados DJZ.

A tabela com a provisão total da reserva estimada de IBNR pelo método *Chain Ladder* é apresentada abaixo:

Tabela 6.2: Reserva de IBNR estimada pelo método *Chain Ladder* para os dados DJZ

Dados DJZ em Milhares de Libras					
w/d	0	1	2	3	4
1	753,50	648,90	311,70	173,50	71,30
2	42,30	648,40	249,70	206,50	65,99
3	715,80	661,10	309,40	196,89	71,13
4	841,60	862,60	364,62	241,56	87,27
5	968,80	925,43	405,28	268,49	97,00

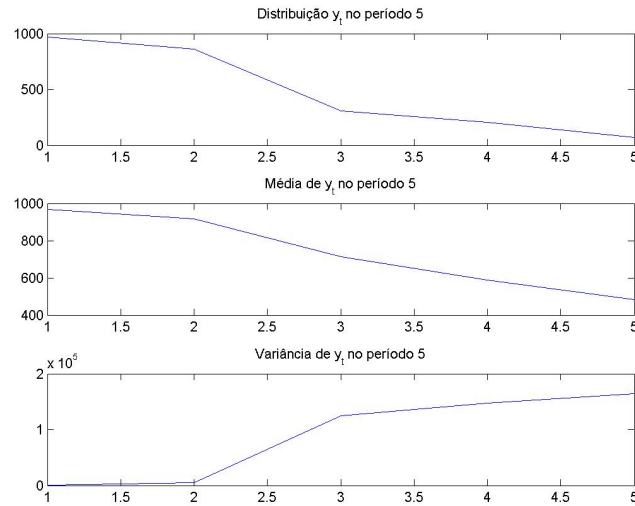
A estimativa total das provisões de IBNR encontrada pelo método *Chain Ladder* foi de 2.723,67 milhares de libras esterlinas.

Este valor serve como referência para o resultado que será estimado pelo método de Espaço de Estados e Filtro de Kalman.

### 6.4.3 Heterocedasticidade da base DJZ

A figura abaixo mostra a distribuição e os momentos de primeira e segunda ordem para  $y_t$  no quinto ano de observação.

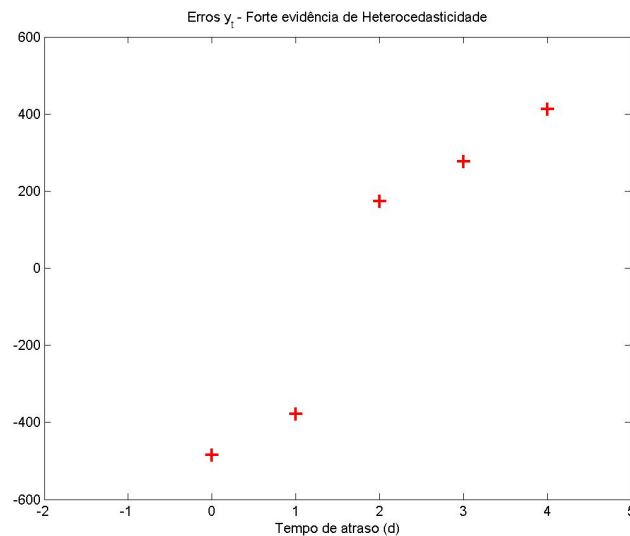
Figura 6.3: Momentos DJZ ao longo do tempo



Da figura 6.3 fica fácil notar que tanto a média quanto a variância mudam ao longo do tempo, ou seja, a hipótese de homocedasticidade não pode ser considerada.

Já a figura 6.4 mostra a distribuição dos erros de  $\hat{y}_t - y_t$ , na qual nota-se forte tendência de crescimento ao longo do tempo.

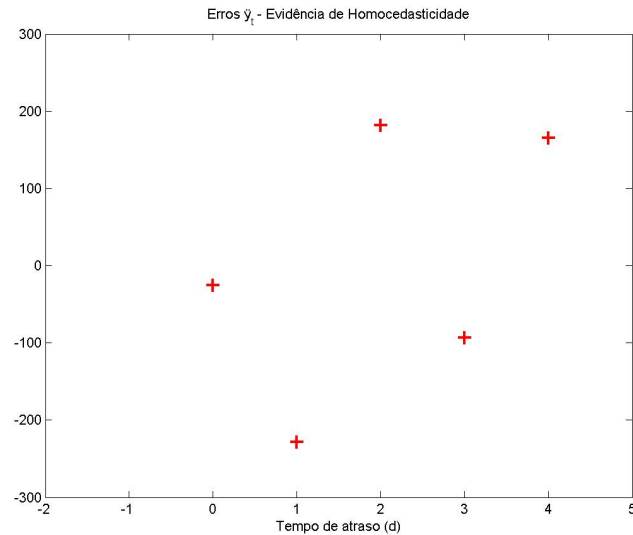
Figura 6.4: Erros DJZ ao longo do tempo



Por outro lado, ao multiplicar-se os dados do vetor  $y_5$  da base DJZ pela função  $\phi(4)$

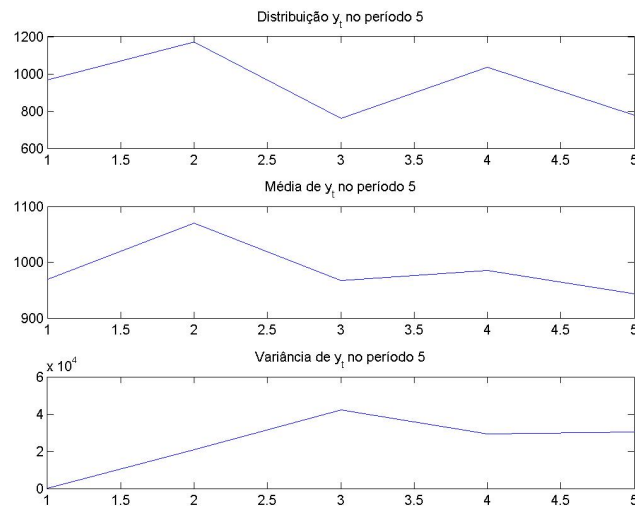
obtem-se uma função ajustada  $\tilde{y}_5$ . O ajuste, na função  $y_5$ , gera uma nova função que mostra forte evidência de homocedasticidade necessária ao modelo.

Figura 6.5: Erros DJZ ao longo do tempo - Sem tendência



Os novos momentos da função ajustada  $\tilde{y}_5$  são descritos na figura 6.6 da página 53.

Figura 6.6: Momentos DJZ ajustados ao longo do tempo



O primeiro momento da nova função mostra uma menor tendência de decaimento em relação ao tempo. Já o segundo momento mostra um fraco indício de estabilização ao longo do tempo, em torno de um valor fixo  $\sigma^2$ .

Diante disso, fica evidente que algum ajuste na matriz de covariância  $U(t)$  é necessário. Segundo [2], é possível ajustar  $U(t)$  como mostrado na equação 6.13 abaixo:

$$\text{var}(\epsilon_d(t)|Y_{t-1}) = \sigma_t^2 E[y_d(t)|Y_{t-1}] \quad (6.13)$$

O que na prática leva a expressão de  $U(t)$  ser escrita na forma de uma matriz diagonal de dimensões  $t \times t$  com seus valores dados por:

$$U(t) = \sigma_t^2 \hat{y}(t) \quad (6.14)$$

Sendo que  $\sigma_t^2$  é a variância homocedástica que deve ser estimada<sup>4</sup> para construir a matriz de covariância heterocedástica  $U(t)$  das observações.

#### 6.4.4 Estimativa dos Parâmetros Calculados pelo Filtro de Kalman para os Dados DJZ

Com base no que foi discutido até este ponto do trabalho, foi construída em MATLAB<sup>®</sup> uma função SST contendo o Filtro de Kalman e a função de Verossimilhança que deve ser minimizada. Para isto a função SST deve ser passada como parâmetro da função *fminunc* do MATLAB<sup>®</sup>.

O valor inicial dos parâmetros, também passados como parâmetro da função *fminunc*, foi alterado arbitrariamente para se testar a eficiência da função perante a minimização da verossimilhança.

A tabela com as variâncias  $\sigma_t^2$  e  $\eta_t^2$  estimadas pela minimização da verossimilhança variando o valor inicial da função *fminunc* é mostrado abaixo:

Tabela 6.3: Parâmetros  $\sigma_t^2$  e  $\eta_t^2$  para DJZ

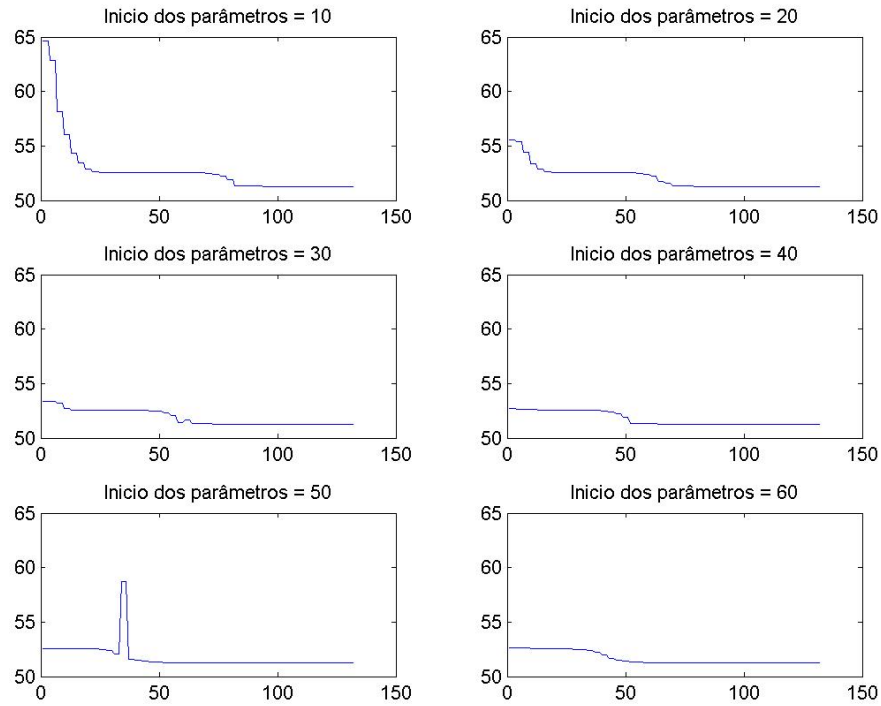
Parâmetros $\sigma_t^2$ e $\eta_t^2$ em Milhares de Libras						
Chute Inicial	10	20	30	40	50	60
$\sigma^2$	31,67	31,67	31,67	31,67	31,67	31,67
$\eta^2$	12098,85	12098,84	12098,85	12098,90	12098,85	12098,85

O "chute" inicial consiste em ajustar os parâmetros  $\psi = [\sigma^2; \eta^2]'$  de forma que possa ser testada a eficiência da função *fminunc* em encontrar o mínimo global da função de verossimilhança. Fica claro pela tabela que a função ficou estável diante dos parâmetros iniciais testados e o mínimo local foi encontrado.

<sup>4</sup>Pela Máximo Verossimilhança em conjunto com um método de otimização.

A figura 6.7 mostra o comportamento da minimização da verossimilhança para cada um dos valores inicial dos parâmetros  $\sigma_t^2$  e  $\eta_t^2$  testados.

Figura 6.7: Minimização dos parâmetros  $\sigma_t^2$  e  $\eta_t^2$  para DJZ



#### 6.4.5 A provisão de IBNR calculada pelo Filtro de Kalman em DJZ

Com os parâmetros  $\sigma^2$  e  $\eta^2$  calculados pelo Filtro de Kalman da função SST foi possível estimar a provisão de IBNR para os dados DJZ.

A tabela 6.4 representa o Triângulo de *Run-off* completo, contendo tanto os dados observados como os calculados, e é mostrada abaixo:

Tabela 6.4: Dados DJZ em Milhares de Libras calculado por SST

Dados DJZ em Milhares de Libras					
w/d	0	1	2	3	4
1	753,50	648,90	311,70	173,50	71,30
2	642,30	648,40	249,70	206,50	70,44
3	715,80	661,10	309,40	161,49	74,26
4	841,60	862,60	367,11	180,07	82,80
5	968,80	687,19	379,20	186,00	85,53



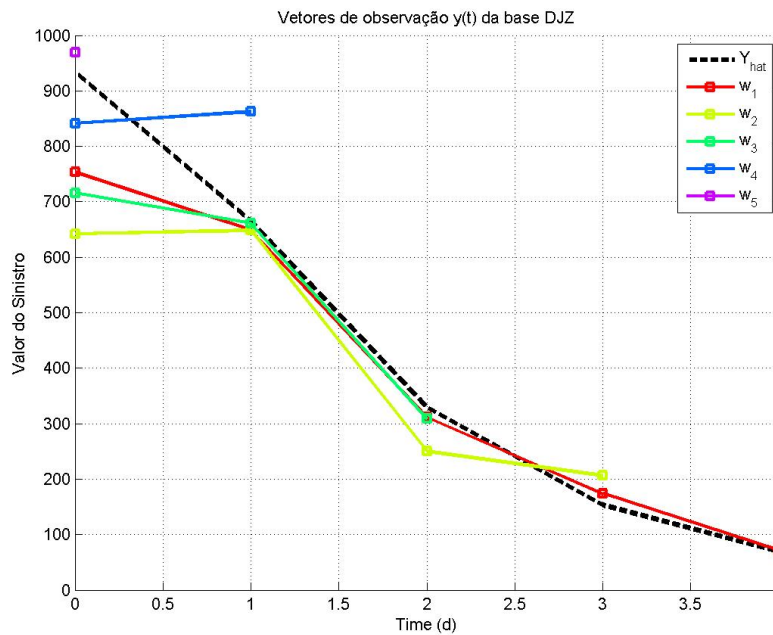
A comparação das provisões de IBNR para os dados de DJZ calculadas pelos dois métodos é mostrada na tabela 6.5:

Tabela 6.5: Resultados dos dados DJZ

IBNR	Total	Desv. Padrão
Espaço de Estados	2.274,09	120,55
Chain Ladder	2.723,67	

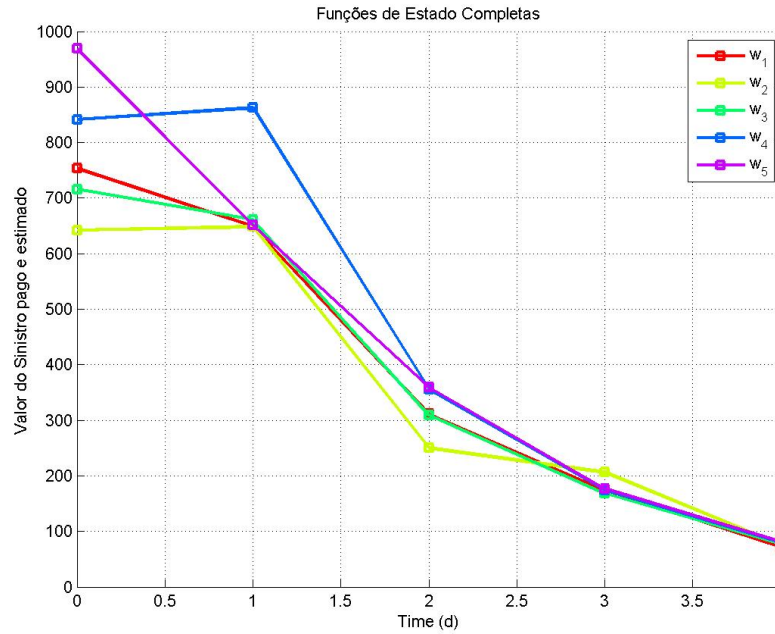
A estimativa de provisão total das reservas de IBNR foi menor do que para o método *Chain Ladder*.

Na figura 6.8 é possível visualizar os sinistros ocorridos, apresentados na tabela 6.1, juntamente com o último  $\hat{y}(t)$  estimado pelo Filtro de Kalman e que será utilizado na estimação. Este aparece em destaque pela linha tracejada negra e se ajusta perfeitamente aos sinistros já ocorridos.

Figura 6.8: Sinistros da base DJZ e  $\hat{y}(t)$  estimado

A figura 6.9 mostra os sinistros ocorridos, juntamente com os estimados pelo modelo de Espaço de Estados e Filtro de Kalman, ou seja, o Triângulo de *Run-off* completo da tabela 6.4.

Fica claro na figura 6.9 que o modelo de Espaço de Estados e Filtro de Kalman se ajustou ao Triângulo de *Run-off* dos dados DJZ mostrando que o modelo pode ser usado como um estimador da provisão de IBNR.

Figura 6.9: Dados DJZ - Triângulo de *Run-off* completo

Na verdade, como a projeção é determinística, os dados do Triângulo de *Run-off* representados por  $\hat{y}(s)$  serão definidos por  $\hat{y}(t)$ .

Como a base com os dados DJZ só possui cinco passos de tempo, esta se torna pequena para estimar os dados e ainda reservar alguma "diagonal" para se averiguar a eficácia do modelo com dados reais. Ficando, portanto, a critério da empresa seguradora escolher qual o método a ser utilizado para o cálculo das provisões de IBNR, se o *Chain Ladder* que foi mais conservador ou o modelo de Espaço de Estados e Filtro de Kalman.

## Capítulo 7

# Estimativa de IBNR, um Caso Brasileiro

No capítulo 6 foram estimadas as reservas de provisão de IBNR com base nos sinistros de uma seguradora do Reino Unido. Estes sinistros, chamados aqui de dados DJZ, foram obtidos do trabalho original de De Jong & Zenwirth. Para o cálculo destas provisões foram utilizados dois métodos, o *Chain Ladder* e um modelo de Espaço de Estados e Filtro de Kalman que foi especificado e descrito no mesmo capítulo.

Nesta parte do trabalho serão utilizadas estas mesmas técnicas e premissas para estimar a reserva de IBNR de uma carteira de seguros de um produto pertencente a um ramo Não Vida utilizando os dados reais de uma seguradora brasileira.

### 7.1 Os Dados SBR

Neste estudo serão utilizados dados reais de uma grande seguradora brasileira para um ramo Não Vida. A partir deste ponto estes dados serão referenciados como "**dados SBR**".

Cada ramo de seguro pode apresentar comportamentos muito diferentes no que tange à abertura de seus sinistros. Ramos Não Vida costumeiramente apresentam "caudas curtas" quando comparados aos demais ramos, excluindo-se claro, os casos de sinistros judiciais.

Classicamente a provisão de reservas de IBNR pode ser dividida em duas componentes, correspondendo aos dois tipos de atrasos que ocorrem nas empresas seguradoras.

- 1 IBNR (Incurred But Not Reported) reservas referentes a sinistros que ocorreram, mas só foram avisadas após a data do ocorrido.
- 2 IBNER (Incurred But Not Enough Reported) reservas referentes a sinistros IBNR que foram avisadas e mesmo após o aviso, o valor contabilizado pode ser alterado.

Assim como em De Jong & Zenwirth[12], este trabalho não realizou a separação destas duas componentes. Sendo assim, para efeito de simplificação todos os cálculos e análises deste capítulo serão realizados com a soma das duas componentes tomadas simplesmente como IBNR.

Considerar o IBNER como IBNR é uma simplificação que pode acarretar alguns problemas, pois diferentes seguradoras podem apresentar diferenças significativas no desenvolvimento de seus sinistros, mesmo para ramos de seguros iguais. Processos mal desenhados e falhas operacionais podem provocar aberturas equivocadas de sinistros que posteriormente são recusados. Causando assim, grandes estornos que prejudicam o decaimento dos avisos, criando vales e até mesmo "sinistros negativos".

Quebras estruturais como as causadas por aspectos macroeconômicos podem afetar a receita da empresa seguradora, influenciando na política de pagamento dos sinistros. Estes tipos de mudanças também afetam de forma significativa o comportamento dos desenvolvimentos dos sinistros.

Em mercados menos "maduros" como o Brasil, pode ocorrer uma discussão maior sobre a cobertura da apólice e o montante a ser pago em caso do sinistro. Podem ainda haver processos diferentes entre seguradoras e a qualidade destes processos alteram o comportamento das provisões. Exemplificando, uma seguradora pode optar por abrir seus sinistros no momento do aviso através de um valor fixo ou um percentual do valor do bem segurado. Já outra pode optar por obter maiores detalhes sobre o sinistro e utilizar tabelas para contabilizá-lo entre outras tantas situações que afetam o fluxo das contabilizações e, portanto, das provisões que vierem a utilizar estes dados.

Os pontos levantados acima, sem dúvida, afetam a informação que será utilizada na modelagem da reserva de IBNR. Os dados SBR não são uma exceção. Estes foram coletados mensalmente de uma carteira de seguros já maturada dentre os meses de abril de 2012 a julho de 2013. A informação anterior a este período sofreu falhas severas de processo e foi desconsiderada.

A seguir na tabela 7.1 são mostrados os dados SBR em milhares de reais:

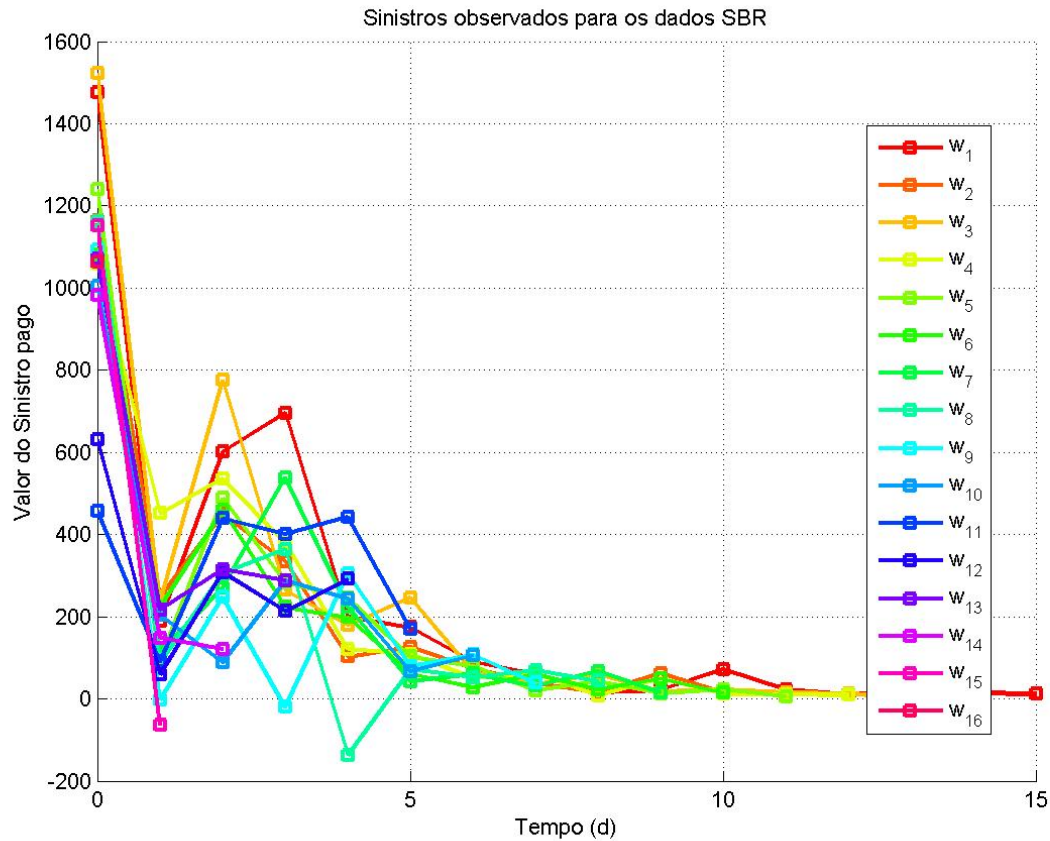
Tabela 7.1: Dados SBR - Milhares de Reais

Dados SBR em Milhares de Reais																
w/d	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1475,33	190,11	600,63	695,47	200,14	174,43	90,74	59,25	17,10	21,25	71,77	21,56	8,95	17,56	14,11	11,49
2	1164,67	247,87	448,87	333,44	103,16	125,80	73,96	38,40	15,11	61,43	15,39	15,51	11,05	8,31	15,00	
3	1522,17	236,47	775,35	264,99	179,16	245,78	64,97	35,40	60,30	16,84	22,99	14,96	11,26	5,69		
4	1059,67	451,22	536,37	381,75	121,02	109,16	50,91	61,57	8,02	49,74	11,33	10,21	9,51			
5	1240,37	94,31	487,93	282,13	244,65	102,22	79,85	20,71	43,81	13,37	23,24	6,25				
6	1081,75	223,58	458,64	223,00	198,59	60,67	28,24	59,96	22,77	49,21	13,98					
7	1092,19	112,31	270,58	538,30	223,10	42,23	60,60	31,91	67,14	14,98						
8	1160,82	88,77	304,95	363,73	-136,66	71,78	52,22	69,52	43,24							
9	1092,32	-2,03	247,96	-19,84	304,08	78,93	107,91	38,98								
10	1003,94	203,07	87,47	286,94	243,64	66,61	104,64									
11	457,21	90,44	438,93	401,16	443,37	170,54										
12	630,17	59,72	308,31	213,21	291,67											
13	1071,93	216,93	315,03	287,74												
14	983,01	148,67	121,05													
15	1152,92	-63,07														
16	1062,68															

Os dados da tabela 7.1 mostrados acima englobam todos os sinistros abertos e efetivamente pagos. Estão inclusos tanto os sinistros que foram abertos e pagos até a data

de aviso conhecidos como IBNR real e os que sofreram correções nos valores ao longo do tempo de atraso  $d$  após terem sido avisados, conhecidos como IBNER.

Figura 7.1: Sinistros já observados para os dados SBR



Na figura 7.1 é mostrado o comportamento dos dados SBR, cujo valores foram mostrados na tabela 7.1, dispostos por ano de ocorrência  $w$  e cada linha é representada por  $w_t$ , conforme apontado pela legenda.

Dois pontos relevantes podem ser notados na figura 7.1. O primeiro refere-se à diferença entre as linhas do gráfico ou ainda os períodos de ocorrência dos sinistros, indicando uma forte evidência de sazonalidade. O segundo refere-se aos valores negativos que são mostrados em vermelho na tabela 7.1. Estes ocorrem pois o sinistro apurado no período foi menor do que os estornos ocorridos em relação aos lançamentos ocorridos no período anterior.

## 7.2 Estimativa pelo Método *Chain Ladder* para os Dados SBR

Como realizado no capítulo 6, utilizando o modelo de *Chain Ladder* foram calculadas as estimativas das provisões de IBNR para os dados SBR.

A estimativa total das provisões de IBNR encontrada pelo método *Chain Ladder* foi de 4,64865 milhões de reais.

A tabela A.1 com a provisão total da reserva estimada de IBNR pelo método *Chain Ladder* é apresentada no apêndice A.

## 7.3 Estimativa dos Parâmetros Calculados pelo Filtro de Kalman para os Dados SBR

Os mesmos passos realizados no capítulo 6 foram realizados nesta etapa para estimar os parâmetros do vetor  $\psi_t = [\sigma_t^2; \eta_t^2]'$  pela minimização da verossimilhança.

O valor inicial dos parâmetros passados para a função *fminunc* foram alterados arbitrariamente para testar a eficiência da função de otimização.

Em um primeiro momento, o valor de  $\psi_t$  foi determinado em  $\psi_t = [1.000; 1.000]'$ , ou seja, um milhão de reais. O resultado foi inesperado, pois a função de otimização retornou variâncias "negativas".

A tabela com as variâncias  $\sigma_t^2$  e  $\eta_t^2$  estimadas pela minimização da verossimilhança variando a quantidade de meses utilizados é mostrado abaixo:

Tabela 7.2: Resultados dos parâmetros  $\sigma_t^2$  e  $\eta_t^2$  calculados pelo Filtro de Kalman sem critério de parada

t	Parâmetros $\sigma_t^2$ e $\eta_t^2$											
	5	6	7	8	9	10	11	12	13	14	15	16
$\sigma$	97,56	-3,07	159,82	1000,00	185,85	1000,00	885,77	0,29	352,17	510,64	757,16	1007,17
$\eta$	5353,91	1002,28	7880,22	1000,00	12450,78	1000,00	-7171,03	1094,56	-1460,84	-3088,13	-5293,18	-9579,29

A falha na otimização ficou nítida, pois não existem variâncias "negativas". O que ocorre é que *fminunc* é uma função *unconstrained*, ou seja, não possui critério de parada. Assim, uma limitação deve ser criada de forma que valores negativos não possam ser utilizados. Outro ponto que vale ressaltar é o de que o "chute" em 1.000 não funcionou bem, pois no período de oito meses e de dez meses não houve minimização.

No trabalho de Christensen *et. al.*[10], uma série de adaptações são citadas para resolver este tipo de situação. Sendo assim, duas mudanças foram realizadas. A primeira, foi a de incluir uma limitação no código da função SST, garantindo que os parâmetros iniciais são da forma  $\psi_t = \log(e^{\psi_t})$ . Isto faz com que a minimização pare sem utilizar

valores negativos. A segunda, foi a de definir o valor inicial para dez milhões de reais, ou  $\psi_t = [10.000; 10.000]'$ , lembrando que os dados da base SBR foram divididos por mil.

Os valores encontrados na nova estimação são mostrados na tabela 7.3.

Tabela 7.3: Resultados dos parâmetros  $\sigma_t^2$  e  $\eta_t^2$  calculados pelo Filtro de Kalman com critério de parada

t	Parâmetros $\sigma_t^2$ e $\eta_t^2$											
	5	6	7	8	9	10	11	12	13	14	15	16
$\sigma$	97,56	185,99	159,82	247,17	185,85	254,44	1054,05	463,66	469,28	765,19	1372,20	2789,93
$\eta$	5353,91	53066,33	7880,22	4842,50	12450,77	14365,94	0,00	8913,54	0,00	0,00	0,00	9181,42

O resultado da nova estimação foi bem melhor que o anterior, por fazer mais sentido em relação ao esperado para as variâncias. Porém, para alguns meses, o valor de  $\eta_t^2$  foi zero.

Se  $\sigma_t^2$  é muito grande em relação a  $\eta_t^2$ , ou ainda, o ruído de medida é muito maior que o ruído do estado, o Filtro de Kalman acompanha a matriz  $X(t)$ , ou seja, o modelo determinístico definido no item 6.2.2, e filtra tudo. Abandonando assim, as medidas que não são considerados confiáveis. Já se o ruído do processo estocástico é muito maior que o ruído da medida, então o modelo é abandonado e somente as medidas são consideradas.

Com base nisso, foi criada uma recursividade no algoritmo que "otimiza" a minimização das estimativas. O valor inicial de  $\psi_t$  foi estipulado ser  $\psi_t = [10.000; 10.000]'$ . Caso o valor de  $\eta_t^2$  encontrado fosse zero, o algoritmo reinicializa e um novo  $\psi_t$  é determinado em  $\psi_t = [1.000; 30.000]'$ . Isto leva a minimização inicializar em valores que acompanhem as medidas em detrimento do modelo.

O novo resultado é mostrado na tabela 7.4 abaixo:

Tabela 7.4: Resultados dos parâmetros  $\sigma_t^2$  e  $\eta_t^2$  calculados pelo Filtro de Kalman otimizado

t	Parâmetros $\sigma_t^2$ e $\eta_t^2$											
	5	6	7	8	9	10	11	12	13	14	15	16
$\sigma$	97,56	185,99	159,82	247,17	185,85	254,44	1.531,83	463,66	780,73	765,19	2.220,94	2.789,93
$\eta$	5.353,91	53.066,33	7.880,22	4.842,49	12.450,78	14.365,93	0,00	8.913,54	0,00	0,00	0,00	9181,42

As variâncias encontradas não são provavelmente as melhores, pois, o método de otimização não garante o mínimo global e sim o local. No entanto, este resultado já é o suficiente para realizar o cálculo da reserva de IBNR.

### 7.3.1 Normalidade dos Resíduos da Estimação

Como comentado no item 5.1.2, para que o ganho de Kalman seja ótimo é necessário que a distribuição dos resíduos  $v(t)$  das observações seja uma Normal.

Aplicando a estatística de Jarque-Bera para os resíduos de cada período de tempo  $t$ , estimados pelo Filtro de Kalman foi possível identificar quando o ganho de Kalman obtido foi ótimo.

### 7.3. Estimativa dos Parâmetros Calculados pelo Filtro de Kalman para os Dados SBR

Nas figuras 7.2 e 7.3 são mostrados os resultados da estatística de Jarque-Bera aplicada aos resíduos  $v_t$  das medidas obtidos pelo Filtro de Kalman. Estes serão importantes para um maior entendimento dos resultados obtidos e mostrados no item 7.3.2.

Figura 7.2: Estatística de Jarque-Bera para os dados SBR com  $5 \leq t \leq 10$

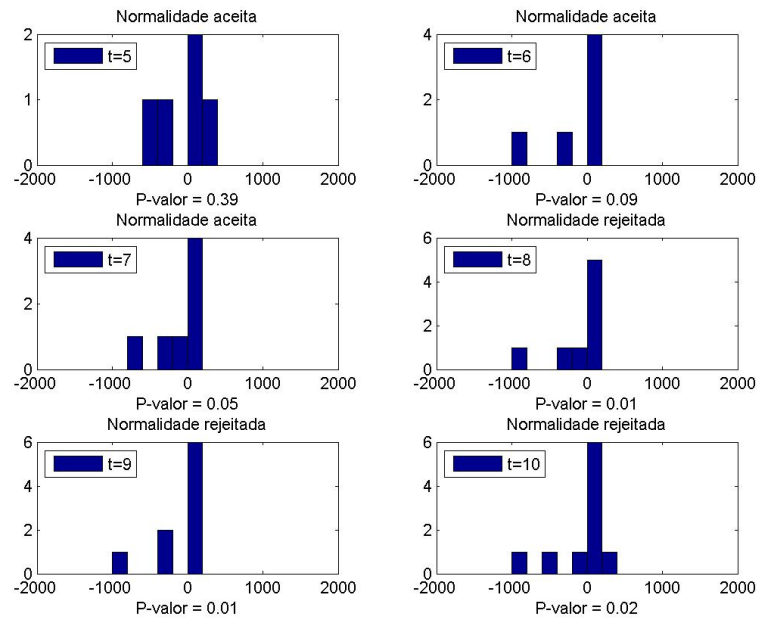
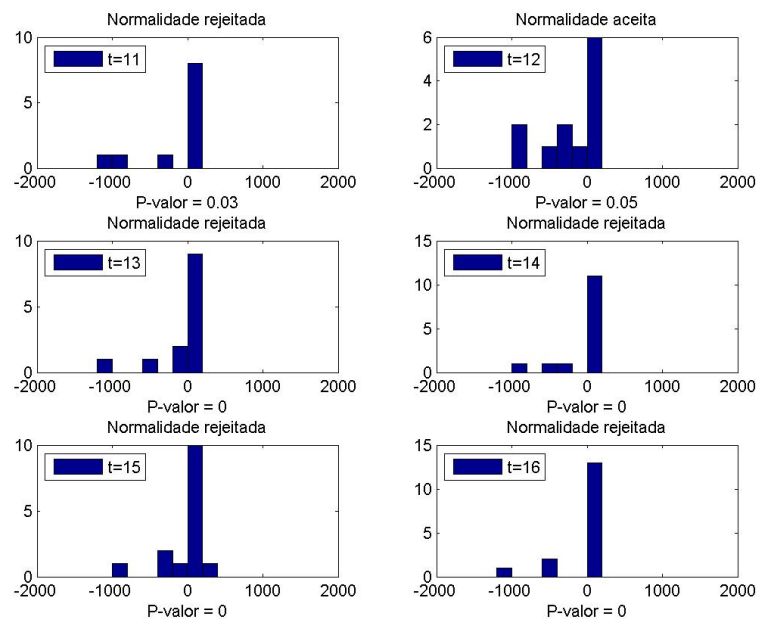


Figura 7.3: Estatística de Jarque-Bera para os dados SBR com  $11 \leq t \leq 16$





### 7.3.2 Resultados da provisão de IBNR calculada pelo Filtro de Kalman

Com os parâmetros  $\sigma_t^2$  e  $\eta_t^2$  calculados pelo Filtro de Kalman foi possível estimar a provisão de IBNR para os dados SBR.

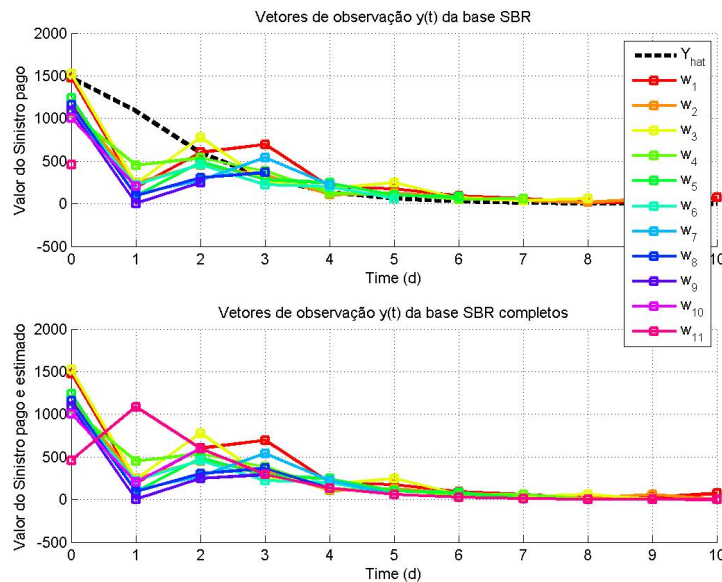
Como foram estimados os valores de provisão para cada um dos períodos disponíveis a partir do quinto mês observado, várias estimativas foram calculadas. Os valores de cada uma destas estimativas podem ser conferidos tanto para o método *Chain Ladder* quanto para o modelo de Espaço de Estados e Filtro de Kalman na tabelas 7.8.

A seguir, o comportamento dos sinistros observados e dos estimados é mostrado nas figuras de 7.4 a 7.9 para os períodos de onze a dezesseis meses.

#### Triângulo com onze meses observados

A figura 7.4 mostra o desenvolvimento dos sinistros observados, bem como suas estimativas calculadas, para o décimo primeiro mês, e é mostrada abaixo:

Figura 7.4: Sinistros da base SBR com onze meses e  $\hat{y}(t)$  estimado

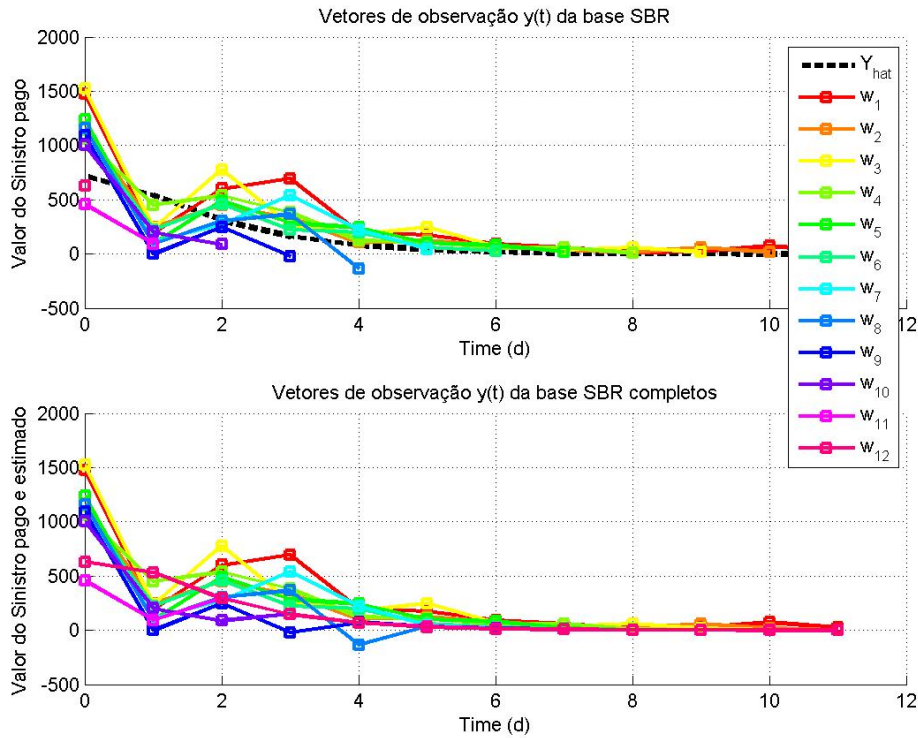


Lembrando que a hipótese de Normalidade foi rejeitada pela estatística de Jarque-Bera para a estimativa da provisão de IBNR para o Triângulo de *Run-off* com onze meses. Vale observar que a linha negra tracejada representando  $\hat{y}(11)$  calculado pelo Filtro de Kalman se distanciou bastante das medidas no primeiro período de atraso, ou ainda  $d = 1$ .

### Triângulo com doze meses observados

A figura 7.5 mostra o desenvolvimento dos sinistros observados, bem como suas estimativas calculadas, para o décimo segundo mês, e é mostrada abaixo:

Figura 7.5: Sinistros da base SBR com doze meses e  $\hat{y}(t)$  estimado

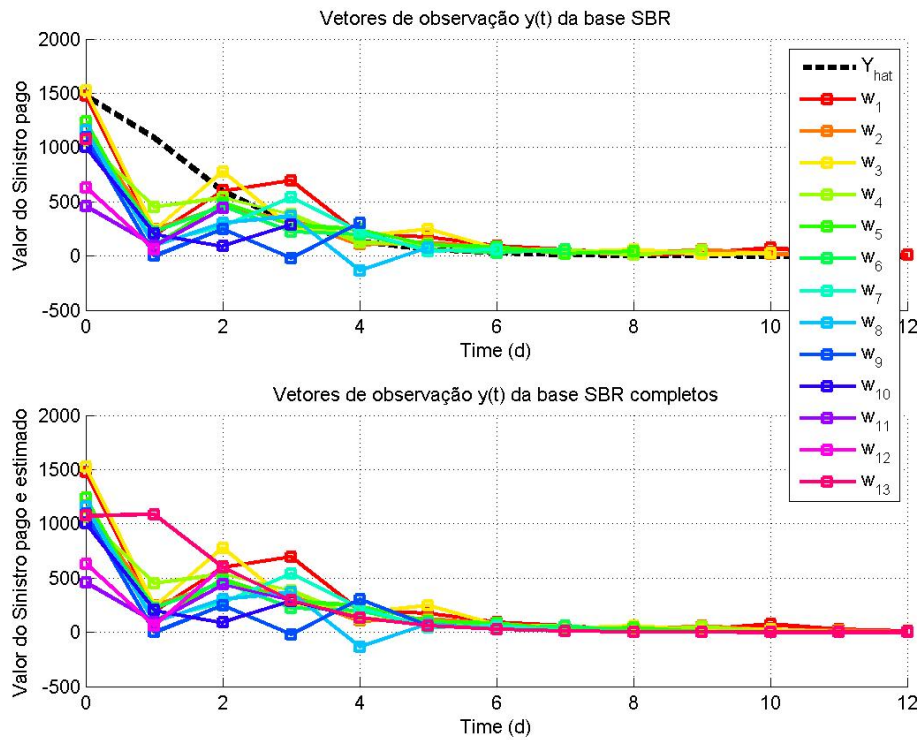


Para o Triângulo de *Run-off* com doze meses, a hipótese de Normalidade foi aceita pela estatística de Jarque-Bera para os resíduos das observações  $v(t)$  obtidos pelo Filtro de Kalman. Diferentemente do que ocorreu para a estimativa em onze meses,  $\hat{y}(12)$  não se distanciou tanto das medidas observadas para o primeiro período de atraso. Para o atraso zero, o valor estimado para  $\hat{y}_0(12)$  ficou muito próximo do último período observado. Isto também se opõe ao que ocorreu na estimação do Triângulo de *Run-off* para o período onze, como mostrado na figura 7.4.

### Triângulo com treze meses observados

A figura 7.6 mostra o desenvolvimento dos sinistros observados, bem como suas estimativas calculadas, para o décimo terceiro mês, e é mostrada abaixo:

Figura 7.6: Sinistros da base SBR com treze meses e  $\hat{y}(t)$  estimado

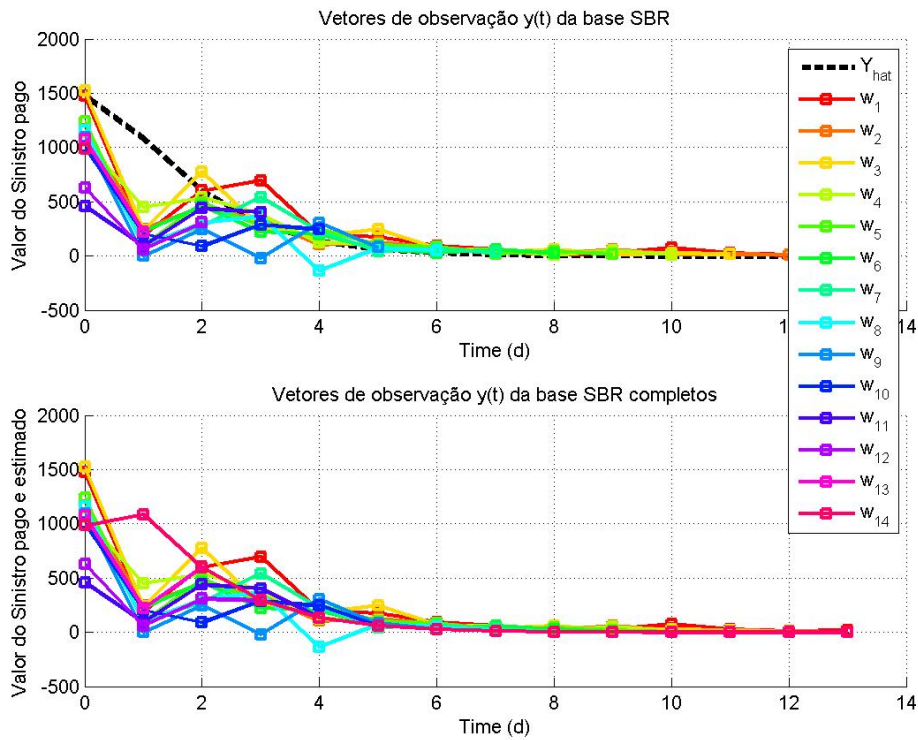


Assim como o Triângulo de *Run-off* com onze meses, a hipótese de Normalidade foi rejeitada pela estatística de Jarque-Bera para a estimativa da provisão de IBNR para o Triângulo de *Run-off* com treze meses. A linha negra tracejada, representando  $\hat{y}(13)$  calculado pelo Filtro de Kalman, distanciou-se bastante das medidas no primeiro período de atraso, ou ainda  $d = 1$ . Isto mostra uma evidência de que as estimativas descartaram as observações e seguiram o modelo determinístico  $\phi(t)$  explicado no item 6.2.2.

### Triângulo com quatorze meses observados

A figura 7.7 mostra o desenvolvimento dos sinistros observados, bem como suas estimativas calculadas, para o décimo quarto mês, e é mostrada abaixo:

Figura 7.7: Sinistros da base SBR com quatorze meses e  $\hat{y}(t)$  estimado

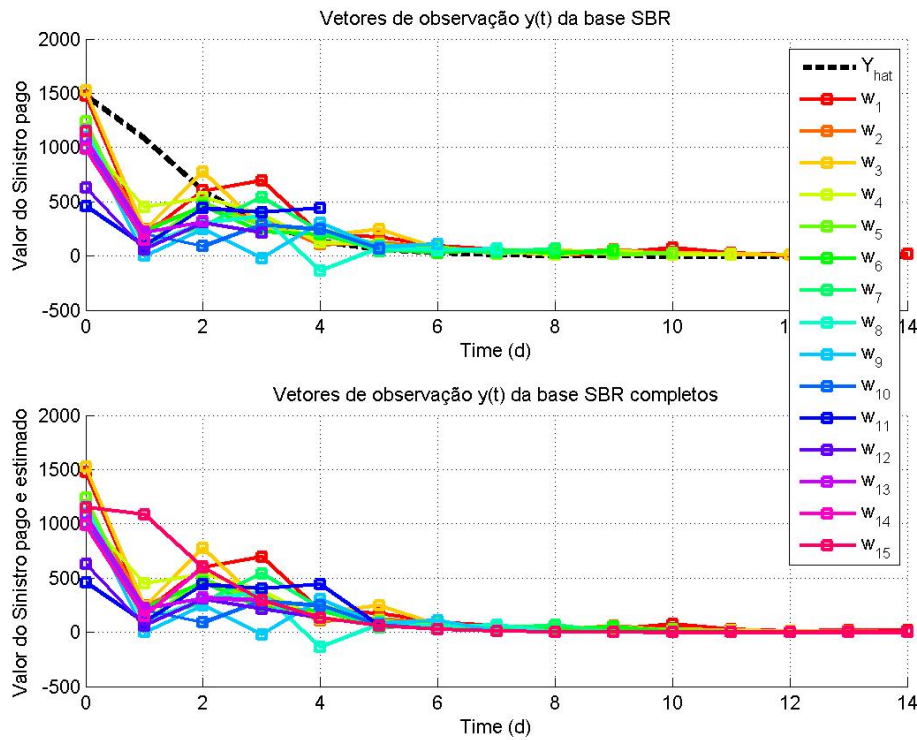


Para o Triângulo de *Run-off* com quatorze meses mostrado na figura 7.7, ocorreu o mesmo processo de estimação ocorrido para o Triângulo de *Run-off* com treze meses. A hipótese de Normalidade foi rejeitada pela estatística de Jarque-Bera para a estimativa da provisão de IBNR. E a linha negra tracejada, representando  $\hat{y}(14)$ , distanciou-se bastante das medidas no primeiro período de atraso, ou ainda  $d = 1$ .

### Triângulo com quinze meses observados

A figura 7.8 mostra o desenvolvimento dos sinistros observados, bem como suas estimativas calculadas, para o décimo quinto mês, e é mostrada abaixo:

Figura 7.8: Sinistros da base SBR com quinze meses e  $\hat{y}(t)$  estimado

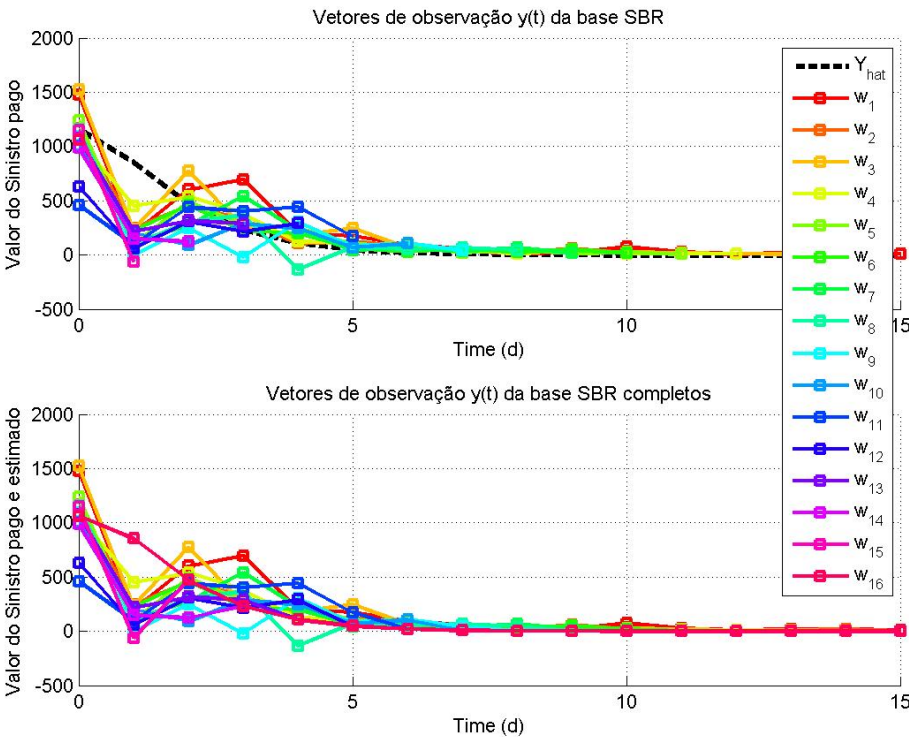


O mesmo processo de estimação ocorrido para o Triângulo de *Run-off* com treze meses e quatorze meses ocorreu para o Triângulo de *Run-off* com quinze meses mostrado na figura 7.8. Todos os três períodos tiveram a hipótese de Normalidade rejeitada pela estatística de Jarque-Bera para a estimativa da provisão de IBNR. E todos os três tiveram a estimativa de  $\hat{y}(t)$ , representado pela linha negra tracejada, distanciou-se bastante das medidas no primeiro período de atraso, ou ainda  $d = 1$ . Pela definição do Filtro de Kalman, estas estimativas não obtiveram o ganho ótimo de Kalman, mas sim um ganho ótimo linear.

**Triângulo com dezesseis meses observados**

Os resultados para dezesseis meses do Triângulo de *Run-off* são mostrados na figura 7.9 abaixo:

Figura 7.9: Sinistros da base SBR com dezesseis meses e  $\hat{y}(t)$  estimado



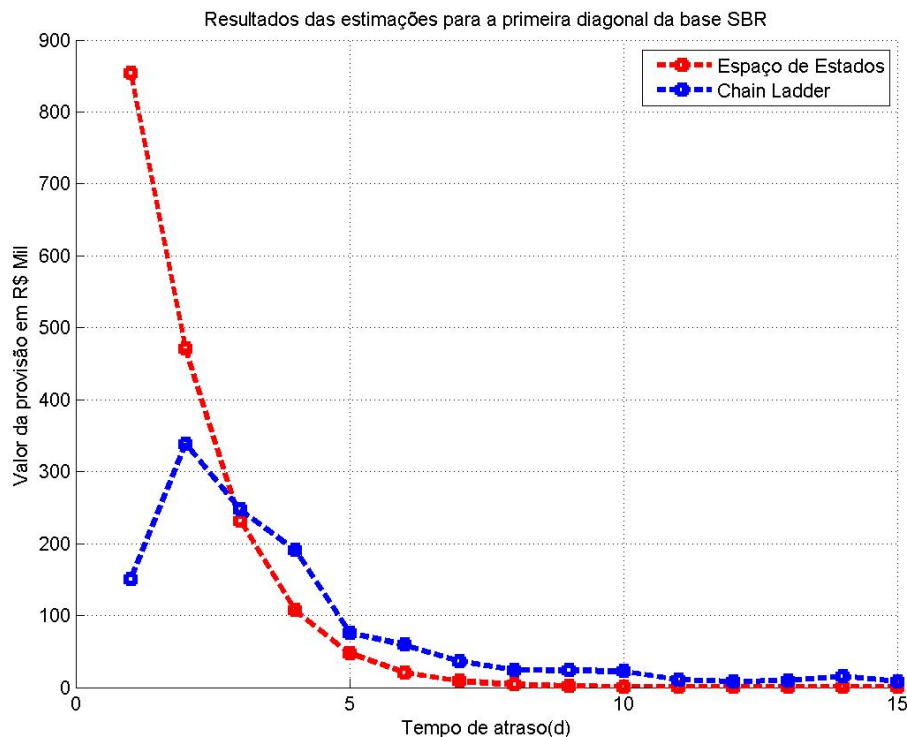
A tabela 7.5 mostra o total, bem como a soma da primeira diagonal, da reserva de IBNR estimada para o décimo sexto mês utilizando ambos os métodos *Chain Ladder* e o modelo de Espaço de Estados e Filtro de Kalman, e é mostrada abaixo:

Tabela 7.5: Resultados dos dados SBR

IBNR	Total	Primeira Diagonal
Espaço de Estados	3.387,95	1.784,16
Chain Ladder	4.648,65	1.222,01

A tabela B.1 representa o Triângulo de *Run-off* completo, contendo tanto os dados observados como os calculados, e é mostrada no apêndice B.

Figura 7.10: Primeira diagonal estimada para a base SBR com dezesesseis meses



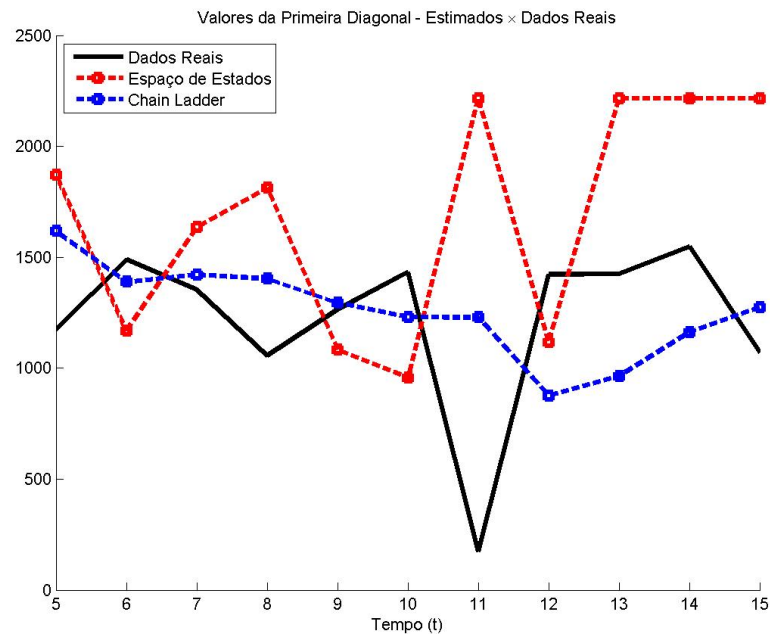
A figura 7.10 mostra o comportamento da primeira diagonal obtida nas estimções utilizando ambos os métodos *Chain Ladder* e o modelo de Espaço de Estados e Filtro de Kalman. É clara a diferença entre os resultados obtidos pelos dois métodos de estimação. Enquanto o método *Chain Ladder* simula um calombo no primeiro período de atraso, o modelo de Espaço de Estados e Filtro de Kalman segue a função determinística  $\phi(t)$  explicada no item 6.2.2 da página 45.

## 7.4 Erros de estimação

Como uma forma de definir qual o modelo mais adequado para o cálculo da provisão de IBNR, se o método *Chain Ladder* ou o modelo de Espaço de Estados e Filtro de Kalman especificado, o ideal seria comparar os dados estimados aos dados reais observados. Infelizmente a base de dados SBR não é grande o bastante para fornecer informação suficiente para estimar os dados do Triângulo de *Run-off* completo e ainda assim, realizar um *Backtest*<sup>1</sup> comparando os valores estimados e os observados em momentos futuros. No entanto, comparar o resultado da primeira diagonal calculada é possível para alguns passos de tempo.

O resultado desta comparação é mostrado em seguida na figura 7.11 e 7.12 além das tabelas 7.6 e 7.7. Os valores obtidos são mostrados tanto em números que representam a soma da diagonal encontrada em cada um dos métodos, além do valor real, como na forma de erro percentual, em relação ao valor real observado no período seguinte.

Figura 7.11: Valores da primeira diagonal estimados para os dados SBR - R\$/Mil



A tabela 7.6 com os valores calculados para a primeira diagonal estimada pelo método *Chain Ladder* e pelo modelo de Espaço de Estados e Filtro de Kalman é mostrada em seguida:

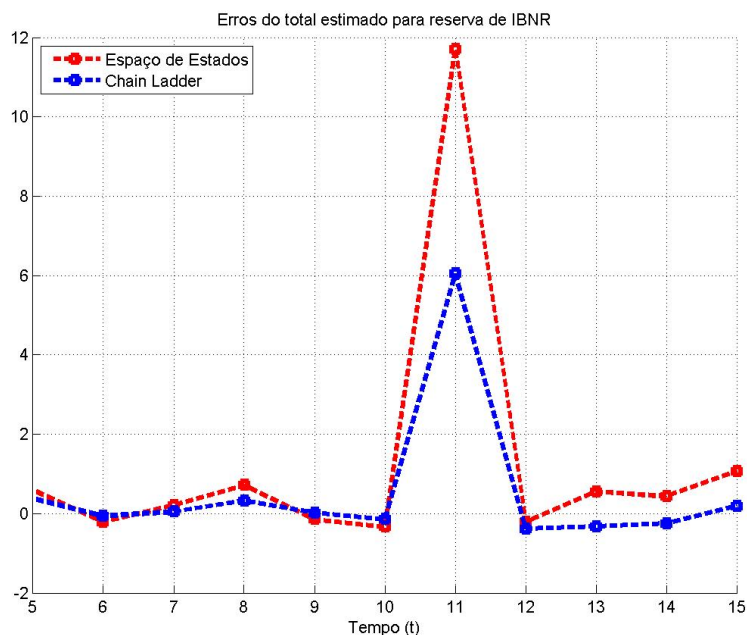
<sup>1</sup>*Backtest* é um termo largamente empregado na Estatística e representa o teste usual aplicado aos modelos especificados utilizando dados reais, como uma forma de validá-los.



Tabela 7.6: Valores da primeira diagonal estimados para os dados SBR -  $R\$/Mil$ 

Primeira diagonal para os dados SBR em Milhares de Reais											
t	5	6	7	8	9	10	11	12	13	14	15
IBNR Real diag1	1173,27	1488,96	1353,09	1056,63	1264,69	1433,44	174,40	1423,02	1424,16	1547,81	1071,70
IBNR SST diag1	1869,87	1168,71	1635,29	1812,74	1081,49	956,34	2216,41	1117,75	2216,82	2216,87	2216,89
IBNR CL diag1	1617,45	1389,78	1420,81	1403,06	1293,38	1229,61	1229,26	874,99	963,89	1162,81	1275,09

Figura 7.12: Erros percentuais da primeira diagonal estimada para os dados SBR



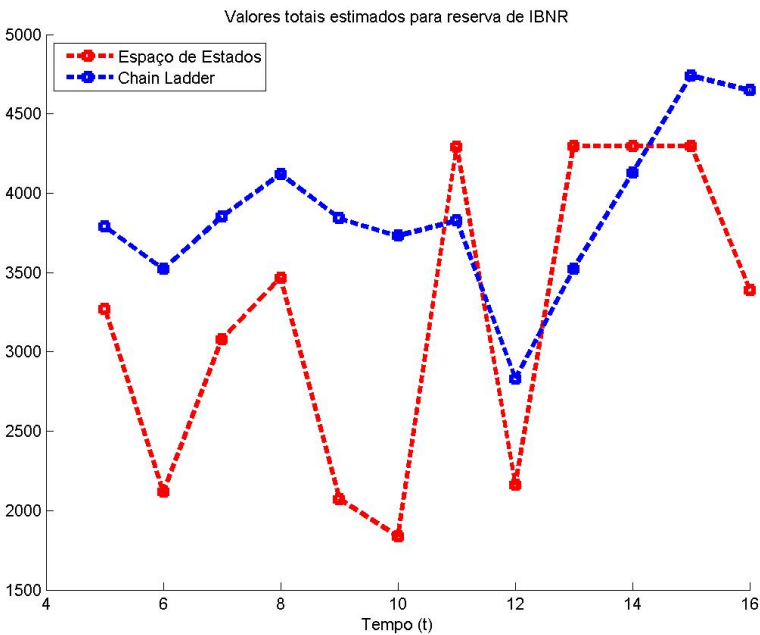
A tabela 7.7 com os erros percentuais referentes aos valores calculados para a primeira diagonal estimada pelo método *Chain Ladder* e pelo modelo de Espaço de Estados e Filtro de Kalman e os dados reais observados no próximo período, é mostrada abaixo:

Tabela 7.7: Erros percentuais da primeira diagonal estimada para os dados SBR

Erros percentuais da primeira diagonal para os dados SBR											
t	5	6	7	8	9	10	11	12	13	14	15
Erro SST	59%	-22%	21%	72%	-14%	-33%	1.171%	-21%	56%	43%	107%
Erro CL	38%	-7%	5%	33%	2%	-14%	605%	-39%	-32%	-25%	19%

No décimo primeiro mês há uma clara ruptura no valor do sinistro observado. Isto causa um grande erro na estimativa, tanto do método *Chain Ladder* quanto do modelo de Espaço de Estados e Filtro de Kalman.

Figura 7.13: Valores totais estimados para os dados SBR -  $R\$/Mil$



A tabela 7.8 com os valores totais estimados para a reserva de IBNR mostrados na figura 7.13 é dada abaixo:

Tabela 7.8: Valores totais estimados para os dados SBR -  $R\$/Mil$

Estimativa de IBNR para os dados SBR em Milhares de Reais												
t	5	6	7	8	9	10	11	12	13	14	15	16
IBNR SST	3269,13	2116,04	3079,47	3467,00	2071,86	1837,76	4291,86	2159,78	4296,53	4297,14	4297,39	3387,95
IBNR CL	3787,68	3520,05	3853,26	4120,38	3840,24	3730,66	3826,81	2827,49	3523,83	4128,24	4739,06	4648,65

O modelo de Espaço de Estados e Filtro de Kalman retorna uma estimativa da reserva de IBNR que oscila muito ao longo do tempo. Isto é uma desvantagem muito grande para modelos de IBNR, uma vez que estes valores devem ser contabilizados e a oscilação pode causar um impacto significativo na decisão de qual método a empresa seguradora irá utilizar.

# Capítulo 8

## Conclusões e Futuros trabalhos

### 8.1 Conclusões

Neste trabalho buscou-se calcular a reserva de provisão dos sinistros do tipo IBNR de uma grande seguradora brasileira, aqui chamados de "**dados SBR**". Testando desta forma, a eficiência do modelo de Espaço de Estados proposto por De Jong & Zenwirth com dados não necessariamente bem comportados, como é o caso dos dados SBR.

Com o objetivo de ilustrar a aplicação da metodologia apresentada, primeiro foi calculada a reserva de provisão dos sinistros IBNR para os dados DJZ fornecidos por De Jong & Zenwirth[12]. Para o cálculo, foi utilizado um modelo de Espaço de Estados e Filtro de Kalman e em contrapartida, o método *Chain Ladder*. Os dados DJZ são estáveis e não apresentam valores negativos, o que os tornam particularmente atraentes para comparar diversos métodos, tornando o trabalho mais didático.

A estimativa das reservas de IBNR, que foram explicadas no capítulo 3, apresentam peculiaridades que possibilitam uma grande gama de métodos a serem utilizados. A prática atuarial possibilita que o método mais adequado seja escolhido para cada conjunto de dados.

Conhecer o comportamento da informação disponível é útil na hora de organizar os dados na forma de Espaço de Estados. De Jong & Zenwirth já tiveram esta preocupação em seu trabalho[12], no entanto, cada caso pode merecer um tratamento especial para obter um melhor modelo. Se o problema não for bem colocado, o modelo não irá descrever o que os sinistros observados mostram, fazendo com que o filtro fique instável.

Este provavelmente foi o ponto que ocorreu com os dados SBR. Ao ocorrerem mudanças nas novas informações, o algoritmo não foi capaz de estimar as variâncias do vetor  $\psi_t$ . O que levou a obtenção de somente mínimos locais, o que tornaram o modelo de Espaço de Estados não Gaussiano. Isto faz com que o ganho de Kalman não seja ótimo e o modelo se resuma a uma regressão linear.

A baixa quantidade de dados dificulta na hora de concluir qual o melhor modelo esco-

lher. Todavia, o Filtro de Kalman não se mostrou estável o bastante para ser considerado um método melhor que o *Chain Ladder*, apesar dos resultados terem apresentado valores próximos aos reais na maior parte dos períodos estimados.

Outro ponto relevante se refere ao fato do Filtro de Kalman ser muito sensível às parametrizações. Isto o torna uma ferramenta instável, caso o modelo de Espaço de Estados não seja corretamente parametrizado. Ponto de atenção, dado que na literatura consultada, não há unanimidade quanto aos parâmetros de inicialização do filtro que devam ser utilizados. Qualquer parametrização pode ser adotada o que torna o método difícil de ser utilizado para uma seguradora que possua uma carteira de muitos produtos de seguros que precisem ser calculados separadamente. O modelo utilizado mostrou-se mais complexo que o cálculo do IBNR pelo método *Chain Ladder* detalhado por Mack [25] aceito como a solução básica para o cálculo de IBNR pelas áreas atuariais. Isto dificulta que empresas seguradoras venha a utilizá-lo no futuro.

## 8.2 Extensões

Neste trabalho foi possível utilizar o modelo de Espaço de Estados e Filtro de Kalman proposto por De Jong & Zenwirth para os dados de uma seguradora brasileira. No entanto, o modelo apresentou instabilidade e algumas sugestões de trabalhos futuros que corrijam os pontos discutidos são apresentados a seguir:

- 1 Testar outro método numérico para maximizar a verossimilhança em contrapartida ao BFGS para encontrar o vetor de parâmetros  $\psi_t$ ;
- 2 Buscar técnicas alternativas à maximização da verossimilhança como de Monte Carlo, para encontrar o vetor de parâmetros  $\psi_t$ ;
- 3 Modificar a função do comportamento da matriz  $X_t$  proposta por De Jong & Zenwirth, como explicado no item 6.2.2 da página 45;
- 4 Realizar a estimativa separadamente para os sinistros do tipo IBNR e IBNER como o trabalho de Liu & Verrall [24].

Os itens acima focam, principalmente, em mudanças no método para maximizar a verossimilhança, pois este foi o ponto que causou maiores problemas nas estimativas. Outras extensões também são sugeridas, como aumentar os *lags* na matriz de parâmetros  $H_t$ , e ainda comparar a abordagem utilizado o Filtro de Kalman como sugerido no trabalho de De Jong & Zenwirth [12], incluindo os índices de inflação e índices de volume de vendas, com o método Bornhuetter-Ferguson[6].

# Referências Bibliográficas

- [1] ATHERINO, R. S. *Um modelo em espaço de estado para estimativa de ibnr*. March 2005. Dissertação (Mestrado em Engenharia Elétrica) - Puc Rio, March 2005.
- [2] ATHERINO, R. S.; FERNANDES, C. A. C. Um modelo em espaço de estado para estimação de reservas ibnr. *Revista Brasileira de Risco e Seguro*, 2007.
- [3] ATHERINO, R. S. *Estimação de reservas ibnr por modelos em espaço de estado*. December 2008. Tese (Doutorado em Engenharia Elétrica) - Puc Rio, December 2008.
- [4] BEARD, R. E. General insurance. by b. benjamin. [heinemann, london 1977]. *Journal of the Institute of Actuaries*, v. 105, p. 107–109, 5 1978.
- [5] BIANCO, A. F. *Filtro de kalman robustos para sistemas dinâmicos singulares em tempo discreto*. 2009. Dissertação (Mestrado em Engenharia Elétrica) - 2009.
- [6] BORNHUETTER, R. L.; FERGUSON, R. E. The actuary and ibnr. *Proceedings of the Casualty Actuarial Society*, v. 59, p. 181–195, 1972.
- [7] BROWN, R. G.; HWANG, P. Y. *Introduction to random signals and applied kalman filtering*. 4th ed.. ed. Wiley, 2012.
- [8] BUENO, R. L. S. *Econometria de séries temporais*. 2<sup>a</sup> ed.. ed. Cengage Learning, 2011.
- [9] BUSSAB, W. O.; MORETTIN, P. *Estatística básica*. 2<sup>a</sup> ed.. ed. Saraiva, 2002.

- [10] CHRISTENSEN, T. M.; HURN, S.; LINDSAY, K. A. The devil is in the detail: Hints for practical optimisation. , n. 32, 2008.
- [11] CONT, R. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, v. 1, p. 223–236, 2001.
- [12] DE JONG, P.; ZEHNWIRTH, B. Claims reserving, state-space models and the kalman filter. *Journal of the Institute of Actuaries*, , n. 110, p. 157–181, 1983.
- [13] DE JONG, P. Forecasting runoff triangles. *North American Actuarial Journal*, , n. 10, p. 28–38, 2006.
- [14] DINO, J. Nasa ames research center, moffett field, calif., history related to the apollo moon program and lunar prospector mission, consultado em jul 2013.
- [15] DURBIN, J.; KOOPMAN, S. J. Time series analysis by state space methods. 2001.
- [16] ENGLAND, P.; VERRALL, R. Stochastic claims reserving in general insurance. *British Actuarial Journal*, v. 8, p. 443–518, 8 2002.
- [17] FENASEG. Projeção dos resultados do mercado de seguros no brasil para 2012 à 2015, jul 2013.
- [18] FINGER, R. J. Modeling loss reserve developments. *Proceedings of the Casualty Actuarial Society*, v. 63, November 1976.
- [19] HARVEY, A. *Forecasting structural time series models and the kalman filter*. Cambridge University Press, 1989.
- [20] HAYKIN, S. *Sistemas de comunicação*. Bookman, 2004.
- [21] KALMAN, R. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, v. 82, p. 35–45, 1960.
- [22] KALMAN, R. E.; BUCY, R. S. New results in linear filtering and prediction theory. *Transactions of the ASME. Series D, Journal of Basic Engineering*, v. 83, p. 95–107, 1961.

- [23] LEVINSON, N.; NOHEL, J.; SATTINGER, D. *Selected of norman levinson*. Contemporary Mathematicians. Birkhäuser Boston, 1998.
- [24] LIU, H.; VERRALL, R. Predictive distributions for reserves which separate. true ibnr and ibner claims. *Astin Bulletin*, v. 39, p. 35–60, 2009.
- [25] MACK, T. Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, v. 23, p. 213–225, 1993.
- [26] MACK, T. The standard error of chain ladder reserve estimates: recursive calculation and inclusion of a tail factor. *ASTIN Bulletin*, v. 29, p. 361–366, 1999.
- [27] MENDENHALL, W.; WACKERLY, D.; SCHEAFFER, R. *Mathematical statistics with applications*. Thomson Higher Education, 2008.
- [28] OLIVEIRA, A. D. *Controle Ótimo de sistemas lineares com saltos markovianos e ruídos multiplicativos sob o critério de média variância ao longo do tempo*. 2011. Tese (Doutorado em Engenharia Elétrica) - Universidade de São Paulo, 2011.
- [29] PIZZINGA, A. H. *Modelos em espaço de estado com restrições nas componentes de interesse*. March 2004. Dissertação (Mestrado em Engenharia Elétrica) - Puc Rio, March 2004.
- [30] RENSHAW, A. E.; VERRALL, R. J. A stochastic model underlying the chain-ladder technique. *BAJ*, p. 903–923, 1998.
- [31] ROSS, S. *Probabilidade - um curso moderno com aplicações*. 8 edição. ed. Bookman, 2010.
- [32] SCHMIDT, K. D.; ZOCHER, M. The bornhuetter-ferguson principle. *Variance Journal*, v. 2, p. 85–110, 2008.
- [33] SCHMIDT, S. F. Applications of state - space methods to navigation problems, in advances in control systems. *Academic Press*, v. 3, p. 293–340, 1966.
- [34] SUSEP. Orientação sobre o cálculo da margem de solvência. 1989.

- [35] SUSEP. Dispõe sobre a provisão para sinistros ocorridos e não avisados - ibnr. 1998.
- [36] TARBEL, T. F. Incurred but not reported claim reserves. *Proceedings of the Casualty Actuarial Society*, v. XX, p. 275–280, 1934.
- [37] WIENER, N. *The extrapolation, interpolation and smoothing of stationary time series*. 1949.
- [38] ZADEH, L. A.; RAGAZZINI, J. R. An extension of wiener's theory of prediction. *Journal of Applied Physics*, New York, v. 21, p. 645–655, 1950.



## Apêndice A

# Resultados para o método *Chain Ladder* para os dados SBR

A tabela A.1 representa o Triângulo de *Run-off* completo, contendo tanto os dados observados como os estimados pelo método *Chain Ladder*.

Pode-se comparar cada um dos termos calculados, com os estimados pelo modelo de Espaço de Estados e Filtro de Kalman mostrados na tabela B.1 do apêndice B.

Tabela A.1: Dados SBR - Reserva estimada pelo método *Chain Ladder* - R\$/Mil

Triângulo de <i>Run-off</i> completo para os dados SBR em Milhares de Reais																
w/d	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1475,33	190,11	600,63	695,47	200,14	174,43	90,74	59,25	17,10	21,25	71,77	21,56	8,95	17,56	14,11	11,49
2	1164,67	247,87	448,87	333,44	103,16	125,80	73,96	38,40	15,11	61,43	15,39	15,51	11,05	8,31	15,00	8,41
3	1522,17	236,47	775,35	264,99	179,16	245,78	64,97	35,40	60,30	16,84	22,99	14,96	11,26	5,69	15,95	10,90
4	1059,67	451,22	536,37	381,75	121,02	109,16	50,91	61,57	8,02	49,74	11,33	10,21	9,51	9,28	13,25	9,05
5	1240,37	94,31	487,93	282,13	244,65	102,22	79,85	20,71	43,81	13,37	23,24	6,25	8,57	8,58	12,26	8,38
6	1081,75	223,58	458,64	223,00	198,59	60,67	28,24	59,96	22,77	49,21	13,98	10,96	7,90	7,91	11,30	7,72
7	1092,19	112,31	270,58	538,30	223,10	42,23	60,60	31,91	67,14	14,98	22,40	11,21	8,08	8,09	11,55	7,89
8	1160,82	88,77	304,95	363,73	-136,66	71,78	52,22	69,52	43,24	23,35	18,64	9,33	6,72	6,73	9,61	6,57
9	1092,32	-2,03	247,96	-19,84	304,08	78,93	107,91	38,98	24,02	21,66	17,29	8,66	6,24	6,25	8,92	6,09
10	1003,94	203,07	87,47	286,94	243,64	66,61	104,64	36,42	26,42	23,82	19,02	9,52	6,86	6,87	9,81	6,70
11	457,21	90,44	438,93	401,16	443,37	170,54	59,39	37,61	27,28	24,59	19,63	9,83	7,08	7,09	10,13	6,92
12	630,17	59,72	308,31	213,21	291,67	75,59	46,84	29,66	21,51	19,39	15,48	7,75	5,58	5,59	7,99	5,46
13	1071,93	216,93	315,03	287,74	191,16	104,74	64,91	41,10	29,81	26,87	21,46	10,74	7,74	7,75	11,07	7,56
14	983,01	148,67	121,05	247,22	151,58	83,05	51,47	32,59	23,64	21,31	17,01	8,52	6,14	6,15	8,78	6,00
15	1152,92	-63,07	338,42	281,86	172,82	94,69	58,68	37,15	26,95	24,29	19,40	9,71	7,00	7,01	10,01	6,84
16	1062,68	150,88	376,83	313,85	192,43	105,44	65,34	41,37	30,01	27,05	21,60	10,81	7,79	7,80	11,14	7,61

## Apêndice B

# Resultados para o modelo de Espaço de Estados e Filtro de Kalman para os dados SBR

A tabela B.1 representa o Triângulo de *Run-off* completo, contendo tanto os dados observados como os estimados pelo modelo de Espaço de Estados e Filtro de Kalman.

Pode-se comparar cada um dos termos calculados, com os estimados pelo método *Chain Ladder* mostrados na tabela A.1 do apêndice A.

Tabela B.1: Dados SBR - Reserva estimada pelo modelo de Espaço de Estados e Filtro de Kalman  $R\$/Mil$

Triângulo de <i>Run-off</i> completo para os dados SBR em Milhares de Reais																
w/d	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1475,33	190,11	600,63	695,47	200,14	174,43	90,74	59,25	17,10	21,25	71,77	21,56	8,95	17,56	14,11	11,49
2	1164,67	247,87	448,87	333,44	103,16	125,80	73,96	38,40	15,11	61,43	15,39	15,51	11,05	8,31	15,00	0,01
3	1522,17	236,47	775,35	264,99	179,16	245,78	64,97	35,40	60,30	16,84	22,99	14,96	11,26	5,69	0,02	0,01
4	1059,67	451,22	536,37	381,75	121,02	109,16	50,91	61,57	8,02	49,74	11,33	10,21	9,51	0,04	0,02	0,01
5	1240,37	94,31	487,93	282,13	244,65	102,22	79,85	20,71	43,81	13,37	23,24	6,25	0,11	0,04	0,02	0,01
6	1081,75	223,58	458,64	223,00	198,59	60,67	28,24	59,96	22,77	49,21	13,98	0,26	0,10	0,04	0,02	0,01
7	1092,19	112,31	270,58	538,30	223,10	42,23	60,60	31,91	67,14	14,98	0,64	0,26	0,10	0,04	0,02	0,01
8	1160,82	88,77	304,95	363,73	-136,66	71,78	52,22	69,52	43,24	1,55	0,63	0,25	0,10	0,04	0,02	0,01
9	1092,32	-2,03	247,96	-19,84	304,08	78,93	107,91	38,98	3,73	1,53	0,62	0,25	0,10	0,04	0,02	0,01
10	1003,94	203,07	87,47	286,94	243,64	66,61	104,64	8,87	3,67	1,50	0,61	0,24	0,10	0,04	0,02	0,01
11	457,21	90,44	438,93	401,16	443,37	170,54	20,79	8,74	3,62	1,48	0,60	0,24	0,10	0,04	0,01	0,01
12	630,17	59,72	308,31	213,21	291,67	47,85	20,53	8,63	3,57	1,46	0,59	0,24	0,09	0,04	0,01	0,01
13	1071,93	216,93	315,03	287,74	107,39	47,41	20,35	8,55	3,54	1,45	0,59	0,24	0,09	0,04	0,01	0,01
14	983,01	148,67	121,05	231,98	106,68	47,09	20,21	8,50	3,52	1,44	0,58	0,23	0,09	0,04	0,01	0,01
15	1152,92	-63,07	471,35	231,20	106,32	46,93	20,14	8,47	3,51	1,43	0,58	0,23	0,09	0,04	0,01	0,01
16	1062,68	853,56	471,01	231,03	106,24	46,90	20,13	8,46	3,50	1,43	0,58	0,23	0,09	0,04	0,01	0,01