Modelo de Espaço de Estados com Efeitos Retrospectivos para Dados Longitudinais Multivariados

Antonieta D'Alcântara de Queiroz Peres

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO GRAU DE DOUTOR
EM ESTATÍSTICA

Área de concentração: **Estatística**

Orientadora: Prof.ª **Dr.ª Denise Aparecida Botter**

São Paulo, junho de 2000

Modelo de Espaço de Estados com Efeitos Retrospectivos para Dados Longitudinais Multivariados

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por Antonieta D'Alcântara de Queiroz Peres e aprovada pela comissão julgadora

São Paulo, junho de 2000

Banca examinadora:

- Profa. Dra. Denise Aparecida Botter (orientadora) (IME-USP)
- Prof. Dr. Bent Jørgensen (Odense University)
- Prof. Dr. Julio da Motta Singer (IME-USP)
- Prof. Dr. Antonio Carlos Ponce de Leon (UERJ)
- Prof. Dr. Gauss Moutinho Cordeiro (UFBA)

À minha mãe.

Agradecimentos

Ao CNPq, pelo apoio institucional por meio dos processos 140.173/95-1 e 201773/88-0;

Aos colegas da Universidade Rural, Antonio, Márcia, Rosane, Tise e Valdomiro, por terem acreditado neste projeto, lutando (literalmente) junto comigo pela minha liberação para vir cursar o doutorado;

À Profa. Dra. Carmem Silvia Cardassi (*In memorian*), com quem gostaria de ter aprendido a lecionar, pelo muito que me ensinou;

À Profa. Dra. Denise Aparecida Botter, pela orientação, pelo incentivo e apoio durante todas as fases deste trabalho;

Ao Prof. Dr. Júlio da Motta Singer, pelo incentivo e por ter lido este trabalho, contribuindo com inúmeras sugestões;

Aos colegas Chico e Cardoso, pela amizade, companheirismo, trocas de idéias e "patos no tucupi";

À colega Eliana, pelo companheirismo e amizade de todas as horas;

Ao Departamento de Estatística Teórica da Universidade de Aarhus, por ter me proporcionado a oportunidade de desenvolver parte deste trabalho na Dinamarca;

Ao Prof. Dr. Bent Jørgensen, pelo apoio, pela oportunidade que me ofereceu de me orientar neste trabalho e pela paciência com que conduziu esta orientação. Ao amigo Bent, pela carinhosa recepção na Dinamarca;

À Helena, Silvia e Vera, pela amizade e carinho, pelos inúmeros papos e jantares que fizeram com que o tempo que passei em Aarhus fosse muito mais agradável;

À Nedir, pelo incentivo, apoio e amizade;

Finalmente, agradeço ao Agenor, pelo carinho e incentivo, pelo sacrifício de tantos finais de semana e, principalmente, por sua imensa paciência.

Resumo

Neste trabalho, apresentamos um modelo de espaço de estados para dados longitudinais, no qual o processo latente é um processo de Markov de ordem $L \geq 1$. Assumimos que, dado o processo latente, a distribuição condicional do processo observado depende do valor presente e de $K \geq 0$ valores passados do processo latente e tanto o processo observado quanto o processo latente têm distribuição na família Tweedie dos modelos de dispersão. O modelo é log-linear em um conjunto de covariáveis que podem variar no tempo e a estimação dos parâmetros de regressão é feita por meio de equações de estimação de Kalman, resolvidas por um algoritmo de escores de Newton. Parâmetros de dispersão são estimados por meio de estimadores de Pearson corrigidos. O modelo admite ainda uma terceira classe de parâmetros, os parâmetros retrospectivos, que definem a ordem do processo latente considerado e a forma como o processo observado está ligado ao processo latente. Para estimação destes parâmetros retrospectivos, propomos um procedimento inspirado na função de verossimilhança perfilada. Mostramos alguns resultados de simulação e uma aplicação do modelo a um conjunto de dados reais.

Abstract

In this work, we consider a state space model for longitudinal data, in which the latent process is an L-order Markov process. We assume that, given the latent process, the conditional distribution of observations depends on the present and on the last K past values of the latent process. We also assume that both observed and latent processes follow distributions on the Tweedie family of dispersion models. The model is log-linear on covariates that may vary on time. The estimation of regression parameter is based on Kalman estimation functions and for dispersion parameters, we use adjusted Pearson estimates. We also define a third class of parameters, called retrospective parameters, that define the order of the latent process as well as the way the observed process is linked to the latent process. For estimating the retrospective parameters, we propose a procedure based on the profile likelihood function. We show some results from simulations and an application of the model to a real data set.

Índice

1 Introdução **1**

1.1 Equações de estimação generalizadas 4

1.2 Modelos de espaço de estados 7

2 Preditores lineares e o filtro de Kalman **10**

2.1 Preditores lineares 10

2.2 Filtro e suavizador de Kalman 18

3 Modelo Tweedie-Tweedie **26**

3.1 Modelo 28

3.2 Momentos 31

3.3 Filtro e suavizador de Kalman 32

3.4 Estimação dos parâmetros 34

3.4.1 Parâmetros de regressão 34

3.4.2 Parâmetros de dispersão 37

3.5 Análise de resíduos 37

3.6 Poluição e mortalidade na cidade de São Paulo 40

3.6.1 Introdução 40

3.6.2 Os dados 41

3.6.3 Análise 43

4 Modelo retrospectivo 56
4.1 Introdução 56
4.2 Modelo 57
4.2.1 Definição 58
4.2.2 Propriedades 59
4.3 Momentos 61
4.4 O filtro e o suavizador de Kalman 65
4.4.1 Modelo de espaço de estados 66
4.5 Estimação dos parâmetros 68
4.5.1 Parâmetros de regressão 71
4.5.2 Parâmetros de dispersão 77
4.5.3 Parâmetros retrospectivos 78
4.6 Resíduos no modelo retrospectivo 81
4.7 Simulações 82
4.8 Outros métodos de estimação 91
4.9 Transferência entre tarefas sincronizatórias 93

5 Conclusões e estudos futuros 103

A Família Tweedie 105

B Cálculos 109
B.1 Esperança condicional 109
B.2 Resíduos 110
B.3 Derivadas do filtro e suavizador de Kalman 111
B.4 Matriz de variabilidade 113

C Gráficos 115

Índice de figuras

2.1 Estrutura de correlação de um modelo de espaço de estados 20

3.1 Simulações de 100 observações de um processo com 3 categorias, com $q = 2$, $p_i = 0, 1, 2$, $\nu_i^2 = 1, \sigma^2 = 0.05$ e $\theta_0 = 2$, sem efeito de covariáveis. As observações estão representadas por linhas sólidas e o processo latente, por linhas pontilhadas. 30

3.2 Evolução do número de óbitos de pessoas com mais de 65 anos de idade, da temperatura e da umidade na cidade de São Paulo entre 1 de janeiro de 1991 a 31 de dezembro de 1993. 41

3.3 Evolução dos níveis de poluentes atmosféricos no período de 1 de janeiro de 1991 a 31 de dezembro de 1993 na cidade de São Paulo. As linhas pontilhadas marcam o nível secundário de qualidade do ar. 42

3.4 Valores observados e valores ajustados, valores observados e valores preditos e processo latente estimado para o modelo 1. 44

3.5 Gráficos de resíduos do modelo 1. 46

3.6 Valores observados e estimados, valores observados e preditos e processo latente estimado para o modelo 2. 47

3.7 Gráficos de resíduos para o modelo 2. 50

3.8 Gráficos de resíduos para o modelo 4. 51

3.9 Valores observados e estimados, valores observados e preditos e processo latente estimado para o modelo final. 53

3.10 Gráficos de resíduos para o modelo final. 55

4.1 Estrutura de correlação do modelo retrospectivo com $K = 0$ e $L = 1$ (modelo Tweedie–Tweedie). 57
4.2 Estrutura de correlação do modelo retrospectivo com $K = 1$ e $L = 1$ e um processo latente univariado. 57
4.3 Estrutura de correlação do modelo retrospectivo com $K = 1$, $L = 1$ e um processo latente bivariado. 61
4.4 Gráficos de $\mathcal{L}_1$ para amostras simuladas com $\psi = (0.0, 1.0)^\top$ (a, b) e com $\psi = (1.0, 0.0)^\top$ (c, d). A linha vertical marca o valor de ψ_2 utilizado na simulação da amostra. 86
4.5 Gráficos de $\mathcal{L}_1$ em função de ψ_2 para amostras simuladas com $\psi = (0.3, 0.7)^\top$ 87
4.6 Gráficos de $\mathcal{L}_1$ em função de ψ_2 para amostras simuladas com $\psi = (0.7, 0.3)^\top$ 87
4.7 Valores observados para o indivíduo 26 (erro variável nas tentativas). 95
4.8 Gráficos de $\mathcal{L}_1$ (a) e $\hat{\sigma}^2$ (b) em função de ψ_2, considerando resposta normal e o modelo 1. 96
4.9 Funções de autocorrelação e autocorrelação parcial dos resíduos obtidos com ajustes do modelo 1 para $\psi_2 = 1.0$ (a, c) e 0.3 (b, d), considerando resposta normal. 97
4.10 Gráficos de $\mathcal{L}_1$ em função de ψ_2, considerando resposta normal e o modelo 1 com $\sigma^2 = 0.0124$ em todos os ajustes. 97
4.11 Funções de autocorrelação e autocorrelação parcial dos resíduos obtidos com ajustes do modelo 1 com $\sigma^2 = 0.0124$ fixo, para $\psi_2 = 1.0$ (lado esquerdo), 0.3 (lado direito), considerando resposta normal. 98
4.12 Gráficos de $\mathcal{L}_1$ (a) e de $\hat{\sigma}^2$ (b) em função de ψ_2, considerando resposta normal e o modelo 2. 99
4.13 Funções de autocorrelação e autocorrelação parcial dos resíduos obtidos com ajustes do modelo 2 para $\psi_2 = 1.0$ (a, c) e 0.3 (b, d) e resposta normal. 100
4.14 Valores observados e estimados com ajustes do modelo 1 com resposta normal (a), do modelo 2 com resposta normal (b) e do modelo 2 com resposta gama (c). Os valores observados são apresentados em linha pontilhada e os estimados em linha cheia. 101
4.15 Gráficos de $\mathcal{L}_1$ e $\hat{\sigma}^2$ em função de ψ_2, considerando resposta gama e o modelo 2. 101

4.16 Funções de autocorrelação (no alto) e autocorrelação parcial (embaixo) dos resíduos obtidos com ajustes do modelo 2 para $\psi_2 = 1.0$ (a, d), $\psi_2 = 0.3$ (b, e) e $\psi_2 = 0.2$ (c, f), resposta gama. 102

C.1 Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 1 116

C.2 Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 2 116

C.3 Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 3 117

C.4 Funções de autocorrelação dos resíduos dos ajustes com $\boldsymbol{\psi}^\top = (0.0, 1.0)$, (a) $\boldsymbol{\psi}^\top = (0.6, 0.4)$ (b), $\boldsymbol{\psi}^\top = (0.7, 0.3)$ (c) e $\boldsymbol{\psi}^\top = (1.0, 0.0)$ (d), para a amostra 3. 117

C.5 Funções de autocorrelação parcial dos resíduos dos ajustes com $\boldsymbol{\psi}^\top = (0.0, 1.0)$, (a) $\boldsymbol{\psi}^\top = (0.6, 0.4)$ (b), $\boldsymbol{\psi}^\top = (0.7, 0.3)$ (c) e $\boldsymbol{\psi}^\top = (1.0, 0.0)$ (d), para a amostra 3. 118

C.6 Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 4 118

C.7 Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 5 119

C.8 Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 6 119

C.9 Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 7 120

C.10 Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 8 120

C.11 Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 9 121

C.12 Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 10 121

C.13 Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 11 122

C.14 Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 12 122

C.15 Comportamento das funções $\mathcal{L}$ (a,d,g), $\mathcal{L}_1$ (c, f, i) e $\mathcal{L}_2$ (b, e, h) ao longo do processo de estimação para a amostra 12. Gráficos (a, b, c): após iteração 1; (d, e, f): após iteração 5 e (g, h , i): após iteração 50 123

C.16 Funções de autocorrelação parcial dos resíduos dos ajustes com $\boldsymbol{\psi}^\top = (0.0, 1.0), (0.8, 0.2), (0.9, 0.1)$ e (1.0,0.0) para a amostra 12. O ponto (0.8,0.2) corresponde ao máximo de $\mathcal{L}_1$ (canto superior direito) e o ponto (1.0,0.0) (no canto inferior direito), foi utilizado na simulação desta amostra. 124

C.17 Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 13 124

Índice de tabelas

3.1 Coeficientes estimados e erros padrões para os modelos 1 a 3 49

3.2 Coeficientes estimados e erros padrões para os modelos 4 ao modelo final. . . 54

4.1 Resíduos condicionais baseados em preditores lineares 83

4.2 Algumas amostras simuladas, modelo normal-gama, $\varphi = (0,1)^\top$, $\sigma^2 = 0.01$. 85

4.3 Algumas amostras simuladas, modelo normal-gama, $\nu_i^2 = 1$, $\sigma^2 = 0.01$ 88

A.1 Família Tweedie 108

Capítulo 1

Introdução

Na prática da Estatística, é comum estarmos interessados na relação entre uma variável resposta e um conjunto de covariáveis observadas sequencialmente. Em geral, processos desta natureza não produzem observações independentes e nestas situações é natural que exista correlação serial entre as observações. Quando a suposição de que as observações provêm de uma distribuição normal é razoável, a análise de tais conjuntos de dados pode ser feita com os métodos estatísticos usuais para análise de dados longitudinais, bem desenvolvidos neste caso (ver, por exemplo, Diggle, Liang e Zeger (1994) ou Singer e Andrade (2000)). Em muitos casos, no entanto, as observações são discretas ou positivas e o uso da distribuição normal é inadequado.

Lima e Sañudo (1996), por exemplo, estudaram o processo de aprendizagem de dois tipos de tarefas, executadas sequencialmente por um grupo de voluntários. O interesse alí eram as diferenças entre o tempo de emissão de um estímulo visual e o tempo da reação do indivíduo (ver mais detalhes na Seção 4.9). Entre as respostas analisadas estavam o erro absoluto e o desvio padrão das observações. Ambas estas medidas requerem distribuições positivas para sua modelagem.

Noutro exemplo, Koyama (1997) analisa um conjunto de dados composto por 1095 observações diárias da concentração de poluentes atmosféricos, condições climáticas e número de óbitos na cidade de São Paulo; entre os objetivos destacamos o estudo do efeito da poluição

no número de óbitos de pessoas com mais de 65 anos de idade. Além da natureza sequencial das observações, a variável resposta neste caso é uma contagem, o que torna inadequada a utilização da distribuição normal para as observações.

A extensão das técnicas baseadas na distribuição normal para outras famílias de distribuições é dificultada pela inexistência de uma classe de distribuições com propriedades similares às da normal multivariada. Devido a isso, nem sempre é possível obter de forma simples estimadores dos parâmetros de interesse para o problema em questão, mesmo no caso em que existe um modelo probabilístico adequado para a distribuição marginal das observações. Uma alternativa para o caso de variável resposta com distribuição marginal na classe de modelos lineares generalizados de Nelder e Wedderburn (1972), foi proposta por Liang e Zeger (1986). Sob este modelo, a estimação dos parâmetros de regressão é feita por meio de equações de estimação generalizadas (EEG) e a matriz de correlações entre as observações é tratada como um parâmetro de perturbação. Um modelo similar foi proposto por Zeger (1988) e utilizado por Koyama (1997) na análise de dados de poluição e mortalidade na cidade de São Paulo. Um dos problemas com esta abordagem, é que a função de verossimilhança não é completamente especificada, o que dificulta a interpretação dos parâmetros e o diagnóstico do modelo.

Modelos de espaço de estados ou modelos lineares dinâmicos (ver Seção 1.2) têm sido também bastante utilizados na modelagem de dados longitudinais, em geral sob o ponto de vista Bayesiano (ver, por exemplo, West e Harrison (1997)). Tais modelos, nos quais os parâmetros de regressão podem variar no tempo, são caracterizados por um par de equações, as *equações de observações* (processo observado) e as *equações de evolução dos parâmetros* (processo latente). Nessa abordagem, é possível acomodar uma grande variedade de distribuições e também descrever várias estruturas de correlação. Os modelos de espaço de estados especificam completamente a distribuição das observações, o que permite uma interpretação mais natural dos parâmetros.

Uma abordagem alternativa para os modelos de espaço de estados foi proposta por Jørgensen *et al.* (1999), que usou um modelo Poisson-gama no qual o processo latente é

modelado segundo um processo de Markov de primeira ordem. Este modelo foi estendido em Jørgensen *et al.*(1996b), incluindo a família Tweedie dos modelos de dispersão introduzidos por Jørgensen (1997). A família Tweedie inclui distribuições importantes, como a normal, a Poisson, a gama e a normal inversa, o que permite a modelagem de uma grande quantidade de fenômenos encontrados na prática. Em contraste com a abordagem Bayesiana usual, o modelo permite a inclusão de covariáveis tanto por meio do processo latente quanto por meio do processo observado. Para a estimação dos parâmetros de regressão, os autores propõem *equações de estimação de Kalman* (ver Seção 3.4.1) que são resolvidas por um algoritmo de escores de Newton. Os autores propõem também uma série de técnicas de diagnóstico que tornam esse método bastante útil na análise de dados longitudinais. Aplicações do modelo Poisson-gama podem ser encontradas em Jørgensen *et al.* (1996a), Skjøth e Lundbye-Christensen (1997) e Botter, Peres e Zandonade (1998).

O objetivo principal deste trabalho é estender o método de Jørgensen *et al.* (1996b) (modelo Tweedie-Tweedie) de modo a considerar uma estrutura de correlação mais ampla gerada por um modelo de Markov de ordem $L \geq 1$ para o processo latente, por exemplo. O modelo que consideramos, além dos parâmetros de regressão e parâmetros de dispersão, admite uma terceira classe de parâmetros, cujos elementos são chamados de *parâmetros retrospectivos*; esses parâmetros definem a ordem do processo latente e a forma de ligação entre o processo observado e o processo latente. Utilizamos um procedimento análogo ao empregado por Jørgensen *et al.* (1996b) para a estimação dos parâmetros de regressão e parâmetros de dispersão e um procedimento baseado na função de verossimilhança perfilada para a estimação dos parâmetros retrospectivos.

Uma introdução à análise de dados longitudinais e uma revisão de métodos de EEG é dada em Diggle *et al.* (1994), por exemplo. Uma extensão dos métodos EEG, que inclui os modelos de dispersão de Jørgensen (1997), é proposta em Artes (1997) e revista em Artes e Jørgensen (1998). Uma interessante revisão de métodos de análise de dados longitudinais que inclui os modelos de espaço de estados é feita em Jørgensen e Tsao (1998).

No Capítulo 2, sistematizamos as propriedades dos preditores lineares, o que permite uma utilização mais objetiva destas informações que se encontram dispersas na literatura. Utilizamos estes resultados para obtenção do filtro e do suavizador de Kalman para um modelo de espaço de estados não gaussiano (para modelos gaussianos, ver por exemplo, Harvey (1981)). Neste capítulo, obtemos também as matrizes de erros quadráticos médios completas para o filtro e o suavizador de Kalman, o que não é, em geral, apresentado na literatura. No Capítulo 3, apresentamos o modelo Tweedie-Tweedie e uma aplicação deste modelo ao conjunto de dados analisado por Koyama (1997). No Capítulo 4, apresentamos o modelo retrospectivo. São obtidos os estimadores dos parâmetros de regressão e suas propriedades, utilizando funções de estimação de Kalman. Para obter estimadores dos parâmetros retrospectivos, utilizamos um procedimento que se baseia na função de verossimilhança perfilada. Apresentamos resultados referentes a algumas amostras simuladas segundo o modelo retrospectivo e um exemplo real com um subconjunto dos dados analisado por Lima e Sañudo (1996). No Apêndice A, apresentamos conceitos básicos sobre a família Tweedie de distribuições e no Apêndice B, apresentamos detalhes de alguns cálculos referentes a resultados do Capítulo 4. No Apêndice C, apresentamos gráficos referentes às simulações.

A seguir, apresentamos alguns conceitos básicos sobre equações de estimação generalizadas e modelos de espaço de estados.

1.1 Equações de estimação generalizadas

Consideramos estimadores que podem ser expressos como soluções de uma equação da forma

$$\mathcal{F}(\mathbf{x};\theta) = \mathbf{0},$$

em que $\mathcal{F}$ representa uma função da amostra $\mathbf{x}$ e de um parâmetro θ, sendo $\mathbf{x}$ e θ associados a um determinado espaço estatístico. A função $\mathcal{F}$ é chamada função de estimação ou função de inferência. Como exemplo, os estimadores de máxima verossimilhança para modelos regulares, são as soluções da equação de estimação

$$u(\mathbf{x};\theta) = \mathbf{0},$$

em que a função de estimação u é a função escore. Outros exemplos conhecidos são os estimadores de mínimos quadrados, os M-estimadores (ver, por exemplo, Huber (1981)) e os estimadores de quase-verossimilhança (ver, por exemplo, Wedderburn (1974)).

A teoria de otimalidade para funções de estimação foi primeiramente apresentada em Godambe (1960) e bastante desenvolvida posteriormente. Em Jørgensen e Laboriau (1994), encontra-se um bom resumo com ênfase em funções de estimação para parâmetros unidimensionais. Para o caso multivariado, Artes (1997) apresenta uma série de resultados estabelecendo condições sob as quais se garante a normalidade assintótica dos estimadores.

Definição 1.1 (*Função de Estimação regular*). *Considere um modelo estatístico* $(\mathcal{X}, \mathcal{A}, \mathcal{P})$, *com* $\mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^k\}$, *sendo* Θ *aberto. Denote por* $p(\cdot, \theta) = \mathrm{d}P_\theta/\mathrm{d}\mu$ *uma versão da derivada de Radon-Nikodym de* P_θ *com respeito a uma medida* σ*-finita* μ. *A função* $\mathcal{F} : \mathcal{X} \times \Theta \to \mathbb{R}^k$ *é uma função de estimação regular se satisfaz as seguintes condições para todo* $\theta = (\theta_1, \ldots, \theta_k)^\top \in \Theta$ *e para* $i, j = 1, \ldots, k$.

i) $\mathrm{E}_\theta\{\mathcal{F}(\theta)\} = \mathbf{0}$; *(*$\mathcal{F}$ *é não viciada)*

ii) A derivada parcial $\partial\mathcal{F}(\mathrm{x};\theta)/\partial\theta_i$ *existe quase certamente;*

iii) A ordem entre integração e diferenciação pode ser trocada como segue,

$$\frac{\partial}{\partial\theta_i}\int_{\mathcal{X}} \mathcal{F}(\mathrm{x};\theta)p(\mathrm{x};\theta)\mathrm{d}\mu(\mathrm{x}) = \int_{\mathcal{X}} \frac{\partial}{\partial\theta_i}\left[\mathcal{F}(\mathrm{x};\theta)p(\mathrm{x};\theta)\right]\mathrm{d}\mu(\mathrm{x});$$

iv) $\mathrm{E}_\theta\left[\{\partial\mathcal{F}_i/\partial\theta_r(\theta)\}\{\partial\mathcal{F}_j/\partial\theta_s(\theta)\}\right] \in \mathbb{R}$ *e a matriz*

$$\mathrm{S}_{\mathcal{F}}(\theta) = \mathrm{E}_\theta\left\{\nabla_\theta\mathcal{F}(\theta)\right\}$$

é não singular.

Em *iv)*, $\mathcal{F}_i$ denota a i-ésima componente do vetor $\mathcal{F}(\theta) = (\mathcal{F}_1(\theta), \ldots, \mathcal{F}_k(\theta))^\top$ e ∇_θ denota o operador gradiente com respeito a θ, isto é,

$$\nabla_\theta\mathcal{F}(\theta) = \frac{\partial}{\partial\theta^\top}\mathcal{F}(\theta).$$

A matriz $\mathrm{S}_{\mathcal{F}}(\theta)$ é chamada matriz sensibilidade de $\mathcal{F}$. Se $\mathcal{F}$ é regular, então a matriz de variabilidade de $\mathcal{F}$, dada por

$$\mathrm{V}_{\mathcal{F}}(\theta) = \mathrm{E}_\theta \left\{ \mathcal{F}(\theta)\mathcal{F}^\top(\theta) \right\}$$

é positiva definida (ver Knudsen (1999)). Quando $\mathcal{F}$ é a função escore, $\mathrm{V}_{\mathcal{F}}(\theta) = -\mathrm{S}_{\mathcal{F}}(\theta)$.

Uma interpretação intuitiva para a sensibilidade e a variabilidade de uma função de estimação é obtida no caso unidimensional. Neste caso, a variabilidade é a variância da função de estimação e a sensibilidade é o valor esperado da derivada da função de estimação em relação ao parâmetro. Alta sensibilidade significa então que uma pequena variação no parâmetro na proximidade do verdadeiro valor provocaria uma grande mudança na função de estimação. Alta sensibilidade e baixa variabilidade são portanto "boas" qualidades de uma função de estimação.

No caso multivariado, podemos comparar duas funções de estimação regulares levando em conta tanto a sensibilidade quanto a variabilidade, utilizando a matriz de informação de Godambe, definida por

$$\mathbf{J}_{\mathcal{F}}(\theta) = \mathrm{S}_{\mathcal{F}}^\top(\theta)\mathrm{V}_{\mathcal{F}}^{-1}(\theta)\mathrm{S}_{\mathcal{F}}(\theta).$$

A informação de Godambe desempenha o papel da informação de Fisher, $\mathbf{I}(\theta)$, para as funções de estimação regulares e coincide com ela quando $\mathcal{F}$ é a função escore. Ela fornece também uma ordenação das funções de estimação. Se $\mathcal{F}_1$ e $\mathcal{F}_2$ são duas funções de estimação regulares, preferimos $\mathcal{F}_1$ a $\mathcal{F}_2$ se para todo $\theta \in \Theta$, temos $\mathbf{J}_{\mathcal{F}_1} \geq \mathbf{J}_{\mathcal{F}_2}$, no sentido que $\mathbf{J}_{\mathcal{F}_1} - \mathbf{J}_{\mathcal{F}_2}$ é positiva semi-definida. Se $\mathcal{F}$ é regular, vale $\mathbf{I}(\theta) \geq \mathbf{J}_{\mathcal{F}}(\theta)$.

Teorema 1.1 *(Jørgensen e Laboriau (1994)). Seja $\theta \in \Theta$ fixo. Se a seqüência de estimadores $\left\{\hat{\theta}_n\right\}_{n=1}^{\infty}$ associada a uma função de estimação regular $\mathcal{F} : \mathcal{X} \times \Theta \to \mathbb{R}^k$ converge em probabilidade para θ, então a seqüência é assintoticamente normal, ou seja,*

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathrm{N}_k \left(\mathbf{0}, \mathbf{J}_{\mathcal{F}}^{-1}(\theta)\right)$$

sob P_θ.

Sob condições que garantem a normalidade assintótica dos estimadores, podemos utilizar um análogo do teste de Wald para testar hipóteses lineares da forma $H_0 : \mathbf{C}\theta = \mathbf{0}$. A estatística de teste é dada por

$$W = (\mathbf{C}\hat{\theta})^\top \left\{\mathbf{C}\mathbf{J}_{\mathcal{F}}^{-1}(\hat{\theta})\mathbf{C}^\top\right\}^{-1} \mathbf{C}\hat{\theta}, \tag{1.1}$$

que, sob a hipótese nula, tem distribuição assintótica χ^2 com s graus de liberdade, em que s é o posto da matriz $\mathbf{C}$.

Para resolver a equação de estimação

$$\mathcal{F}(\mathbf{x};\theta) = \mathbf{0},$$

pode-se utilizar, por exemplo, o algoritmo de Newton-Raphson. Após o passo s, o valor do parâmetro é atualizado por

$$\theta^{(s+1)} = \theta^{(s)} - \left(\nabla_\theta \mathcal{F}^{(s)}\right)^{-1} \mathcal{F}^{(s)},$$

em que $\mathcal{F}^{(s)}$ denota $\mathcal{F}(\mathbf{x}, \theta^{(s)})$. Jørgensen *et al.* (1999) sugerem substituir $\nabla_\theta \mathcal{F}^{(s)}$ por seu valor esperado, a matriz de sensibilidade de $\mathcal{F}$, obtendo o seguinte algoritmo de escores de Newton

$$\theta^{(s+1)} = \theta^{(s)} - \mathbf{S}_{\mathcal{F}}^{-1(s)} \mathcal{F}^{(s)}, \quad s = 0, 1, \ldots .$$

Note que no caso de $\mathcal{F}$ ser a função escore, o algoritmo de escores de Newton é o algoritmo de escores de Fisher. Como no caso da função escore, espera-se que para valores iniciais afastados da solução, o algoritmo de escores de Newton tenha um desempenho melhor que o algoritmo de Newton-Raphson.

1.2 Modelos de espaço de estados

Na sua formulação mais simples, um modelo de espaço de estados ou modelo linear dinâmico, pode ser entendido como um modelo de regressão que permite que os coeficientes de regressão variem no tempo (ver, por exemplo, West e Harrison (1997)). O modelo é caracterizado por

um par de equações chamadas, respectivamente, de *equações de observações* e *equações de evolução dos parâmetros.* As equações de observações para o vetor de respostas $\mathbf{y}_t$ são da forma

$$\mathbf{y}_t = \mathbf{F}_t\theta_t + \boldsymbol{\varepsilon}_t$$

e as equações de evolução dos parâmetros que descrevem a evolução no tempo do vetor de coeficientes de regressão, são da forma

$$\theta_t = \mathbf{G}_t\theta_{t-1} + \boldsymbol{\xi}_t,$$

para $t = 1, \ldots, T$. O vetor de parâmetros de regressão $\boldsymbol{\theta} = (\theta_1^\top, \ldots, \theta_T^\top)^\top$ é também chamado *vetor de estados* ou *processo latente.* As matrizes $\mathbf{F}_t$ e $\mathbf{G}_t$ são não estocásticas e conhecidas. Os vetores $\boldsymbol{\xi}_t$ e $\boldsymbol{\varepsilon}_t$ são chamados *inovações* e são supostos independentes, com $\boldsymbol{\varepsilon}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_t)$ e $\boldsymbol{\xi}_t \sim \mathrm{N}(\mathbf{0}, \mathbf{W}_t)$.

No enfoque Bayesiano, o modelo é completado com uma distribuição a priori para θ_1. Em geral, assume-se $\theta_1 \sim \mathrm{N}(\mathbf{a}_1, \mathbf{R}_1)$. Neste contexto, a inferência para modelos lineares dinâmicos é feita seqüencialmente. Para cada tempo t, as distribuições a priori, preditiva e a posteriori para θ_t são obtidas. Para o caso em que as matrizes $\mathbf{F}_t, \mathbf{G}_t, \boldsymbol{\Sigma}_t$ e $\mathbf{W}_t$ são conhecidas, o filtro de Kalman fornece soluções exatas (ver, por exemplo, Harvey (1981)).

A formulação do modelo de espaço de estados pode ser estendida para uma série de outros modelos como os modelos ARMA(p, q) de Box e Jenkins (1976), por exemplo. Neste caso, o processo latente não é mais tratado necessariamente como parâmetro de regressão e as matrizes $\mathbf{F}_t$ e $\mathbf{G}_t$ podem ser funções de outros parâmetros de interesse.

Do ponto de vista clássico, este problema pode ser tratado como um caso de dados incompletos. O processo latente $\boldsymbol{\theta} = (\theta_1^\top, \cdots, \theta_T^\top)^\top$ é visto como a parte não observada de um hipotético vetor de observações $(\mathbf{Y}^\top, \boldsymbol{\theta}^\top)^\top$ e o algoritmo EM (ver Dempster, Laird e Rubin (1977)) pode então ser utilizado. No passo E, obtém-se a esperança condicional de $\boldsymbol{\theta}$ dado $\mathbf{Y}$ e no passo M obtêm-se as estimativas de máxima verossimilhança para os outros parâmetros de interesse. No caso específico da distribuição normal, tem-se que $\boldsymbol{\theta}|\mathbf{Y} \sim \mathrm{N}\left(\mathrm{E}(\boldsymbol{\theta}|\mathbf{Y}), \boldsymbol{\Sigma}_{\boldsymbol{\theta}|\mathbf{Y}}\right)$. Logo, a esperança condicional $\mathrm{E}(\boldsymbol{\theta}|\mathbf{Y})$ é linear em $\mathbf{Y}$ e pode ser obtida seqüencialmente por

intermédio do filtro de Kalman. Esta metodologia aplicada a um modelo de séries temporais com inovações normais, pode ser encontrada em Shumway (1988).

Uma extensão dos modelos de espaço de estados para distribuições na classe da família exponencial foi apresentada por West, Harrison e Migon (1985). O modelo é descrito a partir das distribuições condicionais das observações dada toda a informação passada. O modelo do sistema de estados é definido pelo estado atual, representado pela distribuição condicional de θ_t dado o passado, e pelo estado futuro, representado pela distribuição condicional de θ_{t+k+1} dado θ_{t+k}. O modelo das observações é definido pela distribuição condicional de Y_{t+k} dado θ_{t+k}. Os autores utilizam distribuições a priori conjugadas, em um tratamento inteiramente Bayesiano do problema.

Uma outra formulação para dados não normais é considerada em Harvey e Fernandes (1989). Eles propõem um modelo de espaço de estados univariado, no qual o mecanismo de transição de θ_t para θ_{t+1} é definido implicitamente. Covariáveis podem ser consideradas no modelo por meio de funções de ligação como nos modelos lineares generalizados de Nelder e Wedderburn (1972). Um algoritmo análogo ao filtro de Kalman é utilizado para construir a função de verossimilhança e a estimação dos parâmetros utiliza técnicas clássicas e Bayesianas.

Capítulo 2

Preditores lineares e o filtro de Kalman

Neste Capítulo, apresentamos a definição e algumas propriedades de preditores lineares e obtemos o filtro e o suavizador de Kalman para um modelo de espaço de estados. Os resultados que apresentamos seguem de perto aqueles discutidos em Jørgensen *et al.* (1999) e em Anderson e Moore (1979).

2.1 Preditores lineares

Definição 2.1 *Sejam* $\mathbf{X}$ *e* $\mathbf{Y}$ *vetores aleatórios definidos no mesmo espaço de probabilidades, com segundos momentos finitos. Um preditor linear de* $\mathbf{X}$ *dado* $\mathbf{Y}$ *(preditor afim) é um estimador da forma* $\mathbf{AY} + \mathbf{b}$*, em que* $\mathbf{A}$ *e* $\mathbf{b}$ *são, respectivamente, uma matriz e um vetor de constantes. O melhor preditor linear não viciado de* $\mathbf{X}$ *dado* $\mathbf{Y}$*,* $\mathbf{m}_{X|Y}$*, é tal que o erro quadrático médio associado a ele é mínimo, isto é,* $\mathbf{m}_{X|Y} = \mathbf{A}^{o}\mathbf{Y} + \mathbf{b}^{o}$ *satisfaz*

$$\mathrm{E}\{||\mathbf{X} - \mathbf{m}_{X|Y}||^2\} = \mathrm{E}\{||\mathbf{X} - \mathbf{A}^{o}\mathbf{Y} - \mathbf{b}^{o}||^2\} \leq \mathrm{E}\{||\mathbf{X} - \mathbf{AY} - \mathbf{b}||^2\}, \quad \text{para todo } \mathbf{A} \text{ e } \mathbf{b}.$$

O resultado seguinte mostra que $\mathbf{A}^{o}$ e $\mathbf{b}^{o}$ dependem apenas dos primeiro e segundo momentos dos vetores $\mathbf{X}$ e $\mathbf{Y}$.

Teorema 2.1 *(Anderson e Moore (1979)). Sejam*

$$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \quad \text{e} \quad \Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}$$

respectivamente, o vetor de médias e a matriz de covariância do vetor aleatório $(\mathbf{X}^\top, \mathbf{Y}^\top)^\top$. *Então, se* Σ_Y *é não singular, o melhor preditor linear de* $\mathbf{X}$ *dado* $\mathbf{Y}$ *é*

$$\mathbf{m}_{X|Y} = \mu_X + \Sigma_{XY}\Sigma_Y^{-1}(\mathbf{Y} - \mu_Y). \tag{2.1}$$

Se Σ_Y *é singular, então o erro quadrático médio associado ao preditor linear*

$$\mathbf{m}_{X|Y} = \mu_X + \Sigma_{XY}(\Sigma_Y^- + \bar{\mathbf{A}})(\mathbf{Y} - \mu_Y),$$

é mínimo. Na expressão acima, $\bar{\mathbf{A}}$ *é qualquer matriz tal que* $\bar{\mathbf{A}}\Sigma_Y = \mathbf{0}$ *e* Σ_Y^- *é qualquer inversa generalizada de* Σ_Y, *isto é, tal que* $\Sigma_Y\Sigma_Y^-\Sigma_Y = \Sigma_Y$.

Demonstração. (Caso Σ_Y não singular).
Queremos encontrar $\mathbf{A}$ e $\mathbf{b}$ que minimizam $\mathrm{E}||\mathbf{X} - \mathbf{A}\mathbf{Y} - \mathbf{b}||^2$. Como $\mathrm{Var}(\mathbf{Z}) = \mathrm{E}(\mathbf{Z}\mathbf{Z}^\top) - \mathrm{E}(\mathbf{Z})\mathrm{E}(\mathbf{Z}^\top)$, temos que

$$\begin{aligned} \mathrm{E}||\mathbf{Z}||^2 &= \mathrm{E}[\mathbf{Z}^\top\mathbf{Z}] = \mathrm{E}\{\mathrm{tr}(\mathbf{Z}\mathbf{Z}^\top)\} = \mathrm{tr}\{\mathrm{E}(\mathbf{Z}\mathbf{Z}^\top)\} \\ &= \mathrm{tr}\{\mathrm{Var}(\mathbf{Z})\} + ||\mathrm{E}(\mathbf{Z})||^2. \end{aligned}$$

Obtemos então,

$$\begin{aligned} \mathrm{E}||\mathbf{X} - \mathbf{A}\mathbf{Y} - \mathbf{b}||^2 &= \mathrm{tr}\{\mathrm{Var}(\mathbf{X} - \mathbf{A}\mathbf{Y} - \mathbf{b})\} + ||\mu_X - \mathbf{A}\mu_Y - \mathbf{b}||^2 \\ &= \mathrm{tr}\Big\{\Sigma_X + \mathbf{A}\Sigma_Y\mathbf{A}^\top - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_Y\mathbf{A}^\top - \mathbf{A}\Sigma_Y\Sigma_Y^{-1}\Sigma_{YX} \\ &\quad + \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX} - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Big\} + ||\mu_X - \mathbf{A}\mu_Y - \mathbf{b}||^2 \\ &= \mathrm{tr}\Big\{\big(\mathbf{A} - \Sigma_{XY}\Sigma_Y^{-1}\big)\Sigma_Y\big(\mathbf{A}^\top - \Sigma_Y^{-1}\Sigma_{YX}\big)\Big\} \\ &\quad + \mathrm{tr}(\Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}) + ||\mu_X - \mathbf{A}\mu_Y - \mathbf{b}||^2. \end{aligned} \tag{2.2}$$

O primeiro termo do lado direito da igualdade acima é não negativo, pois ele é o traço de uma matriz da forma $\mathbf{B}^\top\Sigma_Y\mathbf{B}$ e Σ_Y é positiva definida. O segundo termo não depende

de $\mathbf{A}$ ou $\mathbf{b}$ e o terceiro termo também é não negativo. Logo, o valor mínimo da expressão acima é atingido quando o primeiro e o terceiro termo se anulam o que acontece quando $\mathbf{A} = \Sigma_{XY}\Sigma_Y^{-1}$ e $\mathbf{b} = \mu_X - \mathbf{A}\mu_Y$, ou seja quando

$$\begin{aligned} \mathbf{m}_{X|Y} &= \mathbf{AY} + \mathbf{b} = \Sigma_{XY}\Sigma_Y^{-1}\mathbf{Y} + \mu_X - \Sigma_{XY}\Sigma_Y^{-1}\mu_Y \\ &= \mu_X + \Sigma_{XY}\Sigma_Y^{-1}(\mathbf{Y} - \mu_Y). \end{aligned}$$

□

Da expressão (2.2), obtemos a matriz de erros quadráticos médios de $\mathbf{m}_{X|Y}$,

$$\mathbf{C}_{X|Y} = \Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}.$$

Vamos nos referir daqui em diante ao par $\mathbf{m}_{X|Y}$ e sua matriz de erros quadráticos médios $\mathbf{C}_{X|Y}$ por *preditor linear de* $\mathbf{X}$ *dado* $\mathbf{Y}$. Denotaremos o preditor linear por

$$\mathbf{X}|\mathbf{Y} \sim \left[\mathbf{m}_{X|Y}; \mathbf{C}_{X|Y}\right].$$

De (2.2) obtemos também que se $\mathrm{EQM}_{A,b}$ denota a matriz de erros quadráticos médios associada a $\mathbf{AY} + \mathbf{b}$, então $\mathrm{EQM}_{A,b} - \mathrm{EQM}_{\mathbf{m}_{X|Y}}$ é positiva semi-definida.

Os preditores lineares compartilham uma série de propriedades com a esperança condicional. A seguir, apresentamos algumas dessas propriedades que serão utilizadas na obtenção do filtro de Kalman.

Proposição 2.1 i) $\mathbf{m}_{X|Y}$ *é um estimador não viciado de* $\mathbf{X}$, *isto é,* $\mathrm{E}\left[\mathbf{X} - \mathbf{m}_{X|Y}\right] = 0$, *(ou, equivalentemente,* $\mathrm{E}(\mathbf{m}_{X|Y}) = \mathrm{E}(\mathbf{X})$*);*

ii) $\mathrm{Var}[\mathbf{X} - \mathbf{m}_{X|Y}] = \mathbf{C}_{X|Y}$*;*

iii) $\mathrm{Cov}[\mathbf{X}, \mathbf{m}_{X|Y}] = \mathrm{Var}(\mathbf{m}_{X|Y}) = \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX} = \Sigma_X - \mathbf{C}_{X|Y}$*;*

iv) $\mathrm{Cov}[\mathbf{Y}, \mathbf{m}_{X|Y}] = \Sigma_{YX}$.

Demonstração. i) e ii) decorrem imediatamente de (2.1). Agora,

$$\begin{aligned}\mathrm{Cov}[\mathbf{X}, \mathbf{m}_{X|Y}] &= \mathrm{Cov}[\mathbf{X}, \boldsymbol{\mu}_X + \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}(\mathbf{Y} - \boldsymbol{\mu}_Y)] \\ &= \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\Sigma}_{YX},\end{aligned}$$

e

$$\begin{aligned}\mathrm{Cov}[\mathbf{Y}, \mathbf{m}_{X|Y}] &= \mathrm{Cov}[\mathbf{Y}, \boldsymbol{\mu}_X + \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}(\mathbf{Y} - \boldsymbol{\mu}_Y)] \\ &= \boldsymbol{\Sigma}_Y\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\Sigma}_{YX} = \boldsymbol{\Sigma}_{YX}.\end{aligned}$$

□

No teorema a seguir usaremos o *princípio da ortogonalidade* definido a seguir.

Sejam $\mathbf{X}$ e $\mathbf{Y}$ vetores aleatórios definidos no mesmo espaço de probabilidades. Dizemos que $\mathbf{X}$ e $\mathbf{Y}$ são *ortogonais* e escrevemos $\mathbf{X} \perp \mathbf{Y}$, se $\mathrm{E}(\mathbf{X}\mathbf{Y}^\top) = \mathbf{0}$.

Proposição 2.2 *(Teorema da Projeção). Sejam $\mathbf{X}$ e $\mathbf{Y}$ vetores aleatórios definidos no mesmo espaço de probabilidades e tais que seus segundos momentos são finitos. Então o erro $\mathbf{X} - \mathbf{m}_{X|Y}$ é ortogonal a $\mathbf{Y}$, isto é*

$$\mathrm{E}\left[(\mathbf{X} - \mathbf{m}_{X|Y})\mathbf{Y}^\top\right] = \mathbf{0}.$$

Reciprocamente, se para algum par $(\mathbf{A}, \mathbf{b})$, $\mathrm{E}\left[(\mathbf{X} - \mathbf{A}\mathbf{Y} - \mathbf{b})\mathbf{Y}^\top\right] = \mathbf{0}$ e $\mathrm{E}[\mathbf{X} - \mathbf{A}\mathbf{Y} - \mathbf{b}] = 0$, então $\mathbf{m}_{X|Y} = \mathbf{A}\mathbf{Y} + \mathbf{b}$.

Demonstração. ($\Rightarrow$) Temos que

$$\begin{aligned}\mathrm{E}[(\mathbf{X} - \mathbf{m}_{X|Y})\mathbf{Y}^\top] &= \mathrm{E}\left[\left\{(\mathbf{X} - \boldsymbol{\mu}_X) - \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}(\mathbf{Y} - \boldsymbol{\mu}_Y)\right\}Y^\top\right] \\ &= \boldsymbol{\Sigma}_{XY} - \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\Sigma}_Y = \mathbf{0}.\end{aligned}$$

($\Leftarrow$) Agora, $\mathrm{E}\left[(\mathbf{X} - \mathbf{A}\mathbf{Y} - \mathbf{b})\mathbf{Y}^\top\right] = \mathbf{0}$ implica $\boldsymbol{\Sigma}_{XY} - \mathbf{A}\boldsymbol{\Sigma}_Y = \mathbf{0}$, pois $\mathrm{E}(\mathbf{X} - \mathbf{A}\mathbf{Y} - \mathbf{b}) = \mathbf{0}$. Logo, $\mathbf{A} = \boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}$. Usando novamente $\mathrm{E}(\mathbf{X} - \mathbf{A}\mathbf{Y} - \mathbf{b}) = \mathbf{0}$, obtemos $\mathbf{b} = \boldsymbol{\mu}_X - \mathbf{A}\boldsymbol{\mu}_Y$ e o

resultado segue. □

Observação. Se em analogia a uma interpretação geométrica, considerarmos o espaço linear $\mathcal{V}(\mathbf{Y})$ gerado pelas transformações afins do vetor aleatório $\mathbf{Y}$, o resultado acima diz que podemos interpretar $\mathbf{m}_{X|Y}$ como sendo a *projeção* do vetor aleatório $\mathbf{X}$ em $\mathcal{V}(\mathbf{Y})$ no seguinte sentido

i) $\mathbf{m}_{X|Y} \in \mathcal{V}(\mathbf{Y})$;

ii) $\mathbf{X} - \mathbf{m}_{X|Y} \perp \mathcal{V}(\mathbf{Y})$;

iii) $\mathbf{X} \in \mathcal{V}(\mathbf{Y}) \Rightarrow \mathbf{m}_{X|Y} = \mathbf{X}$. Neste caso, $\mathbf{C}_{X|Y} = \mathbf{0}$;

iv) $\mathbf{X} - \boldsymbol{\mu}_X \perp \mathcal{V}(\mathbf{Y}) \Rightarrow \mathbf{m}_{X|Y} = \boldsymbol{\mu}_X$. Neste caso, $\mathbf{C}_{X|Y} = \boldsymbol{\Sigma}_X$.

Como corolário do teorema da projeção, obtemos a seguinte propriedade,

Proposição 2.3 *Se* $\mathrm{E}[\mathbf{X}|\mathbf{Y}]$ *é linear em* $\mathbf{Y}$, *então*

$$\mathbf{X}|\mathbf{Y} \sim [\mathrm{E}[\mathbf{X}|\mathbf{Y}]; \mathrm{E}\left(\mathrm{Var}[\mathbf{X}|\mathbf{Y}]\right)].$$

Demonstração. A parte linear é imediata. Agora, $\mathrm{E}(\mathbf{C}_{X|Y}) = \mathrm{Var}(\mathbf{X} - \mathbf{m}_{X|Y}) = \mathrm{Var}\{\mathbf{X} - \mathrm{E}(\mathbf{X}|\mathbf{Y})\} = \mathrm{E}\{\mathrm{Var}(\mathbf{X}|\mathbf{Y})\}$. □

Análogamente à esperança e à variância condicionais, o preditor linear também satisfaz

Proposição 2.4 *(Linearidade). Para* $\mathbf{A}$, $\mathbf{B}$ *e* $\mathbf{D}$ *matrizes e vetores de constantes e* $\mathbf{X}, \mathbf{Y}$ *e* $\mathbf{Z}$ *vetores aleatórios, temos*

$$\begin{aligned}\mathbf{AX} + \mathbf{BY} + \mathbf{D}|\mathbf{Z} \sim [\ & \mathbf{A}\mathbf{m}_{X|Z} + \mathbf{B}\mathbf{m}_{Y|Z} + \mathbf{D}; \\ & \mathbf{A}\mathbf{C}_{X|Z}\mathbf{A}^\top + \mathbf{B}\mathbf{C}_{Y|Z}\mathbf{B}^\top + \mathbf{A}\mathbf{C}_{XY|Z}\mathbf{B}^\top + \mathbf{B}\mathbf{C}_{YX|Z}\mathbf{A}^\top],\end{aligned}$$

em que $\mathbf{C}_{XY|Z} = \boldsymbol{\Sigma}_{XY} - \boldsymbol{\Sigma}_{XZ}\boldsymbol{\Sigma}_Z^{-1}\boldsymbol{\Sigma}_{ZY}$.

Demonstração. De (2.1), obtemos

$$\mathbf{m}_{AX+BY+D|Z} = \boldsymbol{\mu}_{AX+BY+D} + \mathrm{Cov}(\mathbf{AX}+\mathbf{BY}+\mathbf{D},\mathbf{Z})\boldsymbol{\Sigma}_Z^{-1}(\mathbf{Z}-\boldsymbol{\mu}_Z).$$

Como $\mathrm{Cov}(\mathbf{AX}+\mathbf{BY}+\mathbf{D},\mathbf{Z}) = \mathbf{A}\boldsymbol{\Sigma}_{XZ} + \mathbf{B}\boldsymbol{\Sigma}_{YZ}$, obtemos

$$\begin{aligned}
\mathbf{m}_{AX+BY+D|Z} &= \mathbf{A}\boldsymbol{\mu}_X + \mathbf{B}\boldsymbol{\mu}_Y + \mathbf{D} + (\mathbf{A}\boldsymbol{\Sigma}_{XZ} + \mathbf{B}\boldsymbol{\Sigma}_{YZ})\,\boldsymbol{\Sigma}_Z^{-1}(\mathbf{Z}-\boldsymbol{\mu}_Z) \\
&= \mathbf{A}\boldsymbol{\mu}_X + \mathbf{A}\boldsymbol{\Sigma}_{XZ}\boldsymbol{\Sigma}_Z^{-1}(\mathbf{Z}-\boldsymbol{\mu}_Z) + \mathbf{B}\boldsymbol{\mu}_Y + \mathbf{B}\boldsymbol{\Sigma}_{YZ}\boldsymbol{\Sigma}_Z^{-1}(\mathbf{Z}-\boldsymbol{\mu}_Z) + \mathbf{D} \\
&= \mathbf{A}\mathbf{m}_{X|Z} + \mathbf{B}\mathbf{m}_{Y|Z} + \mathbf{D}.
\end{aligned}$$

Temos também $\mathrm{Var}(\mathbf{AX}+\mathbf{BY}+\mathbf{D}) = \mathbf{A}\boldsymbol{\Sigma}_X\mathbf{A}^\top + \mathbf{B}\boldsymbol{\Sigma}_Y\mathbf{B}^\top + \mathbf{A}\boldsymbol{\Sigma}_{XY}\mathbf{B}^\top + \mathbf{B}\boldsymbol{\Sigma}_{YX}\mathbf{A}^\top$. Logo,

$$\begin{aligned}
\mathbf{C}_{AX+BY+D|Z} &= \boldsymbol{\Sigma}_{AX+BY+D} - \boldsymbol{\Sigma}_{(AX+BY+D),Z}\boldsymbol{\Sigma}_Z^{-1}\boldsymbol{\Sigma}_{Z,(AX+BY+D)} \\
&= \mathbf{A}\boldsymbol{\Sigma}_X\mathbf{A}^\top + \mathbf{B}\boldsymbol{\Sigma}_Y\mathbf{B}^\top + \mathbf{A}\boldsymbol{\Sigma}_{XY}\mathbf{B}^\top + \mathbf{B}\boldsymbol{\Sigma}_{YX}\mathbf{A}^\top \\
&\quad - \left(\mathbf{A}\boldsymbol{\Sigma}_{XZ} + \mathbf{B}\boldsymbol{\Sigma}_{YZ}\right)\boldsymbol{\Sigma}_Z^{-1}\left(\boldsymbol{\Sigma}_{ZX}\mathbf{A}^\top + \boldsymbol{\Sigma}_{ZY}\mathbf{B}^\top\right) \\
&= \mathbf{A}\boldsymbol{\Sigma}_X\mathbf{A}^\top - \mathbf{A}\boldsymbol{\Sigma}_{XZ}\boldsymbol{\Sigma}_Z^{-1}\boldsymbol{\Sigma}_{ZX}\mathbf{A}^\top + \mathbf{B}\boldsymbol{\Sigma}_Y\mathbf{B}^\top - \mathbf{B}\boldsymbol{\Sigma}_{YZ}\boldsymbol{\Sigma}_Z^{-1}\boldsymbol{\Sigma}_{ZY}\mathbf{B}^\top \\
&\quad + \mathbf{A}\boldsymbol{\Sigma}_{XY}\mathbf{B}^\top + \mathbf{B}\boldsymbol{\Sigma}_{YX}\mathbf{A}^\top - \mathbf{A}\boldsymbol{\Sigma}_{XZ}\boldsymbol{\Sigma}_Z^{-1}\boldsymbol{\Sigma}_{ZY}\mathbf{B} - \mathbf{B}\boldsymbol{\Sigma}_{YZ}\boldsymbol{\Sigma}_Z^{-1}\boldsymbol{\Sigma}_{ZX}\mathbf{A} \\
&= \mathbf{A}\mathbf{C}_{X|Z}\mathbf{A}^\top + \mathbf{B}\mathbf{C}_{Y|Z}\mathbf{B}^\top + \mathbf{A}\mathbf{C}_{XY|Z}\mathbf{B}^\top + \mathbf{B}\mathbf{C}_{YX|Z}\mathbf{A}^\top.
\end{aligned}$$

□

Observação. Notar que na expressão acima $\boldsymbol{\Sigma}_{YX} + \boldsymbol{\Sigma}_{XZ}\boldsymbol{\Sigma}_Z^{-1}\boldsymbol{\Sigma}_{ZY} = \mathrm{Cov}(\mathbf{X},\mathbf{Y}-\mathbf{m}_{Y|Z})$.

Proposição 2.5 *Se* $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_k$ *são não correlacionados, então*

$$\begin{aligned}
\mathbf{X}|\mathbf{Y}_1,\mathbf{Y}_2,\ldots,\mathbf{Y}_k \sim [\ & \mathbf{m}_{X|Y_1} + \mathbf{m}_{X|Y_2} + \ldots + \mathbf{m}_{X|Y_k} - (k-1)\boldsymbol{\mu}_X; \\
& \mathbf{C}_{X|Y_1} + \mathbf{C}_{X|Y_2} + \ldots + \mathbf{C}_{X|Y_k} - (k-1)\boldsymbol{\Sigma}_X].
\end{aligned}$$

Para $k = 2$*, o resultado acima pode ser escrito como: Se* $\mathbf{Y}$ *e* $\mathbf{Z}$ *são não correlacionados, então* $\mathbf{X}|\mathbf{Y},\mathbf{Z} \sim \left[\mathbf{m}_{X|Y,Z}; \mathbf{C}_{X|Y,Z}\right]$*, em que*

$$\mathbf{m}_{X|Y,Z} = \mathbf{m}_{X|Y} + \boldsymbol{\Sigma}_{XZ}\boldsymbol{\Sigma}_Z^{-1}(\mathbf{Z}-\boldsymbol{\mu}_Z), \quad \mathbf{C}_{X|Y,Z} = \mathbf{C}_{X|Y} - \boldsymbol{\Sigma}_{XZ}\boldsymbol{\Sigma}_Z^{-1}\boldsymbol{\Sigma}_{ZX}.$$

Demonstração. O resultado segue diretamente de (2.1), aplicada à matriz diagonal em blocos $\Sigma_{\mathbf{Y}}$, em que $\mathbf{Y}^\top = (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_k^\top)^\top$. □

Proposição 2.6 *Sejam* $\mathbf{X}$, $\mathbf{Y}$ *e* $\mathbf{Z}$ *vetores aleatórios com segundos momentos finitos. Então, o preditor linear conjunto de* $(\mathbf{X},\mathbf{Y})$ *dado* $\mathbf{Z}$ *é*

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \Bigg| \mathbf{Z} \sim \left[\begin{pmatrix} \mathbf{m}_{X|Z} \\ \mathbf{m}_{Y|Z} \end{pmatrix} ; \begin{pmatrix} \mathbf{C}_{X|Z} & \mathbf{C}_{XY|Z} \\ \mathbf{C}_{YX|Z} & \mathbf{C}_{Y|Z} \end{pmatrix} \right], \tag{2.3}$$

em que $\mathbf{C}_{XY|Z}$ *é como na Proposição 2.4.*

Demonstração. Note que

$$\mathrm{Cov}\left\{ \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}, \mathbf{Z} \right\} = \begin{pmatrix} \Sigma_{XZ} \\ \Sigma_{YZ} \end{pmatrix} \text{ e } \mathrm{Var} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}.$$

O resultado segue, então, diretamente de (2.1). □

Proposição 2.7 *Sejam* $\mathbf{X}$, $\mathbf{Y}$ *e* $\mathbf{Z}$ *como na Proposição 2.6. O preditor linear de* $\mathbf{X}$ *dados* $\mathbf{Y}$ *e* $\mathbf{Z}$ *(preditor linear condicional) é*

$$\mathbf{X} \Bigg| \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} \sim \left[\mathbf{m}_{X|Z} + \mathbf{C}_{XY|Z}\mathbf{C}_{Y|Z}^{-1}(\mathbf{Y} - \mathbf{m}_{Y|Z}) \; ; \; \mathbf{C}_{X|Z} - \mathbf{C}_{XY|Z}\mathbf{C}_{Y|Z}^{-1}\mathbf{C}_{YX|Z} \right].$$

Demonstração. Seja $\mathcal{V}(\mathbf{Y},\mathbf{Z})$ o espaço linear gerado por $\mathbf{Y}$ e $\mathbf{Z}$. Pelo teorema da projeção (Proposição 2.2), temos $(\mathbf{Y} - \mathbf{m}_{X|Z}) \perp \mathbf{Z}$ e $\mathbf{m}_{Y|Z} \in \mathcal{V}(\mathbf{Z})$, o que implica que $\mathcal{V}(\mathbf{Y},\mathbf{Z}) = \mathcal{V}(\mathbf{Y} - \mathbf{m}_{Y|Z}, \mathbf{Z})$. Logo,

$$\mathbf{X} \Bigg| \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} = \mathbf{X} \Bigg| \begin{pmatrix} \mathbf{Y} - \mathbf{m}_{Y|Z} \\ \mathbf{Z} \end{pmatrix}.$$

A Proposição 2.5 implica então que

$$\mathbf{X} \Bigg| \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} = \left[\mathbf{m}_{X|Z} + \mathbf{m}_{X|(Y-\mathbf{m}_{Y|Z})} - \boldsymbol{\mu}_X ; \mathbf{C}_{X|Z} + \mathbf{C}_{X|(Y-\mathbf{m}_{Y|Z})} - \Sigma_X \right].$$

Aplicando agora (2.1) ao preditor linear de $\mathbf{X}$ dado $(\mathbf{Y} - \mathbf{m}_{Y|Z})$, obtemos

$$\begin{aligned}
\mathbf{m}_{X|(Y-\mathbf{m}_{Y|Z})} &= \boldsymbol{\mu}_X + \Sigma_{X,(Y-\mathbf{m}_{Y|Z})}\Sigma^{-1}_{Y-\mathbf{m}_{Y|Z}}(\mathbf{Y} - \mathbf{m}_{Y|Z}) \text{ e} \\
\mathbf{C}_{X|(Y-\mathbf{m}_{Y|Z})} &= \Sigma_X - \Sigma_{X,(Y-\mathbf{m}_{Y|Z})}\Sigma^{-1}_{Y-\mathbf{m}_{Y|Z}}\Sigma_{(Y-\mathbf{m}_{Y|Z}),X}.
\end{aligned}$$

Pelo item ii) da Proposição 2.1, temos que $\Sigma_{Y-\mathbf{m}_{Y|Z}} = \mathbf{C}_{Y|Z}$ e da Proposição 2.4, temos que $\Sigma_{X,(Y-\mathbf{m}_{Y|Z})} = \mathbf{C}_{XY|Z}$. □

Proposição 2.8 *(Corolário da Proposição 2.6). Se dado* $\mathbf{Z}$, $\mathbf{X}$ *e* $\mathbf{Y}$ *são não correlacionados e se ou* $\mathrm{E}[\mathbf{X}|\mathbf{Z}]$ *ou* $\mathrm{E}[\mathbf{Y}|\mathbf{Z}]$ *forem lineares em* $\mathbf{Z}$, *então*

$$\mathbf{X} \left| \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} \right. \equiv \mathbf{X}|\mathbf{Z} \sim [\mathbf{m}_{X|Z}; \mathbf{C}_{X|Z}].$$

Demonstração. Temos

$$\mathrm{Cov}(\mathbf{X}, \mathbf{Y} - \mathbf{m}_{Y|Z}) = \mathrm{E}\left[\mathrm{Cov}(\mathbf{X}, \mathbf{Y} - \mathbf{m}_{Y|Z}|\mathbf{Z})\right] + \mathrm{Cov}\left[\mathrm{E}(\mathbf{X}|\mathbf{Z}), \mathrm{E}(\mathbf{Y} - \mathbf{m}_{Y|Z}|\mathbf{Z})\right].$$

Se $\mathrm{E}(\mathbf{Y}|\mathbf{Z})$ é linear em $\mathbf{Z}$, então a Proposição 2.3 nos dá $\mathbf{m}_{Y|Z} = \mathrm{E}(\mathbf{Y}|\mathbf{Z})$ e o segundo termo do lado direito da igualdade acima se anula. Se $\mathrm{E}(\mathbf{X}|\mathbf{Z})$ for linear em $\mathbf{Z}$, digamos $\mathrm{E}(\mathbf{X}|\mathbf{Z}) = \mathbf{A}\mathbf{Z} + \mathbf{b}$, temos

$$\begin{aligned}
\mathrm{Cov}\left[\mathbf{A}\mathbf{Z} + \mathbf{b}, \mathrm{E}(\mathbf{Y} - \mathbf{m}_{Y|Z})|\mathbf{Z}\right] &= \mathrm{Cov}\left[\mathbf{A}\mathbf{Z} + \mathbf{b}, \mathrm{E}(\mathbf{Y}|\mathbf{Z}) - \mathbf{m}_{Y|Z}\right] \\
&= \mathrm{Cov}\left[\mathbf{A}\mathbf{Z} + \mathbf{b}, \mathbf{Y} - \mathbf{m}_{Y|Z}\right] = \mathbf{A}\Sigma_{ZY} - \mathbf{A}\Sigma_{Z,\mathbf{m}_{Y|Z}} = \mathbf{0},
\end{aligned}$$

pois $\mathrm{Cov}(\mathbf{Z}, \mathrm{E}(\mathbf{Y}|\mathbf{Z})) = \mathrm{Cov}(\mathbf{Z}, \mathbf{Y})$ e $\Sigma_{Z,\mathbf{m}_{Y|Z}} = \mathrm{Cov}\left(\mathbf{Z}, \boldsymbol{\mu}_Y + \Sigma_{YZ}\Sigma_Z^{-1}(\mathbf{Z} - \boldsymbol{\mu}_Z)\right) = \mathrm{Cov}(\mathbf{Z}, \mathbf{Z} - \boldsymbol{\mu}_Z)\Sigma_Z^{-1}\Sigma_{ZY} = \Sigma_{ZY}$. □

Proposição 2.9 *Suponha que* $\mathbf{Y}|\mathbf{Z} \sim [\mathbf{m}_{Y|Z}; \mathbf{C}_{Y|Z}]$ *e que* $\mathbf{X}|\mathbf{Y}, \mathbf{Z} \sim [\mathbf{A}\mathbf{Y} + \mathbf{B}\mathbf{Z} + \mathbf{D}; \mathbf{C}_{X|Y,Z}]$, *em que* $\mathbf{A}, \mathbf{B}$ *e* $\mathbf{D}$ *são matrizes de constantes e* $\mathbf{X}$, $\mathbf{Y}$ *e* $\mathbf{Z}$ *são vetores aleatórios. Então o preditor conjunto de* $(\mathbf{X}, \mathbf{Y})$ *dado* $\mathbf{Z}$ *é*

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \Bigg| \mathbf{Z} \sim \left[\begin{pmatrix} \mathbf{A}\mathbf{m}_{Y|Z} + \mathbf{B}\mathbf{Z} + \mathbf{D} \\ \mathbf{m}_{Y|Z} \end{pmatrix}; \begin{pmatrix} \mathbf{C}_{X|Y,Z} + \mathbf{A}\mathbf{C}_{Y|Z}\mathbf{A}^\top & \mathbf{A}\mathbf{C}_{Y|Z} \\ \mathbf{C}_{Y|Z}\mathbf{A}^\top & \mathbf{C}_{Y|Z} \end{pmatrix} \right].$$

Em particular, $\mathbf{X}|\mathbf{Z} \sim [\mathbf{A}\mathbf{m}_{Y|Z} + \mathbf{B}\mathbf{Z} + \mathbf{D}\ ;\ \mathbf{C}_{X|Y,Z} + \mathbf{A}\mathbf{C}_{Y|Z}\mathbf{A}^\top]$.

Demonstração. Pela Proposição 2.6, temos

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \Bigg| \mathbf{Z} \sim \left[\begin{pmatrix} \mathbf{m}_{X|Z} \\ \mathbf{m}_{Y|Z} \end{pmatrix} ; \begin{pmatrix} \mathbf{C}_{X|Z} & \mathbf{C}_{XY|Z} \\ \mathbf{C}_{YX|Z} & \mathbf{C}_{Y|Z} \end{pmatrix} \right],$$

e pela Proposição 2.7, temos que

$$\mathbf{X} \Bigg| \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} \sim \left[\mathbf{m}_{X|Z} + \mathbf{C}_{XY|Z}\mathbf{C}_{Y|Z}^{-1}(\mathbf{Y} - \mathbf{m}_{Y|Z}) \; ; \; \mathbf{C}_{X|Z} - \mathbf{C}_{XY|Z}\mathbf{C}_{Y|Z}^{-1}\mathbf{C}_{YX|Z} \right].$$

Portanto, $\mathbf{m}_{X|Z} + \mathbf{C}_{XY|Z}\mathbf{C}_{Y|Z}^{-1}(\mathbf{Y} - \mathbf{m}_{Y|Z}) = \mathbf{AY} + \mathbf{BZ} + \mathbf{D}$. Pela recíproca do teorema da projeção (Proposição 2.2), temos então, necessariamente, que $\mathbf{A} = \mathbf{C}_{XY|Z}\mathbf{C}_{Y|Z}^{-1}$, pois $\mathbf{A}$ é a constante que multiplica $\mathbf{Y}$ na parte linear do preditor. Pós multiplicando ambos os lados por $\mathbf{C}_{Y|Z}$, obtemos $\mathbf{C}_{XY|Z} = \mathbf{AC}_{Y|Z}$ e podemos reescrever

$$\mathbf{X}|\mathbf{Y}, \mathbf{Z} \sim [\mathbf{m}_{X|Z} + \mathbf{A}(\mathbf{Y} - \mathbf{m}_{Y|Z}); \mathbf{C}_{X|Z} - \mathbf{AC}_{YX|Z}].$$

Logo, $\mathbf{C}_{X|Y,Z} = \mathbf{C}_{X|Z} - \mathbf{AC}_{YX|Z}$, ou $\mathbf{C}_{X|Z} = \mathbf{C}_{X|Y,Z} + \mathbf{AC}_{YX|Z}$. Como $\mathbf{C}_{YX|Z} = \mathbf{C}_{XY|Z}^{\top}$, temos que $\mathbf{C}_{YX|Z} = \mathbf{C}_{Y|Z}\mathbf{A}^{\top}$ o que nos dá $\mathbf{C}_{X|Z} = \mathbf{C}_{X|Y,Z} + \mathbf{AC}_{Y|Z}\mathbf{A}^{\top}$. Aplicando a Definição (2.1) e observando que tanto $\mathbf{m}_{Y|Z}$ quanto $\mathbf{m}_{X|Z}$ são lineares em $\mathbf{Z}$, obtemos também, após algum desenvolvimento algébrico,

$$\begin{aligned} \mathbf{B} &= \Sigma_{XZ}\Sigma_{Z}^{-1} - \mathbf{C}_{XY|Z}\mathbf{C}_{Y|Z}^{-1}\Sigma_{XZ}\Sigma_{Z}^{-1}, \\ \mathbf{D} &= \boldsymbol{\mu}_X - \Sigma_{XY}\Sigma_{Y,Z}^{-1}\boldsymbol{\mu}_Y - \Sigma_{XZ}\Sigma_{Y,Z}^{-1}\boldsymbol{\mu}_Z. \end{aligned}$$

□

Como o preditor linear $\mathbf{X}|\mathbf{Y}, \mathbf{Z}$ é, por definição, linear no vetor $(\mathbf{Y}^{\top}, \mathbf{Z}^{\top})^{\top}$, as hipóteses da Proposição 2.9 são sempre satisfeitas. Este resultado será útil quando as constantes $\mathbf{A}$, $\mathbf{B}$ e $\mathbf{C}$ forem conhecidas.

2.2 Filtro e suavizador de Kalman

Filtros e suavizadores de Kalman aparecem em geral na literatura em conexão com processos que variam no tempo. Neste contexto, podem ser pensados como mecanismos que tornam

possível predizer um *sinal* a partir de medidas feitas com *ruído*. Para predizer o sinal, digamos s, no tempo t, um *filtro* utiliza a informação disponível até o tempo t enquanto um *suavizador* utiliza também a informação que se torna disponível após o tempo t. O filtro e o suavizador de Kalman são preditores lineares obtidos recursivamente.

Nesta seção, vamos descrever o filtro e o suavizador de Kalman para o seguinte modelo de espaço de estados

$$\mathbf{Y}_t = \mathbf{F}_t\Theta_t + \varepsilon_t, \tag{2.4}$$

$$\Theta_t = \mathbf{G}_t\Theta_{t-1} + \xi_t, \tag{2.5}$$

em que $\mathbf{Y}_t$ é um vetor de respostas de dimensão d, Θ_t, o processo latente no tempo t, é um vetor de dimensão M e $\mathbf{F}_t$ e $\mathbf{G}_t$ são matrizes não estocásticas. Todas as inovações têm esperança zero e satisfazem as seguintes condições,

$$\mathrm{Cov}(\xi_t, \xi_s) = 0, \ s \neq t; \tag{2.6}$$

$$\mathrm{Cov}(\varepsilon_t, \varepsilon_s) = 0, \ s \neq t; \tag{2.7}$$

$$\mathrm{Cov}(\xi_t, \varepsilon_s) = 0, \forall s, t; \tag{2.8}$$

$$\mathrm{Var}(\xi_t) = \lambda_t \ \text{ é diagonal}; \tag{2.9}$$

$$\mathrm{Var}(\varepsilon_t) = \Lambda_t \text{ é diagonal}; \tag{2.10}$$

$$\mathrm{Cov}(\xi_t, \Theta_{t-1}) = 0; \tag{2.11}$$

$$\mathrm{Cov}[\varepsilon_t, \Theta_t] = 0; \tag{2.12}$$

$$\mathrm{Cov}(\xi_t, \mathbf{Y}^{t-1}) = 0, \text{em que } \mathbf{Y}^t \text{ denota o vetor } (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_t^\top)^\top; \tag{2.13}$$

$$\mathrm{E}(\mathbf{Y}_{t+1}, \cdots, \mathbf{Y}_n | \Theta_{t+1}) \text{ é linear em } \Theta_{t+1}; \tag{2.14}$$

$$\mathrm{Cov}(\mathbf{Y}_{t+1}, \ldots, \mathbf{Y}_n, \Theta_t | \Theta_{t+1}) = 0. \tag{2.15}$$

A estrutura de dependência deste modelo de espaços de estados pode ser representada pelo esquema na Figura 2.1 em que as setas representam a dependência entre as variáveis e ilustram a seguinte estrutura de correlação: duas quantidades, digamos A e B, são condicionalmente independentes dado uma terceira, digamos C, se o único *caminho* ligando A a

$$
\begin{array}{ccccccccc}
 & & \mathbf{Y}_{t-1} & & \mathbf{Y}_t & & & & \mathbf{Y}_{t+s} \\
 & & \uparrow & & \uparrow & & & & \uparrow \\
\cdots \rightarrow & \Theta_{t-2} & \rightarrow & \Theta_{t-1} & \rightarrow & \Theta_t & \rightarrow \cdots \rightarrow & & \Theta_{t+s} \quad \cdots
\end{array}
$$

Figura 2.1: Estrutura de correlação de um modelo de espaço de estados

B passa por C. Por exemplo, $\mathbf{Y}_{t-1}$ e $\mathbf{Y}_t$ são condicionalmente independentes dado Θ_{t-1} ou Θ_t e, dado o processo latente Θ, as observações $\mathbf{Y}$ são independentes. Além disso, dado o processo latente, a distribuição condicional de $\mathbf{Y}_t$ depende apenas de Θ_t.

Tanto o filtro quanto o suavizador de Kalman para o modelo (2.4)–(2.5) podem ser obtidos diretamente a partir dos resultados de Harvey (1981) para a distribuição normal. No entanto, seguimos aqui a dedução proposta por Jørgensen *et al.* (1999), com o propósito de tornar mais claro o cálculo da matriz de erros quadráticos na Seção 2.2. A matriz completa de erros quadráticos será utilizada na Seção 4.5 para a obtenção da matriz de variabilidade do modelo retrospectivo.

O filtro de Kalman para o modelo (2.4)–(2.5) é o preditor linear de Θ_t, dados $\mathbf{Y}_1, \ldots, \mathbf{Y}_t$. O suavizador de Kalman é o preditor de Θ_t dada toda a amostra $\mathbf{Y} = \mathbf{Y}^n = (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_n^\top)^\top$ e é obtido a partir do filtro.

A fim de obter o suavizador de Kalman, vamos seguir os seguintes passos: primeiramente, vamos obter $\mathbf{Y}_t|\mathbf{Y}^{t-1}$ e $\Theta_t|\mathbf{Y}^{t-1}$. A partir destes dois preditores, vamos então obter $\Theta_t|\mathbf{Y}^t$ e $\left.\begin{pmatrix} \Theta_t \\ \Theta_{t+1} \end{pmatrix}\right| \mathbf{Y}^t$. Finalmente, vamos obter $\Theta_t|\mathbf{Y}$ iterativamente a partir de $\Theta_n|\mathbf{Y}^n$.

Denotando o filtro de Kalman no tempo t por $[\mathbf{m}_t, \mathbf{C}_t]$, obtemos

Proposição 2.10 *O preditor linear de Θ_t dado o passado do processo observado, $\Theta_t|\mathbf{Y}^{t-1}$, é dado por $[\mathbf{G}_t\mathbf{m}_{t-1}, \mathbf{B}_t]$, em que $\mathbf{B}_t = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t^\top + \mathrm{Var}(\boldsymbol{\xi}_t)$ e $\Theta_{t-1}|\mathbf{Y}^{t-1} \sim [\mathbf{m}_{t-1}; \mathbf{C}_{t-1}]$.*

Demonstração. Utilizando o modelo de espaços de estados, podemos escrever $\Theta_t|\mathbf{Y}^{t-1} =$

$(\mathbf{G}_t\Theta_{t-1} + \xi_t)|\mathbf{Y}^{t-1}$. De (2.11) e (2.13) temos $\mathrm{Cov}[\xi_t, \Theta_{t-1}] = \mathbf{0}$ e $\mathrm{Cov}[\xi_t, \mathbf{Y}^{t-1}] = \mathbf{0}$. Como $\mathrm{E}(\xi_t) = \mathbf{0}$, iniciando com $\mathbf{m}_0$ e $\mathbf{C}_0$, o resultado segue da Proposição 2.4. □

Proposição 2.11 *O preditor linear do processo observado dado o passado,* $\mathbf{Y}_t|\mathbf{Y}^{t-1}$, *é* $[\mathbf{f}_t, \mathbf{Q}_t]$, *em que* $\mathbf{f}_t = \mathbf{F}_t\mathbf{G}_t\mathbf{m}_{t-1}$ *e* $\mathbf{Q}_t = \mathbf{F}_t\mathbf{B}_t\mathbf{F}_t^\top + \Lambda_t$.

Demonstração. Podemos escrever $\mathbf{Y}_t|\mathbf{Y}^{t-1} = (\mathbf{F}_t\Theta_t + \varepsilon_t)|\mathbf{Y}^{t-1}$. De (2.10), (2.12) e (2.13) temos que $\mathrm{Var}(\varepsilon_t) = \Lambda_t$, $\mathrm{Cov}[\varepsilon_t, \Theta_t] = \mathbf{0}$ e $\mathrm{Cov}[\varepsilon_t, \mathbf{Y}^{t-1}] = \mathbf{0}$. Como $\mathrm{E}[\varepsilon_t] = \mathbf{0}$, o resultado segue diretamente das Proposições 2.10 e 2.4. □

Proposição 2.12 *Seja* $[\mathbf{m}_t, \mathbf{C}_t]$ *o filtro no tempo* t. *Então,*

$$\mathbf{m}_t = \mathbf{G}_t\mathbf{m}_{t-1} + \mathbf{B}_t\mathbf{F}_t^\top\mathbf{Q}_t^{-1}(\mathbf{Y}_t - \mathbf{f}_t) \quad \text{e} \quad \mathbf{C}_t = \mathbf{B}_t - \mathbf{B}_t\mathbf{F}_t^\top\mathbf{Q}_t^{-1}\mathbf{F}_t\mathbf{B}_t^\top.$$

Demonstração. Primeiramente, note que $\mathbf{Y}^t$ pode ser escrito como $(\mathbf{Y}^{t-1\top}, \mathbf{Y}_t^\top)^\top$. Então podemos escrever $\Theta_t|\mathbf{Y}^t = \Theta_t|\mathbf{Y}^{t-1}, \mathbf{Y}_t$. Da Proposição 2.2, temos que $\mathbf{f}_t \in \mathcal{V}(\mathbf{Y}^{t-1})$. Logo, $\mathcal{V}(\mathbf{Y}^{t-1}, \mathbf{Y}_t) = \mathcal{V}(\mathbf{Y}^{t-1}, \mathbf{Y}_t - \mathbf{f}_t)$. Temos também que $\mathbf{Y}^{t-1}$ e $\mathbf{Y}_t - \mathbf{f}_t$ são ortogonais e portanto, não correlacionados, uma vez que $\mathrm{E}(\mathbf{Y}_t - \mathbf{f}_t) = \mathbf{0}$. Tomando $\mathbf{X} = \Theta_t$, $\mathbf{Y} = \mathbf{Y}^{t-1}$ e $\mathbf{Z} = \mathbf{Y}_t - \mathbf{f}_t$, na notação da Proposição 2.5, obtemos $\mathbf{m}_{X|Y} = \mathbf{G}_t\mathbf{m}_{t-1}$, $\mathbf{C}_{X|Y} = \mathbf{B}_t$, $\Sigma_Z = \mathrm{Var}(\mathbf{Y}_t - \mathbf{f}_t) = \mathbf{Q}_t$, $\Sigma_{XZ} = \mathrm{Cov}[\Theta_t, \mathbf{Y}_t - \mathbf{f}_t] = \mathbf{B}_t\mathbf{F}_t^\top$ e o resultado segue. □

Proposição 2.13 *O preditor conjunto de* Θ_t, Θ_{t+1} *dado* $\mathbf{Y}^t$ *é*

$$\begin{pmatrix} \Theta_t \\ \Theta_{t+1} \end{pmatrix} \Bigg| \mathbf{Y}^t \sim \left[\begin{pmatrix} \mathbf{m}_t \\ \mathbf{G}_{t+1}\mathbf{m}_t \end{pmatrix}; \begin{pmatrix} \mathbf{C}_t & \mathbf{C}_t\mathbf{G}_{t+1}^\top \\ \mathbf{G}_{t+1}\mathbf{C}_t & \mathbf{B}_{t+1} \end{pmatrix} \right].$$

Demonstração. Basta tomar, na notação da Proposição 2.6, $\mathbf{X} = \Theta_t$, $\mathbf{Y} = \Theta_{t+1}$ e $\mathbf{Z} = \mathbf{Y}^t$. Então, $\mathbf{m}_{X|Z} = \mathbf{m}_t$, $\mathbf{C}_{X|Z} = \mathbf{C}_t$, $\mathbf{m}_{Y|Z} = \mathbf{G}_{t+1}\mathbf{m}_t$ e $\mathbf{C}_{Y|Z} = \mathbf{B}_{t+1}$. Logo,

$$\begin{aligned} \mathbf{C}_{XY|Z} &= \mathrm{Var}[\Theta_t]\mathbf{G}_{t+1}^\top - \mathrm{Cov}[\Theta_t, \mathbf{Y}^t]\mathrm{Var}(\mathbf{Y}^t)^{-1}\mathrm{Cov}[\mathbf{Y}^\top, \Theta_t]\mathbf{G}_{t+1}^\top = \\ &\quad \mathbf{C}_{X|Z}\mathbf{G}_{t+1}^\top = \mathbf{C}_t\mathbf{G}_{t+1}^\top. \end{aligned}$$

□

Proposição 2.14 *O suavizador de Kalman é* $\Theta_t|\mathrm{Y} \sim [\mathrm{m}_t^*; \mathrm{C}_t^*]$, *em que*

$$\begin{aligned}
\mathrm{m}_t^* &= \mathrm{m}_t + \mathrm{D}_t\mathrm{B}_{t+1}^{-1}(\mathrm{m}_{t+1}^* - \mathrm{G}_{t+1}\mathrm{m}_t), \\
\mathrm{C}_t^* &= \mathrm{C}_t - \mathrm{D}_t\mathrm{B}_{t+1}^{-1}\mathrm{D}_t^\top + \mathrm{D}_t\mathrm{B}_{t+1}^{-1}\mathrm{C}_{t+1}^*\mathrm{B}_{t+1}^{-1}\mathrm{D}_t^\top
\end{aligned}$$

e $\mathrm{D}_t = \mathrm{C}_t\mathrm{G}_{t+1}^\top$.

Demonstração. De (2.15), temos que Θ_t e $(\mathrm{Y}_{t+1}, \ldots, \mathrm{Y}_n)$ são não correlacionados dado Θ_{t+1} e de (2.14) temos que $\mathrm{E}(\mathrm{Y}_{t+1}, \ldots, \mathrm{Y}_n|\Theta_{t+1})$ é linear em Θ_{t+1}. Logo, pela Proposição 2.7, obtemos que $\Theta_t|\Theta_{t+1}, \mathrm{Y}^n = \Theta_t|\Theta_{t+1}, \mathrm{Y}^t, \mathrm{Y}_{t+1}, \ldots, \mathrm{Y}_n \equiv \Theta_t|\Theta_{t+1}, \mathrm{Y}^t$. Utilizamos então a Proposição 2.6 para obter o preditor linear condicional,

$$\Theta_t|\Theta_{t+1}, \mathrm{Y}^t \sim [\mathrm{m}_t + \mathrm{C}_t\mathrm{G}_{t+1}^\top\mathrm{B}_{t+1}^{-1}(\Theta_{t+1} - \mathrm{G}_{t+1}\mathrm{m}_t); \mathrm{C}_t\mathrm{G}_{t+1}^\top\mathrm{B}_{t+1}^{-1}\mathrm{G}_{t+1}\mathrm{C}_t^\top].$$

Tomando $\mathrm{X} = \Theta_t$, $\mathrm{Y} = \Theta_{t+1}$ e $\mathrm{Z} = \mathrm{Y}^n$ na notação da Proposição 2.8, temos $\mathrm{A} = \mathrm{C}_t\mathrm{G}_{t+1}^\top\mathrm{B}_{t+1}^{-1}$, $\mathrm{B} = 0$, e $\mathrm{D} = \mathrm{m}_t - \mathrm{C}_t\mathrm{G}_{t+1}^\top\mathrm{B}_{t+1}^{-1}\mathrm{G}_{t+1}\mathrm{m}_t$. Logo, podemos começar uma iteração a partir $\Theta_n|\mathrm{Y}^n$ obtido na Proposição 2.12. Denotando $\Theta_{t+1}|\mathrm{Y}^n$ por $[\mathrm{m}_{t+1}^*; \mathrm{C}_{t+1}^*]$, obtemos,

$$\begin{aligned}
\Theta_t|\mathrm{Y}^n &\sim \Big[\mathrm{C}_t\mathrm{G}_{t+1}^\top\mathrm{B}_{t+1}^{-1}m_{t+1}^* + \mathrm{m}_t - \mathrm{C}_t\mathrm{G}_{t+1}^\top\mathrm{B}_{t+1}^{-1}\mathrm{G}_{t+1}\mathrm{m}_t \\
&\quad \mathrm{C}_t - \mathrm{C}_t\mathrm{G}_{t+1}^\top\mathrm{B}_{t+1}^{-1}\mathrm{G}_{t+1}\mathrm{C}_t + \mathrm{C}_t\mathrm{G}_{t+1}^\top\mathrm{B}_{t+1}^{-1}\mathrm{C}_{t+1}^*\mathrm{B}_{t+1}^{-1}\mathrm{G}_{t+1}\mathrm{C}_t\Big] \\
&\sim \Big[\mathrm{m}_t + \mathrm{D}_t\mathrm{B}_{t+1}^{-1}(\mathrm{m}_{t+1}^* - \mathrm{G}_{t+1}\mathrm{m}_t)\,; \\
&\quad \mathrm{C}_t - \mathrm{D}_t\mathrm{B}_{t+1}^{-1}\mathrm{D}_t^\top + \mathrm{D}_t\mathrm{B}_{t+1}^{-1}\mathrm{C}_{t+1}^*\mathrm{B}_{t+1}^{-1}\mathrm{D}_t^\top\Big].
\end{aligned}$$

□

A matriz de erros quadráticos médios

Considere um processo observado $\mathrm{Y} = (\mathrm{Y}_1^\top, \ldots, \mathrm{Y}_n^\top)^\top$, com Y_t denotando um vetor de dimensão d, e um processo latente univariado $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_n)^\top$, em que θ_0 é um valor

inicial. Nesta seção, vamos obter a matriz completa de erros quadráticos médios relativa ao preditor linear de $\boldsymbol{\theta}$ dado $\mathbf{Y}$, isto é, vamos obter $\boldsymbol{\theta}|\mathbf{Y} \sim [\mathbf{m}^*; \mathbf{C}^*]$, em que $\mathbf{C}^* = \mathrm{E}\left\{(\boldsymbol{\theta} - \mathbf{m}^*)(\boldsymbol{\theta} - \mathbf{m}^*)^\top\right\}$.

Seja $[m_t^*; C_t^*]$ o preditor linear $\theta_t|\mathbf{Y}$, que pode ser obtido a partir dos resultados da Seção 2.2. Como consequência da Proposição 2.6, temos que $\mathbf{m}^* = [m_0^*, \ldots, m_n^*]$ e que a diagonal de $\mathbf{C}^*$ é dada por $(C_0^*, \ldots, C_n^*)^\top$. Para obtermos os termos fora da diagonal da matriz $\mathbf{C}^*$ precisamos calcular ainda $\mathrm{E}\{(\theta_t - m_t^*)(\theta_{t+s} - m_{t+s}^*)\}$, que denotaremos $C_{t,t+s}^*$.

Considere agora um processo latente M-variado, construído a partir de $\boldsymbol{\theta}$ de forma que $\boldsymbol{\Theta}_t = (\theta_{t-M+1}, \ldots, \theta_t)^\top$, com $\boldsymbol{\Theta}_0 = (\theta_{-M+1}, \ldots, \theta_0)$, em que $\theta_{-M+1}, \ldots, \theta_{-1}$ são valores iniciais adicionais. Suponha que este processo latente, juntamente com o processo observado $\mathbf{Y}$, satisfaça o esquema na Figura 2.1 e as condições de (2.6) a (2.15). Seja $[\mathbf{m}_t^*; \mathbf{C}_t^*]$ o preditor linear $\boldsymbol{\Theta}_t|\mathbf{Y}$. Também como consequência da Proposição 2.6, temos que $\mathbf{m}_t^* = (m_{t-M+1}^*, \ldots, m_t^*)^\top$ e que a diagonal de $\mathbf{C}_t^* = (C_{t-M+1}^*, \ldots, C_t^*)$ para $t > M-1$. Logo, podemos obter $[m_t^*; C_t^*]$ a partir de $[\mathbf{m}_t^*; \mathbf{C}_t^*]$.

Vamos denotar por $\mathbf{C}_{t,t+s}^*$ a matriz $\mathrm{E}\{(\boldsymbol{\Theta}_t - \mathbf{m}_t^*)(\boldsymbol{\Theta}_{t+s} - \mathbf{m}_{t+s}^*)\}$, de forma que $\mathbf{C}_{t,t}^* = \mathbf{C}_t^*$ e $C_{t,t+s}^* = \mathbf{C}_{t,t+s}^*(M, M)$, o M-ésimo componente de $\mathbf{C}_{t,t+s}^*$. Como $\boldsymbol{\Theta}_t$ foi construído a partir de um *empilhamento* do processo latente univariado, vários termos nas matrizes $\mathbf{C}_{t,t+s}^*$ se repetem, o que pode ser levado em conta no cálculo da matriz $\mathbf{C}^*$ a partir de $\boldsymbol{\Theta}_t|\mathbf{Y}$. Por exemplo, no caso em que n é um múltiplo de M, digamos $n = k \cdot M$, temos

$$\mathbf{C}^* = \begin{pmatrix} \mathbf{C}_M^* & \mathbf{C}_{M,2M}^* & \cdots & \mathbf{C}_{M,k\cdot M}^* \\ \mathbf{C}_{2M,M}^* & \mathbf{C}_{2M}^* & \cdots & \\ & \ddots & & \\ C_{k\cdot M,M}^* & \cdots & & \mathbf{C}_{k\cdot M}^* \end{pmatrix}. \tag{2.16}$$

Para obtermos os blocos fora da diagonal, $\mathbf{C}_{t,t+s}^*$, vamos seguir o mesmo procedimento descrito em Jørgensen *et al.* (1999) isto é, vamos utilizar o preditor $\boldsymbol{\Theta}_t|\boldsymbol{\Theta}_{t+s}, \mathbf{Y}^n$.

Temos que $\mathbf{m}_t^* \in \mathcal{V}(\mathbf{Y}^n)$ e $\boldsymbol{\Theta}_{t+s} - \mathbf{m}_{t+s}^* \perp \mathcal{V}(\mathbf{Y}^n)$, em que $\mathcal{V}(Y)$ denota o espaço linear gerado por $\mathbf{Y}$. Logo, $\mathbf{C}_{t,t+s}^* = \mathrm{Cov}(\boldsymbol{\Theta}_t - \mathbf{m}_t^*, \boldsymbol{\Theta}_{t+s} - \mathbf{m}_{t+s}^*) = \mathrm{Cov}(\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_{t+s} - \mathbf{m}_{t+s}^*)$.

Dado $\boldsymbol{\theta}$, a distribuição condicional de $\mathbf{Y}_t$ depende apenas de $\boldsymbol{\Theta}_t$, conforme a Figura 2.1. Denotando independência por $\perp\!\!\!\perp$, temos que $\boldsymbol{\Theta}_t \perp\!\!\!\perp Y_{t+s}, \ldots, Y_n|\boldsymbol{\Theta}_{t+s}, \mathbf{Y}^{t+s-1}$, e

$Y_{t+s}, \ldots, Y_n \perp\!\!\!\perp \mathbf{Y}^{t+s-1}|\Theta_{t+s}$. Logo, $\mathrm{E}\left(Y_{t+s}, \ldots, Y_n|\Theta_{t+s}, \mathbf{Y}^{t+s-1}\right) = \mathrm{E}\left(Y_{t+s}, \ldots, Y_n|\Theta_{t+s}\right)$, que é linear em Θ_{t+s} (por (2.14)) e portanto, em $(\Theta_{t+s}, \mathbf{Y}^{t+s-1})$. Logo, aplicando a Proposição 2.8, obtemos

$$\Theta_t|\Theta_{t+s-1}, \mathbf{Y}^n \equiv \Theta_t|\Theta_{t+s-1}, \mathbf{Y}^{t+s-1}.$$

Vamos denotar $\Theta_t|\mathbf{Y}^{t+s-1}$ por $[\mathbf{m}_t^{(s-1)}, \mathbf{C}_t^{(s-1)}]$. Uma vez que $\Theta_t|\mathbf{Y}^{t-1} \sim [\mathbf{G}_t\mathbf{m}_{t-1}, \mathbf{B}_t]$, obtemos da Proposição 2.7 que a parte linear de $\Theta_t|\Theta_{t+s}, \mathbf{Y}^{t+s-1}$ é dada por

$$\mathbf{m}_t^{(s-1)} + \mathbf{C}_{\Theta_t,\Theta_{t+s}|\mathbf{Y}^{t+s-1}} \cdot \mathbf{C}^{-1}_{\Theta_{t+s}|\mathbf{Y}^{t+s-1}}(\Theta_{t+s} - \mathbf{G}_{t+s}\mathbf{m}_{t+s-1}) =$$

$$\mathbf{m}_t^{(s-1)} + \mathrm{Cov}\left(\Theta_t, \Theta_{t+s} - \mathbf{G}_{t+s}\mathbf{m}_{t+s-1}\right)\mathbf{B}_{t+s}^{-1}\left(\Theta_{t+s} - \mathbf{G}_{t+s}\mathbf{m}_{t+s-1}\right).$$

Logo, a parte linear de $\Theta_t|\Theta_{t+s}, \mathbf{Y}^n$ é $A_1\Theta_{t+s} + B_1\mathbf{Y}^n + C_1$, em que

$$A_1 = \mathrm{Cov}\left(\Theta_t, \Theta_{t+s} - \mathbf{G}_{t+s}\mathbf{m}_{t+s-1}\right)\mathbf{B}_{t+s}^{-1}$$

e B_1 C_1 são constantes. Desta forma, pela Proposição 2.9, temos

$$\left.\begin{pmatrix}\Theta_t\\ \Theta_{t+s}\end{pmatrix}\right|\mathbf{Y}^n \sim \left[\begin{pmatrix}A_1\mathbf{m}_{t+s}^* + B_1\mathbf{Y}^n + C_1\\ \mathbf{m}_{t+s}^*\end{pmatrix}; \begin{pmatrix}\cdots & A_1\mathbf{C}_{t+s}^*\\ \mathbf{C}_{t+s}^*A_1^\top & \cdots\end{pmatrix}\right],$$

em que $A_1\mathbf{C}_{t+s}^* = \mathrm{Cov}(\Theta_t, \Theta_{t+s} - \mathbf{m}_{t+s}^*) = \mathrm{Cov}(\Theta_t, \Theta_{t+s} - \mathbf{G}_{t+s}\mathbf{m}_{t+s-1})\mathbf{B}_{t+s}^{-1}\mathbf{C}_{t+s}^*$ e

$$\begin{aligned}\mathrm{Cov}(\Theta_t, \Theta_{t+s} - \mathbf{G}_{t+s}\mathbf{m}_{t+s-1}) &= \mathrm{Cov}(\Theta_t, \Theta_{t+s}) - \mathrm{Cov}(\Theta_t, \mathbf{m}_{t+s-1})\mathbf{G}_{t+s}^\top = \\ &\quad \mathrm{Cov}(\Theta_t, \Theta_{t+s-1} - \mathbf{m}_{t+s-1})\mathbf{G}_{t+s}^\top.\end{aligned}$$

Para calcular $\mathrm{Cov}(\Theta_t, \Theta_{t+s-1} - \mathbf{m}_{t+s-1})$, podemos usar $\Theta_t|\mathbf{Y}^t$ e $\Theta_t|\mathbf{Y}^{t-1}$. Condicionando em Θ_{t+s-1}, temos que $Y_{t+s-1}, \ldots, Y_n \perp\!\!\!\perp \mathbf{Y}^{t+s-2}|\Theta_{t+s-1}$ e também que $Y_{t+s}, \ldots, Y_n \perp\!\!\!\perp \mathbf{Y}^{t+s-1}|\Theta_{t+s-1}$, com $\mathrm{E}(Y_{t+s}, \ldots, Y_n|\Theta_{t+s-1})$ e $\mathrm{E}(Y_{t+s-1}, \ldots, Y_n|\Theta_{t+s-1})$ lineares em Θ_{t+s-1}. Desta forma, a Proposição 2.8 implica tanto $\Theta_t|\Theta_{t+s-1}, \mathbf{Y}^n \equiv \Theta_t|\Theta_{t+s-1}, \mathbf{Y}^{t+s-1}$ quanto $\Theta_t|\Theta_{t+s-1}, \mathbf{Y}^n \equiv \Theta_t|\Theta_{t+s-1}, \mathbf{Y}^{t+s-2}$.

Utilizando o filtro e a Proposição 2.7, temos que a parte linear de $\Theta_t|\Theta_{t+s-1}, \mathbf{Y}^{t+s-1}$ é dada por

$$\mathbf{m}_t^{(s-1)} + \mathrm{Cov}(\Theta_t, \Theta_{t+s-1} - \mathbf{m}_{t+s-1})\mathbf{C}_{t+s-1}^{-1}(\Theta_{t+s-1} - \mathbf{G}_{t+s-1}\mathbf{m}_{t+s-2}).$$

Vamos denotar por A_2 a matriz $\mathrm{Cov}(\Theta_t, \Theta_{t+s-1} - \mathbf{m}_{t+s-1})\mathbf{C}_{t+s-1}^{-1}$ que aparece na expressão acima. Usando $\Theta_t|\mathbf{Y}^{t-1}$ e a Proposição 2.7, obtemos que a parte linear de $\Theta_t|\Theta_{t+s-1}, \mathbf{Y}^{t+s-2}$ é

$$\mathbf{m}_t^{(s-2)} + \mathrm{Cov}(\Theta_t, \Theta_{t+s-1} - \mathbf{G}_{t+s-1}\mathbf{m}_{t+s-2})\mathbf{B}_{t+s-1}^{-1}(\Theta_{t+s-1} - \mathbf{G}_{t+s-1}\mathbf{m}_{t+s-2}).$$

Denotando por A_3 a constante $\mathrm{Cov}(\Theta_t, \Theta_{t+s-1} - \mathbf{G}_{t+s-1}\mathbf{m}_{t+s-2})\mathbf{B}_{t+s-1}^{-1}$, temos que, pela recíproca do teorema da projeção (Proposição 2.2), as constantes A_2 e A_3 são iguais. Logo, obtemos

$$\mathrm{Cov}(\Theta_t, \Theta_{t+s-1} - \mathbf{m}_{t+s-1}) = \mathrm{Cov}(\Theta_t, \Theta_{t+s-1} - \mathbf{G}_{t+s-1}\mathbf{m}_{t+s-2})\mathbf{B}_{t+s-1}^{-1}\mathbf{C}_{t+s-1}.$$

Como $\mathrm{Cov}(\Theta_t, \Theta_{t+s-1} - \mathbf{G}_{t+s-1}\mathbf{m}_{t+s-2}) = \mathrm{Cov}(\Theta_t, \Theta_{t+s-2} - \mathbf{m}_{t+s-2})\mathbf{G}_{t+s-1}^\top$, obtemos

$$\begin{aligned}
&\mathrm{Cov}(\Theta_t, \Theta_{t+s-1} - \mathbf{m}_{t+s-1}) = \mathrm{Cov}(\Theta_t, \Theta_{t+s-2} - \mathbf{m}_{t+s-2})\mathbf{G}_{t+s-1}^\top\mathbf{B}_{t+s-1}^{-1}\mathbf{C}_{t+s-1} \\
= \;& \mathrm{Cov}(\Theta_t, \Theta_{t+s-3} - \mathbf{m}_{t+s-3})\mathbf{G}_{t+s-2}^\top\mathbf{B}_{t+s-2}^{-1}\mathbf{C}_{t+s-2}\mathbf{G}_{t+s-1}^\top\mathbf{B}_{t+s-1}^{-1}\mathbf{C}_{t+s-1} \\
= \;& \mathbf{C}_t\mathbf{G}_{t+1}^\top\mathbf{B}_{t+1}^{-1}\mathbf{C}_{t+1}\cdots\mathbf{G}_{t+s-1}^\top\mathbf{B}_{t+s-1}^{-1}\mathbf{C}_{t+s-1}.
\end{aligned}$$

Combinando este resultado com o fato

$$\mathrm{Cov}(\Theta_t, \Theta_{t+s} - \mathbf{m}_{t+s}^*) = \mathrm{Cov}(\Theta_t, \Theta_{t+s-1} - \mathbf{m}_{t+s-1})\mathbf{G}_{t+s}^\top\mathbf{B}_{t+s}^{-1}\mathbf{C}_{t+s}^*,$$

obtemos

$$\mathbf{C}_{t,t+s}^* = \mathbf{C}_t\left(\prod_{i=1}^{s-1}\mathbf{G}_{t+i}^\top\mathbf{B}_{t+i}^{-1}\mathbf{C}_{t+i}\right)\mathbf{G}_{t+s}^\top\mathbf{B}_{t+s}^{-1}\mathbf{C}_{t+s}^*.$$

Em particular, $\mathbf{C}_{t-1,t}^* = \mathbf{C}_{t,t-1}^{*\top} = \mathbf{C}_{t-1}\mathbf{G}_t^\top\mathbf{B}_t^{-1}\mathbf{C}_t^*$.

Capítulo 3

Modelo Tweedie-Tweedie

Neste capítulo, vamos apresentar o modelo de espaço de estados proposto em Jørgensen *et al.* (1996b) para dados longitudinais multivariados. Este modelo, ao qual vamos nos referir como modelo Tweedie-Tweedie, estende o modelo Poisson-gama proposto em Jørgensen *et al.* (1999) para a família Tweedie, que pertence à classe dos modelos exponenciais de dispersão (ver Apêndice A). O modelo Tweedie-Tweedie é um modelo de regressão log-linear para dados multivariados longitudinais com covariáveis variando no tempo. O vetor de respostas é um vetor d-dimensional tal que suas d componentes (que chamaremos *categorias*) não têm, necessariamente, a mesma distribuição mas refletem uma mesma tendência subjacente. O processo latente é um processo de Markov de primeira ordem e as observações são condicionalmente independentes dado o processo latente, tanto no tempo quanto entre as categorias.

Na Seção 3.1 apresentamos o modelo e na Seção 3.2 seus momentos. Na Seção 3.3 apresentamos o filtro e o suavizador de Kalman associados a esse modelo. O algoritmo de escores de Newton é apresentado na Seção 3.4 e técnicas de diagnóstico na Seção 3.5. No Exemplo 1, apresentamos brevemente uma aplicação do modelo Poisson-gama, um caso particular do modelo Tweedie-Tweeedie, a um conjunto de dados no qual a variável resposta é unidimensional. Na Seção 3.6 apresentamos a análise completa dos dados deste exemplo.

Exemplo 1 Vamos considerar o conjunto de dados analisado por Koyama (1997) e mencionado no Capítulo 1. Este conjunto de dados consiste do número diário de óbitos de pessoas com idade superior a 65 anos na cidade de São Paulo, no período de 1 de janeiro de 1991 a 31 de dezembro de 1993, dos valores diários de variáveis metereológicas (temperatura e umidade) e dos valores diários de níveis de poluição (concentrações de Dióxido de Enxôfre - SO_2, Partículas inaláveis, Monóxido de Carbono - CO, Ozônio - O_3 e Dióxido de Nitrogênio - NO_2).

Seja Y_t o número de óbitos no dia $t, t = 1, \ldots, 1095$. Vamos assumir um modelo log-linear para Y_t, satisfazendo

$$\log \mathrm{E}(Y_t) = \alpha_0 + \alpha_1 x_{1t} + \cdots + \alpha_{l_1} x_{l_1 t} + \beta_1 z_{1t} + \cdots + \beta_{l_2} z_{l_2 t} = \mathbf{x}_t^\top \boldsymbol{\alpha} + \mathbf{z}_t^\top \boldsymbol{\beta},$$

em que $\mathbf{z}_t = (z_{1t}, \ldots, z_{l_1 t})^\top$ é um vetor com os níveis de concentração dos poluentes atmosféricos e $\mathbf{x}_t = (x_{1t}, \ldots, x_{l_1 t})^\top$ é um vetor com as outras covariáveis consideradas no modelo (variáveis metereológicas, ciclos e tendências, por exemplo). Vamos admitir que as observações podem ser correlacionadas no tempo mas vamos supor que, dado um processo latente univariado $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_{1095})^\top$, as observações são independentes. Vamos assumir que a distribuição de Y_t dado $\boldsymbol{\theta}$ só depende de θ_t e satisfaz

$$Y_t | \theta_t \sim \mathrm{Po}(a_t \theta_t),$$

em que $\mathrm{Po}(\mu)$ denota a distribuição de Poisson com média μ e $a_t = \exp(\alpha_0 + \alpha_1 x_{1t} + \cdots + \alpha_{l_1} x_{l_1 t})$. Admitimos também que o processo latente $\boldsymbol{\theta}$ é um processo de Markov de ordem 1, caracterizado pela seguinte distribuição de transição,

$$\theta_t | \theta_{t-1} \sim \mathrm{Ga}\left(b_t \theta_{t-1}, \frac{\sigma^2}{\theta_{t-1}}\right),$$

em que $\mathrm{Ga}(\mu, \delta^2)$ denota a distribuição gama com média μ e coeficiente de variação δ, $b_t = \exp(\beta_1 \Delta z_{1t} + \cdots + \beta_{l_2 t} \Delta z_{l_2 t})$ e $\Delta z_{jt} = z_{jt} - z_{j,t-1}, j = 1, \ldots, l_2$, o que nos dá $\mathrm{E}(\theta_t) = \exp(\beta_1 z_{1t} + \cdots + \beta_{l_2} z_{l_2 t})$. O valor θ_0 é uma constante inicial, que, neste caso, assumiremos igual a 1.

Este modelo é um caso particular do modelo Tweedie-Tweedie que descrevemos com detalhes na próxima seção.

3.1 Modelo

Vamos considerar um modelo log-linear para uma variável resposta multivariada d-dimensional e um conjunto de h unidades amostrais independentes. No contexto do Exemplo ??, poderíamos considerar, por exemplo, o número de óbitos por faixa etária como variável resposta (d= número de faixas etárias consideradas) e um conjunto de dados semelhante para uma segunda cidade ($h = 2$).

Sejam $\mathbf{Y}_{jt} = (Y_{1jt}, \ldots, Y_{djt})^\top, t = 1, \ldots, n_j, j = 1, \ldots, h$, um vetor de observações de dimensão d tomadas em tempos igualmente espaçados e

$$\mathbf{Y} = (\mathbf{Y}_{11}^\top, \ldots, \mathbf{Y}_{1n_1}^\top, \ldots, \mathbf{Y}_{h1}^\top, \ldots, \mathbf{Y}_{hn_h}^\top)^\top$$

o vetor completo de observações. Vamos denotar ainda por $\mathbf{Y}_{ij}$ o vetor $(Y_{ij1}, \ldots, Y_{ijn_j})^\top$ e por $\mathbf{Y}_j$ o vetor $(\mathbf{Y}_{j1}^\top, \ldots, \mathbf{Y}_{jn_j}^\top)^\top$, em que $i = 1, \ldots, d$, $j = 1, \ldots, h$ e $t = 1, \ldots, n_j$. Suponha que a $\mathbf{Y}_j$ esteja associado um processo latente univariado $\boldsymbol{\theta}_j = (\theta_{j0}, \theta_{j1}, \ldots, \theta_{jn_j})^\top$, em que $\theta_{j0}, j = 1, \ldots, h$, são valores iniciais. Assumimos que, dado $\boldsymbol{\theta}_j$, as $n_j d$ componentes de $\mathbf{Y}_j$ são condicionalmente independentes e que a distribuição de $\mathbf{Y}_{jt}$ depende apenas de θ_{jt}.

Sejam $\mathbf{x}_{jt} \in \mathbb{R}^{l_1}$ e $\mathbf{z}_{jt} \in \mathbb{R}^{l_2}$ vetores distintos de covariáveis e sejam $\boldsymbol{\alpha}_{ij} \in \mathbb{R}^{l_1}$ e $\boldsymbol{\beta}_j \in \mathbb{R}^{l_2}$ os parâmetros de regressão associados a $\mathbf{x}_{jt}$ e $\mathbf{z}_{jt}$, respectivamente. Vamos supor que a distribuição da i-ésima categoria de $\mathbf{Y}_{jt}$, dado θ_{jt}, segue um modelo exponencial de dispersão da família Tweedie,

$$Y_{ijt}|\theta_{jt} \sim \mathrm{Tw}_{p_i}\left(a_{ijt}\theta_{jt}, \frac{\nu_i^2}{\theta_{jt}^{p_i-1}}\right), \tag{3.1}$$

para $i = 1, \ldots, d, t = 1, \ldots, n_j$ e $j = 1, \ldots, h$, em que $a_{ijt} = \exp(\mathbf{x}_{jt}^\top \boldsymbol{\alpha}_{ji})$ e ν_i^2 é um parâmetro de dispersão.

A notação $\mathrm{Tw}_p(\mu, \delta^2)$ em (3.1) denota o modelo exponencial de dispersão da família Tweedie com parâmetro de forma p (ver Apêndice A). Supõe-se o parâmetro p conhecido mas como ele pode variar entre as d categorias, estas podem ter diferentes distribuições. O processo latente é um processo de Markov definido pela seguinte distribuição de transição

$$\theta_{jt}|\theta_{j,t-1} \sim \mathrm{Tw}_q\left(b_{jt}\theta_{j,t-1}, \frac{\sigma^2}{\theta_{j,t-1}^{q-1}}\right) \tag{3.2}$$

para $t = 1, \ldots, n_j$, em que o parâmetro q é conhecido e satisfaz $q \geq 2$. Modelos Tweedie com $q \geq 2$ correspondem a distribuições positivas (ver Apêndice A), o que exclui, por exemplo, a distribuição normal.

Em (3.2), temos $b_{jt} = \exp(\Delta \mathbf{z}_{jt}^\top \boldsymbol{\beta}_j)$, em que $\Delta \mathbf{z}_{jt} = \mathbf{z}_{jt} - \mathbf{z}_{j,t-1}$. Por conveniência, supõe-se $\mathbf{z}_{j0} = 0$. Devido à definição de $\Delta \mathbf{z}_{jt}$, a constante não pode ser incluída no modelo por meio do processo latente.

As covariáveis em $\mathbf{x}_{jt}$ representam fatores moduladores que têm um efeito imediato em Y_{ijt}, relativo ao valor de θ_{jt} e são chamadas "covariáveis de curto–prazo". A estrutura Markoviana do processo latente permite que uma mudança no nível de $\mathbf{z}_{jt}$ no tempo t tenha um efeito em tempos subsequentes. Devido a esse efeito de propagação, as covariáveis em $\mathbf{z}_t$ são chamadas de "covariáveis de longo-prazo". Outra característica que diferencia as covariáveis de curto-prazo das de longo–prazo é que seus efeitos não variam entre as d categorias de $\mathbf{Y}_{jt}$. O processo latente pode ser interpretado como um passeio aleatório e é, essencialmente, não estacionário.

Suponhamos ainda que a cada unidade amostral esteja associado um vetor de covariáveis constantes no tempo, $\mathbf{w}_j$, e que a distribuição marginal de θ_{j0} seja dada por

$$\theta_{j0} \sim \mathrm{Tw}_r(g_j, \omega^2),$$

em que $g_j = \exp(\mathbf{w}_j^\top \boldsymbol{\gamma})$, sendo $\boldsymbol{\gamma} \in \mathbb{R}^{l_3}$ um terceiro parâmetro de regressão e ω^2 um parâmetro de dispersão que reflete a variabilidade entre as unidades amostrais. Este modelo é log-linear nas covariáveis $\mathbf{x}, \mathbf{z}$ e $\mathbf{w}$ e satisfaz

$$\log \mathrm{E}(Y_{ijt}) = \mathbf{x}_{jt}^\top \boldsymbol{\alpha}_{ij} + \mathbf{z}_{jt}^\top \boldsymbol{\beta}_j + \mathbf{w}_j^\top \boldsymbol{\gamma},$$

(ver Seção 3.2). Para uma única unidade amostral, podemos considerar $\theta_0 = g = \exp(\gamma)$, de forma que γ seja o termo constante no modelo log-linear definido em (3.1).

O parâmetro p satisfaz $p \geq 1$ ou $p \leq 0$. Para $p = 0$, temos a distribuição normal; para $p = 1$, a distribuição de Poisson; para $p = 2$ a distribuição gama e para $p = 3$, a normal inversa. Para $1 < p < 2$, temos distribuições Poisson compostas, que são distribuições mistas,

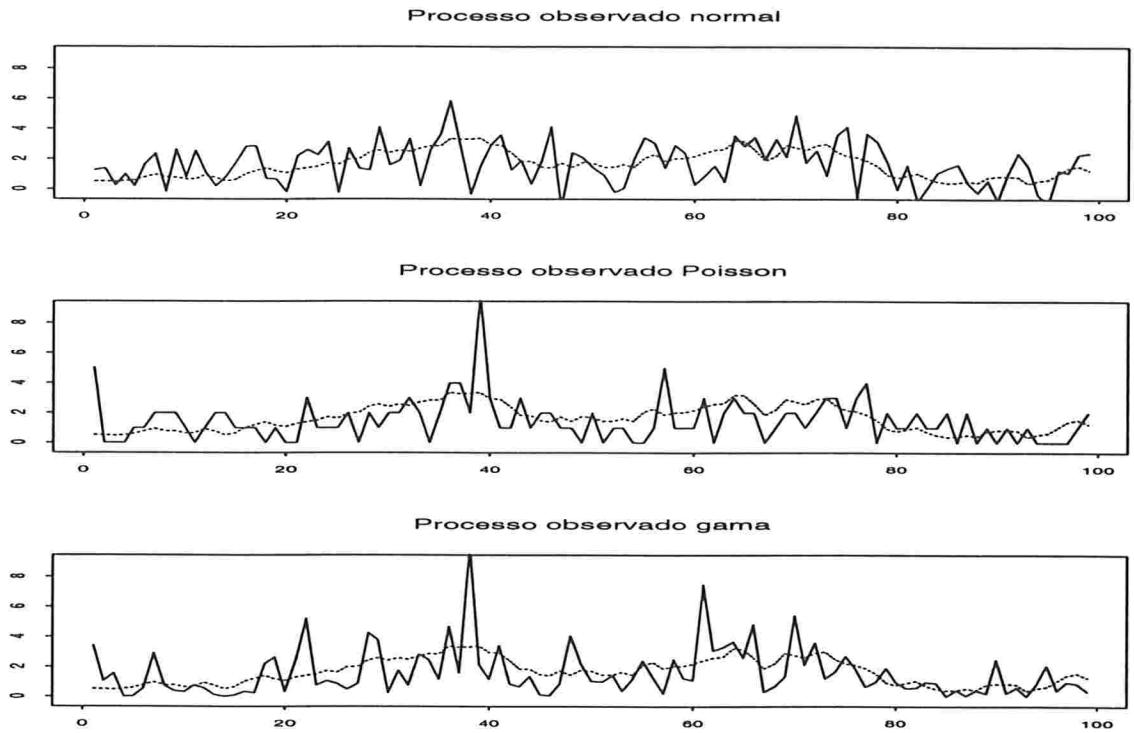

Figura 3.1: Simulações de 100 observações de um processo com 3 categorias, com $q = 2$, $p_i = 0, 1, 2$, $\nu_i^2 = 1, \sigma^2 = 0.05$ e $\theta_0 = 2$, sem efeito de covariáveis. As observações estão representadas por linhas sólidas e o processo latente, por linhas pontilhadas.

positivas, com probabilidade positiva no ponto 0 (ver Jørgensen (1997)). No caso em que $p_1 = \cdots = p_d = 1$ e $q = 2$, temos o modelo Poisson-gama de Jørgensen *et al.* (1999), com

$$Y_{it}|\theta_t \sim \mathrm{Po}(a_{it}\theta_t)$$

e

$$\theta_t|\theta_{t-1} \sim \mathrm{Ga}\left(b_t\theta_{t-1}, \frac{\sigma^2}{\theta_{t-1}}\right).$$

No Exemplo ??, temos o modelo Poisson-gama para $d = h = 1$.

3.2 Momentos

Nesta seção, vamos apresentar os momentos do modelo definido por (3.1)–(3.2). Os resultados serão utilizados na obtenção do filtro e do suavizador de Kalman na Seção 3.3.

No modelo $Y \sim \mathrm{Tw}_p(\mu, \delta^2)$, temos $\mathrm{E}(Y) = \mu$ e $\mathrm{Var}(Y) = \mu^p\delta^2$ (ver Apêndice A). Logo, de (3.1) obtemos que a esperança condicional do processo observado dado o processo latente é $\mathrm{E}(Y_{ijt}|\theta_{jt}) = a_{ijt}\theta_{jt}$ e a variância condicional é $\mathrm{Var}(Y_{ijt}|\theta_{jt}) = \nu_i^2 a_{ijt}^{p_i}\theta_{jt}$. Desta forma, dado θ_{jt}, apenas as covariáveis curto prazo influenciam Y_{ijt}.

Para o processo latente, temos que a esperança condicional de θ_{jt} dado o passado $\theta_{j,t-1}$ é $\mathrm{E}(\theta_{jt}|\theta_{j,t-1}) = b_{jt}\theta_{j,t-1}$ e a variância condicional é $\mathrm{Var}(\theta_{jt}|\theta_{j,t-1}) = \sigma^2 b_{jt}^q\theta_{j,t-1}$. Os momentos marginais do processo latente são obtidos por recursão. A média, log-linear nas covariáveis, é dada por

$$\mathrm{E}(\theta_{jt}) = \tau_{jt} = g_j b_{j1} \cdots b_{jt} = \exp(\mathbf{z}_{jt}^\top \boldsymbol{\beta}_j + \mathbf{w}_j^\top \boldsymbol{\gamma}).$$

A variância de θ_{jt} é obtida a partir de $\mathrm{Var}(\theta_{j0}) = \omega^2 g_j^r$,

$$\mathrm{Var}(\theta_{jt}) = \sigma^2\phi_{jt}\tau_{jt} + \omega^2\tau_{jt}^2 g_j^{r-2},$$

em que $\phi_{jt} = b_{jt}^{q-1} + b_{jt}b_{jt-1}^{q-1} + \ldots + b_{jt}b_{jt-1}\ldots b_{j1}^{q-1}$.

Para o processo observado, obtemos

$$\mathrm{E}(Y_{ijt}) = a_{ijt}\tau_{jt} = \exp(\mathbf{x}_{jt}^\top \boldsymbol{\alpha}_{ij} + \mathbf{z}_{jt}^\top \boldsymbol{\beta}_j + \mathbf{w}_j^\top \boldsymbol{\gamma}) \tag{3.3}$$

e a variância marginal é dada por

$$\mathrm{Var}(\mathbf{Y}_{jt}) = \Lambda_{jt}\tau_{jt} + \mathbf{a}_{jt}\mathbf{a}_{jt}^{\top}\mathrm{Var}(\theta_{jt}),$$

em que $\Lambda_{jt} = \mathrm{diag}(\nu_1^2 a_{1jt}^{p_1}, \ldots, \nu_d^2 a_{djt}^{p_d})$ e $\mathbf{a}_{jt} = (a_{1jt}, \ldots, a_{djt})^{\top}$. O segundo termo na expressão da variância mostra uma superdispersão com relação ao modelo (3.1) e mostra também que a correlação entre categorias é positiva. Embora correlação positiva entre as categorias seja uma limitação do modelo, não é difícil encontrar, na prática, conjuntos de dados para os quais esta hipótese seja adequada. A covariância entre observações de uma mesma unidade amostral, separadas por um intervalo de tempo $s > 0$, é

$$\mathrm{Cov}(\mathbf{Y}_{jt}, \mathbf{Y}_{jt+s}) = \mathbf{a}_{jt}\mathbf{a}_{jt+s}^{\top} b_{jt+1} \cdots b_{jt+s}\mathrm{Var}(\theta_{jt}).$$

A Figura 3.1 mostra simulações do modelo Tweedie-Tweedie para $q = 2$ (um processo latente gama) e três categorias, com $p_1 = 0$ (normal), $p_2 = 1$ (Poisson) e $p_3 = 2$ (gama) para uma única unidade amostral ($h = 1$), sem efeito de covariáveis.

3.3 Filtro e suavizador de Kalman

Nesta seção, vamos apresentar o modelo (3.1)–(3.2) na forma de espaço de estados e apresentar os correspondentes filtro e suavizador de Kalman para este modelo. Os detalhes dos cálculos podem ser obtidos a partir dos resultados da Seção 2.2.

No contexto do modelo Tweedie-Tweedie, o filtro de Kalman prediz uma observação futura $\mathbf{Y}_{j,t+1}$ e o valor atual do processo latente θ_{jt}, com base nas observações até o tempo t, $\mathbf{Y}_{j1}, \ldots, \mathbf{Y}_{jt}$. O suavizador prediz o processo latente θ_{jt} com base em toda a informação disponível, $\mathbf{Y}_{j1}, \ldots, \mathbf{Y}_{jn_j}$. Para simplificar a notação, vamos suprimir o índice j no que segue e denotar por n o número total de observações.

Para obter a formulação em modelo de espaço de estados, vamos antes definir as inovações para o processo latente,

$$\xi_t = \theta_t - \mathrm{E}(\theta_t|\theta_s, s < t) = \theta_t - b_t\theta_{t-1},$$

e para o processo observado,

$$\varepsilon_t = \mathbf{Y}_t - \mathrm{E}(\mathbf{Y}_t|\theta_s, s \leq t) = \mathbf{Y}_t - \mathbf{a}_t\theta_t,$$

para $t = 1, 2, \ldots, n$. O modelo pode então ser escrito na forma

$$\mathbf{Y}_t = \mathbf{a}_t\theta_t + \varepsilon_t \tag{3.4}$$

$$\theta_t = b_t\theta_{t-1} + \xi_t, \tag{3.5}$$

em que tanto ε_t e θ_t quanto ξ_t e θ_{t-1} são não correlacionados. Temos também que todas as $2n$ inovações são não correlacionadas e portanto a estrutura de correlação é a mesma do modelo de espaço de estados guassiano usual (ver, por exemplo, West e Harrison (1997)).

As variâncias das inovações são dadas por

$$\mathrm{Var}(\xi_t) = \mathrm{E}(\xi_t^2) = \mathrm{E}\left\{\mathrm{Var}(\theta_t|\boldsymbol{\theta}^{t-1})\right\} = \sigma^2 b_t^{q-1}\tau_t$$

e

$$\mathrm{Var}(\varepsilon) = \mathrm{E}(\varepsilon_t\varepsilon_t^\top) = \mathrm{E}\left\{\mathrm{Var}(\mathbf{Y}_t|\boldsymbol{\theta}^{t-1})\right\} = \Lambda_t\tau_t,$$

em que $\boldsymbol{\theta}^{t-1}$ denota $(\theta_0, \ldots, \theta_{t-1})^\top$.

Este modelo tem a estrutura de correlação ilustrada na Figura 2.1 e satisfaz as condições (2.6) a (2.15). Logo, podemos utilizar os resultados da Seção 2.2 para obter as recursões para o filtro e o suavizador de Kalman, apresentadas a seguir.

Sejam $\theta_{t-1}|\mathbf{Y}^{t-1} \sim [m_{t-1}; C_{t-1}]$ e $D_t = b_tC_{t-1} + \sigma^2 b_t^{q-1}\tau_{t-1}$, em que $\mathbf{Y}^{t-1} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_{t-1})^\top$. Então,

$$\mathbf{Y}_t|\mathbf{Y}^{t-1} \sim [\mathbf{f}_t; \mathbf{Q}_t],$$

em que $\mathbf{f}_t = \mathbf{a}_t b_t m_{t-1}$ e $\mathbf{Q}_t = b_t D_t \mathbf{a}_t\mathbf{a}_t^\top + \Lambda_t\tau_t$.

Iniciando com $m_0 = g$ e $C_0 = \omega^2 g^r$, obtemos para o processo latente,

$$\theta_t|\mathbf{Y}^t \sim [m_t; C_t],$$

em que $m_t = b_t\{m_{t-1} + D_t\mathbf{a}_t^\top\mathbf{Q}_t^{-1}(\mathbf{Y}_t - \mathbf{f}_t)\}$ e $C_t = b_tD_t(1 - b_tD_t\mathbf{a}_t^\top\mathbf{Q}_t^{-1}\mathbf{a}_t)$.

O suavizador é dado pela seguinte recursão, para $t = n-1, \ldots, 0$,

$$\theta_t | \mathbf{Y}^n \sim [m_t^*; C_t^*],$$

em que

$$m_t^* = m_t + \frac{C_t}{D_{t+1}}(m_{t+1}^* - b_{t+1} m_t) \quad \text{e} \quad C_t^* = \sigma^2 \frac{C_t}{D_{t+1}} b_{t+1}^{q-2} \tau_{t+1} + \frac{C_t^2}{D_{t+1}^2} C_{t+1}^*.$$

A recursão é iniciada com $m_n^* = m_n$ e $C_n^* = C_n$.

3.4 Estimação dos parâmetros

Consideremos novamente h unidades amostrais independentes, satisfazendo o modelo definido na Seção 3.1. Nosso objetivo é estimar os parâmetros desconhecidos $\boldsymbol{\alpha}_{ij}, \boldsymbol{\beta}_j, \boldsymbol{\gamma}, \nu_1^2, \ldots, \nu_d^2, \sigma^2$ e ω^2 para esse modelo. Sejam

$$\mathbf{m}^* = (m_{10}^*, \ldots, m_{1n_1}^*, \ldots, m_{h0}^*, \ldots, m_{hn_h}^*)^\top$$

o vetor de valores suavizados e

$$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_{11}^\top, \ldots, \boldsymbol{\alpha}_{d1}^\top, \ldots, \boldsymbol{\alpha}_{1h}^\top, \ldots, \boldsymbol{\alpha}_{dh}^\top)^\top \quad \text{e} \quad \boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_h^\top)^\top,$$

os vetores de parâmetros de regressão.

3.4.1 Parâmetros de regressão

Para obter os estimadores dos parâmetros de regressão $\boldsymbol{\alpha}, \boldsymbol{\beta}$ e $\boldsymbol{\gamma}$, Jørgensen *et al.* (1999) propuseram um método inspirado no algoritmo EM de Dempster, Laird e Rubin (1977), que descrevemos a seguir. O vetor $(\mathbf{Y}^\top, \boldsymbol{\theta}^\top)^\top$ é tratado como um vetor de observações hipotético e o processo latente $\boldsymbol{\theta}$ é tratado como a parte não observada desse vetor. A esperança condicional $\mathrm{E}(\boldsymbol{\theta}|\mathbf{Y})$ do passo E do algoritmo EM é substituída pelo suavizador de Kalman. No passo M é feita apenas uma iteração do algoritmo de escores de Newton definido na Seção 1.1.

A função de verossimilhança de $(\mathbf{Y}^\top, \boldsymbol{\theta}^\top)^\top$ fica completamente especificada pelas equações (3.1)–(3.2). Logo, caso pudéssemos observar $\boldsymbol{\theta}$, obteríamos as seguintes funções escore para os parâmetros de regressão,

$$\mathbf{u}(\boldsymbol{\alpha}_{ij}) = \sum_{t=1}^{n_j} \frac{1}{a_{ijt}^{p_i-1}} \left(Y_{ijt} - a_{ijt}\theta_{jt}\right) \mathbf{x}_{jt}, \tag{3.6}$$

$$\mathbf{u}(\boldsymbol{\beta}_j) = \sum_{t=1}^{n_j} \frac{1}{b_{jt}^{q-1}} \left(\theta_{jt} - b_t\theta_{jt-1}\right) \Delta\mathbf{z}_{jt}, \tag{3.7}$$

$$\mathbf{u}(\boldsymbol{\gamma}) = \sum_{t=1}^{h} \frac{1}{g_j^{r-1}} \left(\theta_{j0} - g_j\right) \mathbf{w}_j. \tag{3.8}$$

Seja $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)^\top$, em que $\boldsymbol{\eta}_1 = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ refere-se aos parâmetros de regressão e $\boldsymbol{\eta}_2 = (\nu_1^2, \ldots, \nu_d^2, \sigma^2, \omega^2)^\top$ refere-se aos parâmetros de dispersão. O estimador $\hat{\boldsymbol{\eta}}_1$ é definido como sendo a solução das *equações de estimação de Kalman*, $\mathcal{F}_{\eta_1}(\boldsymbol{\eta}) = \mathbf{0}$, em que $\mathcal{F}_{\eta_1}$ é um vetor com as componentes

$$\mathcal{F}_\alpha(\boldsymbol{\eta}) = \mathbf{X}^\top \mathbb{K}_a^{-1}(\mathbf{Y} - \mathbf{A}\mathbf{m}^*), \tag{3.9}$$

$$\mathcal{F}_\beta(\boldsymbol{\eta}) = \Delta\mathbf{Z}^\top \mathbb{K}_b^{-1}\mathbf{B}\mathbf{m}^*, \tag{3.10}$$

$$\mathcal{F}_\gamma(\boldsymbol{\eta}) = \mathbf{W}^\top \mathbb{K}_g^{-1}(\mathbf{m}_0^* - \mathbf{g}), \tag{3.11}$$

em que

i) $\mathbf{m}_0^* = (m_{10}, \ldots, m_{h0})^\top$ e $\mathbf{g} = (g_1, \ldots, g_h)^\top$;

ii) as matrizes $\mathbf{A}$ e $\mathbf{B}$ são definidas de forma que os elementos de $\mathbf{Y} - \mathbf{A}\mathbf{m}^*$ sejam $\mathbf{Y}_t - \mathbf{a}_t m^*$ e os elementos de $\mathbf{B}\mathbf{m}^*$ sejam $m_t^* - b_t m_{t-1}^*$;

iii) as matrizes $\mathbf{X}, \Delta\mathbf{Z}$ e $\mathbf{W}$ são matrizes de especificação contendo as covariáveis $\mathbf{x}, \Delta\mathbf{z}$ e $\mathbf{w}$, respectivamente;

iv) as matrizes $\mathbb{K}_a, \mathbb{K}_b$ e $\mathbb{K}_g$ são definidas por

$$\mathbb{K}_a = \operatorname{diag}\left\{\nu_1^2 a_{111}^{p_1-1}, \ldots, \nu_d^2 a_{dhn_h}^{p_d-1}\right\},$$

$$\mathbb{K}_b = \operatorname{diag}\left\{b_{11}^{q-1}, \ldots, b_{1n_1}^{q-1}, \ldots, b_{h1}^{q-1}, \ldots, b_{hn_h}^{q-1}\right\},$$

$$\mathbb{K}_g = \operatorname{diag}\left\{g_1^{r-1}, \ldots, b_h^{r-1}\right\}.$$

A função $\mathcal{F}_{\eta_1}$ corresponde às funções escore em (3.6),(3.7) e (3.8) na forma matricial, com o suavizador $\mathbf{m}^*$ substituindo $\boldsymbol{\theta}$.

Suponha que $\mathcal{F}_{\eta_2}$ seja uma função de estimação regular para $\boldsymbol{\eta}_2$ e seja

$$\mathrm{S}_{\eta} = \begin{pmatrix} \mathrm{S}_{\eta_1} & \mathrm{S}_{\eta_1\eta_2} \\ \mathrm{S}_{\eta_2\eta_1} & \mathrm{S}_{\eta_2} \end{pmatrix}$$

a matriz de sensibilidade associada a $\mathcal{F}_{\eta} = (\mathcal{F}_{\eta_1}^{\top}, \mathcal{F}_{\eta_2}^{\top})^{\top}$. Como a esperança marginal de $\mathbf{Y}$ não depende de $\boldsymbol{\eta}_2$, obtemos que $\mathrm{S}_{\eta_1\eta_2} = \mathbf{0}$ o que nos dá a seguinte estrutura para a inversa da matriz de sensibilidade

$$\mathrm{S}_{\eta}^{-1} = \begin{pmatrix} \mathrm{S}_{\eta_1}^{-1} & \mathbf{0} \\ \mathrm{S}^{2.1} & \mathrm{S}_{\eta_2}^{-1} \end{pmatrix},$$

em que $\mathrm{S}^{2.1} = -\mathrm{S}_{\eta_2}^{-1}\mathrm{S}_{\eta_2\eta_1}\mathrm{S}_{\eta_1}^{-1}$. Desta forma, o algoritmo de escores de Newton para os parâmetros de regressão pode ser obtido como se os parâmetros de dispersão fossem conhecidos (ver detalhes em Knudsen (1999) ou na Seção 4.5.1).

O processo de estimação dos parâmetros de regressão é iniciado ajustando-se o modelo log-linear marginal (3.1) às observações. Os valores iniciais dos parâmetros de dispersão podem ser obtidos usando-se um estimador de momentos com base neste ajuste inicial. O algoritmo para estimação de $\boldsymbol{\eta}_1$ pode ser resumido como

- Passo 0:
  obter $\boldsymbol{\eta}_1^{(0)}$ pelo modelo log-linear marginal;
  obter $\boldsymbol{\eta}_2^{(0)}$ usando um estimador de momentos.

- Passo E: Calcular $\mathbf{m}^*$.

- Passo M: atualizar $\boldsymbol{\eta}_1$ por meio de

$$\boldsymbol{\eta}_1^{(s+1)} = \boldsymbol{\eta}_1^{(s)} - \mathrm{S}_{\eta_1}^{-1}\left(\boldsymbol{\eta}^{(s)}\right)\mathcal{F}_{\eta_1}\left(\boldsymbol{\eta}^{(s)}\right).$$

- atualizar $\boldsymbol{\eta}_2$ por meio das equações (3.12), (3.13) e (3.14).

A convergência do processo de estimação pode ser monitorada de diversas maneiras. A idéia básica é interromper as iterações quando $||\boldsymbol{\eta}^{(s)} - \boldsymbol{\eta}^{(s-1)}||$ for suficientemente pequeno.

Nas aplicações, consideramos o seguinte critério,

$$\max_k \frac{\left|\eta(k)^{(s+1)} - \eta(k)^{(s)}\right|}{|\eta(k)^{(s)}|} < 10^{-4},$$

em que $\eta(k)$ denota o k-ésimo componente do vetor η. Uma outra possibilidade é considerar a matriz de informação de Godambe e interromper as iterações quando

$$(\eta^{(s)} - \eta^{(s-1)})^\top \mathbf{J}^{(s)}(\eta^{(s)} - \eta^{(s-1)})$$

for pequeno (ver Knudsen(1999) para a justificativa deste procedimento). Este critério, no entanto, é difícil de ser utilizado pois exige que a matriz de informação de Godambe seja avaliada a cada iteração, tornando o processo de estimação lento.

3.4.2 Parâmetros de dispersão

Os parâmetros de dispersão podem ser estimados por meio de *estimadores de Pearson*, calculados a partir dos quadrados dos resíduos obtidos a partir do suavizador de Kalman, as inovações estimadas (ver Seção 3.5). Estes estimadores são viciados mas o vício pode ser facilmente corrigido utilizando-se as matrizes de erros quadráticos $\mathbf{C}$ e $\mathbf{C}^*$ associadas ao filtro e ao suavizador de Kalman, respectivamente. Obtemos então os seguintes *estimadores ajustados de Pearson* para os parâmetros de dispersão,

$$\hat{\nu}_i^2 = \frac{1}{N}\sum_{j=1}^{h}\sum_{t=1}^{n_j}\frac{\hat{\varepsilon}_{ijt}^{*2}}{\tau_{jt}a_{ijt}^{p_i}} + \frac{1}{N}\sum_{j=1}^{h}\sum_{t=1}^{n_j}\frac{a_{ijt}^2 C_{jt}^*}{\tau_{jt}a_{ijt}^{p_i}}, \tag{3.12}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{j=1}^{h}\sum_{t=1}^{n_j}\frac{\hat{\xi}_{jt}^{*2}}{\tau_{jt}b_{jt}^{q-1}} + \frac{1}{N}\sum_{j=1}^{h}\sum_{t=1}^{n_j}\frac{C_{jt}^* + b_{jt}^2 C_{jt-1}^* - 2b_{jt}C_{jt-1}C_{jt}^*/D_{jt}}{\tau_{jt}b_{jt}^{q-1}}, \tag{3.13}$$

e

$$\hat{\omega}^2 = \frac{1}{h}\sum_{j=1}^{h}\frac{(m_{j0}^* - g_j)^2}{g_j^r} + \frac{1}{h}\sum_{j=1}^{h}\frac{C_{j0}^*}{g_j^r}. \tag{3.14}$$

3.5 Análise de resíduos

Uma importante característica do método proposto por Jørgensen *et al.* (1996b) é a possibilidade de se verificar a adequação do modelo por meio de uma análise de resíduos, que

podem ser definidos tanto para o processo observado quanto para o processo latente. A idéia básica é utilizar gráficos de resíduos padronizados de forma semelhante à utilizada em modelos lineares generalizados. A estrutura de correlação pode ser verificada por meio das funções de autocorrelação e autocorrelação parcial dos resíduos.

Podem-se definir tanto para o processo observado quanto para o processo latente, dois tipos de resíduos, os *resíduos marginais* e os *resíduos condicionais*. Os resíduos marginais são as diferenças entre os valores observados e a média marginal estimada. Os resíduos condicionais são definidos como sendo as predições das inovações ε e ξ, calculadas a partir do filtro ou do suavizador de Kalman. Todos os resíduos têm média 0 mas nem todos são não correlacionados. A estrutura de correlação dos resíduos será apresentada em detalhes na Seção 4.6 e será omitida aqui. A definição dos resíduos condicionais para o processo observado que apresentamos a seguir difere daquela proposta por Jørgensen *et al.* (1996b) mas conserva as mesmas propriedades e interpretação.

As predições das inovações com base no filtro de Kalman são

$$\hat{\varepsilon}_{jt} = \mathbf{Y}_{jt} - \mathbf{a}_{jt} m_{jt} \quad \text{e} \quad \hat{\xi}_{jt} = m_{jt} - b_{jt} m_{jt}$$

e as predições com base no suavizador de Kalman são

$$\hat{\varepsilon}^*_{jt} = \mathbf{Y}_{jt} - \mathbf{a}_{jt} m^*_{jt} \quad \text{e} \quad \hat{\xi}_{jt} = m^*_{jt} - b_{jt} m^*_{jt}.$$

Os resíduos condicionais com base no filtro de Kalman são não correlacionados e devido a isso são bastante úteis para construção de gráficos de resíduos. Os resíduos padronizados (variância unitária) serão denotados $\hat{\bar{\xi}}_{jt}$, etc. Os resíduos $\hat{\varepsilon}_{jt}$ são padronizados componente a componente.

Diferentes tipos de gráficos de resíduos podem ser utilizados para investigar várias características do modelo. Temos,

- *Gráficos da função de autocorrelação e da função de autocorrelação parcial dos resíduos condicionais $\hat{\bar{\xi}}_{jt}$ e $\hat{\bar{\varepsilon}}_{jt}$*

  São utilizados para verificar a estrutura de correlação. Como esses resíduos são não correlacionados, estes gráficos devem apresentar um padrão de "ruído branco".

- *Diagrama de dispersão dos resíduos condicionais padronizados em função do logaritmo dos valores ajustados*

  É útil na verificação da forma da função de variância. Este gráfico pode ser construído para os resíduos do processo observado e para resíduos do processo latente. Uma forma de "megafone" indica que o parâmetro de forma na função de variância correspondente está incorreto (p_i e q). No caso de várias unidades amostrais, deve-se utilizar os resíduos $m^*_{j0} - g_j$ para verificar a forma da distribuição de θ_{j0}, pois os resíduos calculados a partir do filtro de Kalman neste caso são constantes.

- *Diagramas de dispersão dos resíduos padronizados baseados no suavizador em função de cada uma das covariáveis*

  São utilizados para detectar não linearidade.

- *Diagrama de dispersão do logaritmo dos valores observados em função do logaritmo das predições baseadas no suavizador* $\left(\log(Y_{ijt})\ vs\ \log(a_{ijt}m^*_{jt})\right)$

  É utilizado para verificar a adequação da ligação logarítmica. Este gráfico deve apresentar uma relação linear. Uma curvatura ou outra forma indica que a hipótese de ligação logarítmica é inadequada.

- *Gráfico dos resíduos marginais do processo latente ao longo do tempo (gráfico de série temporal)*

  Este gráfico é útil para identificar covariáveis não incluídas no modelo pois o processo latente tende a absorver o efeito destas covariáveis. É especialmente útil para detectar efeitos sistemáticos como ciclos e tendências.

- *Gráficos das funções de correlação cruzada entre os resíduos e as covariáveis*

  Estes gráficos podem ajudar a identificar possíveis defasagens das covariáveis a serem incluídas no modelo.

3.6 Poluição e mortalidade na cidade de São Paulo

Nesta seção, vamos utilizar o modelo Poisson-gama, que corresponde ao modelo Tweedie-Tweedie com $p = 1$ e $q = 2$, para analisar os dados descritos no Exemplo 1.

3.6.1 Introdução

A associação entre poluição atmosférica e saúde pública na cidade de São Paulo tem sido objeto de vários estudos e políticas públicas. Sabe-se que condições climáticas e poluição atmosférica estão bastante relacionados, principalmente devido à *inversão térmica*, um fenômeno que ocorre com baixas temperaturas dificultando a dispersão de poluentes. Devido a isso, embora se pense que a emissão de poluentes do ar seja mais ou menos uniforme durante o ano, a poluição do ar tende a ser maior no inverno.

Neste estudo, desejamos examinar a relação entre poluição atmosférica e mortalidade de pessoas maiores de 65 anos na cidade de São Paulo. Este grupo etário responde por cerca de 42% do total de mortes; aproximadamente 11% dessas mortes tiveram causas respiratórias no período investigado.

Os dados que analisamos foram coletados antes da introdução do sistema de rodízio pela prefeitura de São Paulo e incluímos no estudo tanto as variáveis climáticas quanto variáveis associadas à poluição.

Este conjunto de dados foi analisado anteriormente por Singer e Conceição (1994) que usaram um modelo de regressão normal e por Koyama (1997) que usou o modelo log-linear para a distribuição de Poisson, proposto em Zeger (1988).

No nosso caso, seguindo Jørgensen *et al.* (1999), usamos as variáveis metereológicas como variáveis curto-prazo e as de poluição como longo-prazo. Desta forma, o processo latente pode ser interpretado como um efeito potencial da poluição na mortalidade na população. A estrutura Markoviana do processo latente permite, então, que uma elevação no nível dos poluentes atmosféricos em um dado dia tenha um efeito em dias subseqüentes. Incluímos um ciclo anual e efeitos defasados de covariáveis. A inclusão desses efeitos foi sugerida pela

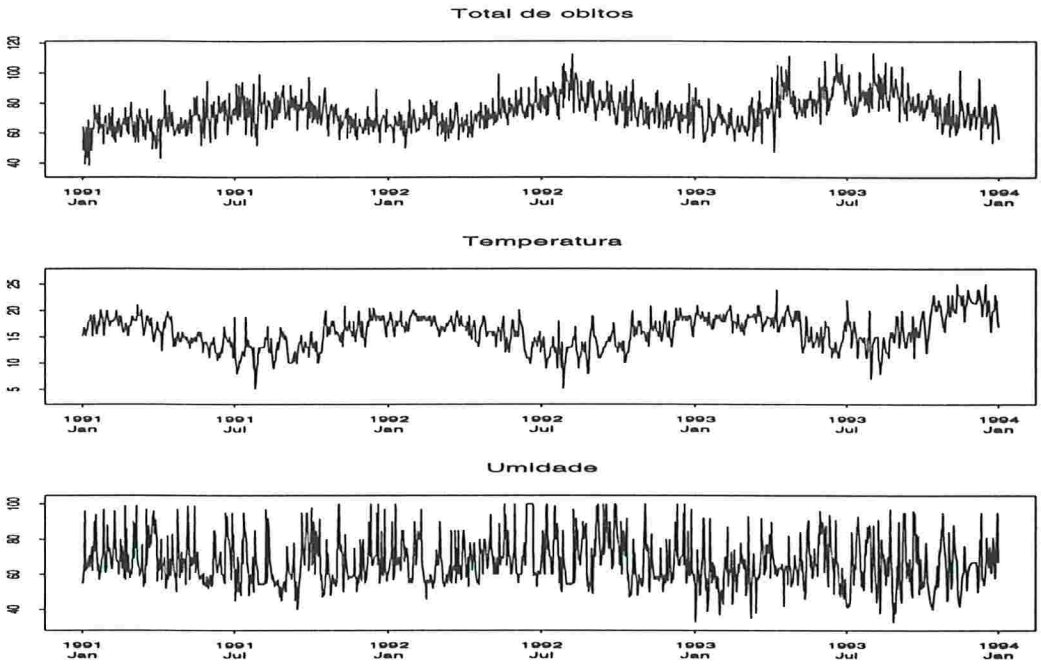

Figura 3.2: Evolução do número de óbitos de pessoas com mais de 65 anos de idade, da temperatura e da umidade na cidade de São Paulo entre 1 de janeiro de 1991 a 31 de dezembro de 1993.

própria análise, por meio das técnicas de diagnóstico descritas na Seção 3.5. Nosso modelo final difere dos modelos obtidos nas análises anteriores, embora todos eles tenham encontrado efeitos significantes da concentração de poluentes na mortalidade. A inclusão de um ciclo anual no nosso modelo permitiu a retirada de vários fatores considerados nas outras análises, reduzindo consideravelmente o número de parâmetros estimados e melhorando o ajuste.

3.6.2 Os dados

As variáveis meteorológicas que consideramos são temperatura (°C) e umidade (%); as variáveis de poluição atmosférica são Enxôfre - SO_2, ($\mu g/m^3$), Partículas (total de partículas em suspensão - PM_{10}), Monóxido de Carbono - CO (ppm), Ozônio - O_3 (ppb) e Dióxido de Nitrogênio - NO_2, ($\mu g/m^3$). Todas as covariáveis são as médias diárias de medidas tomadas em diferentes localidades da cidade.

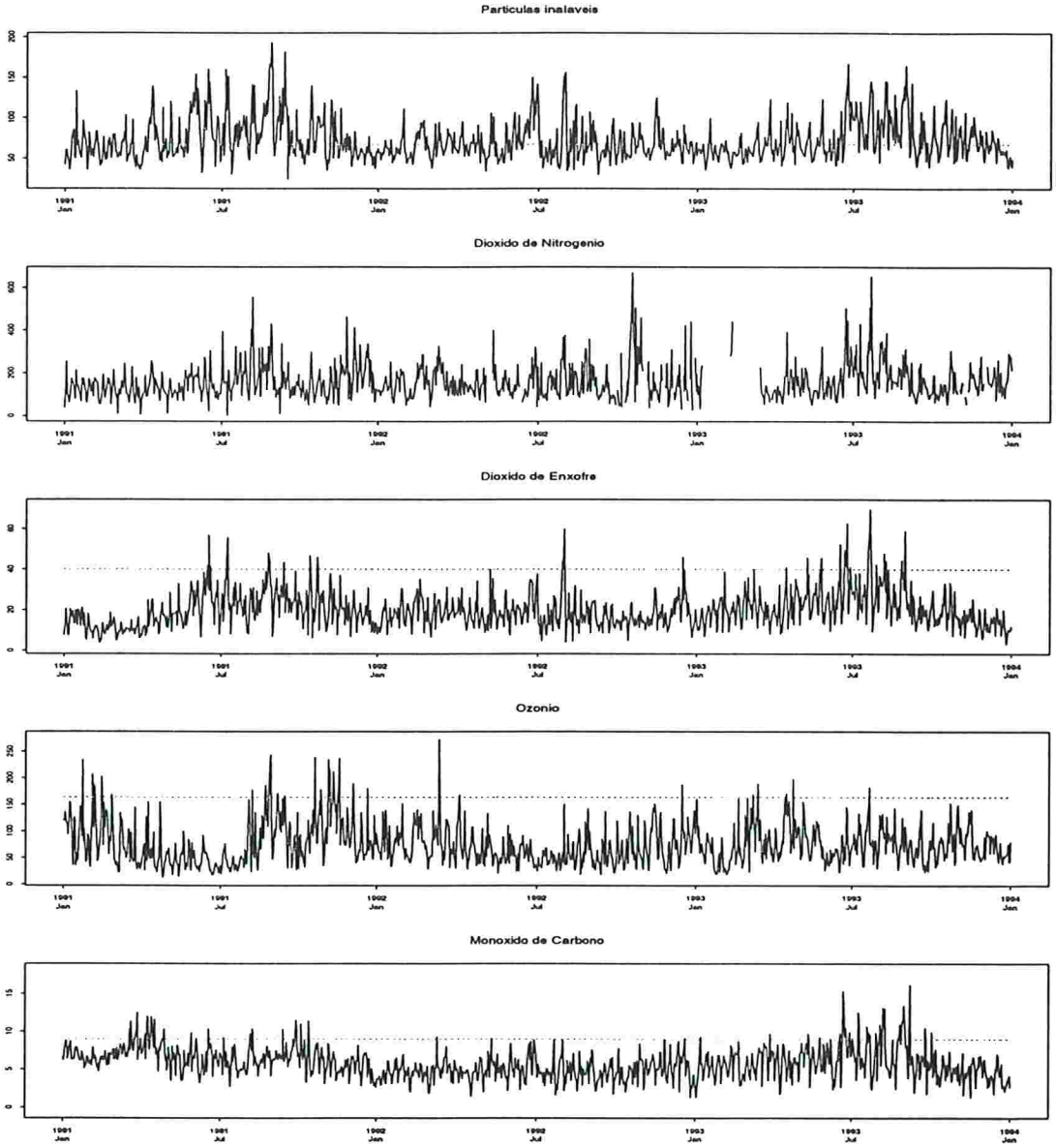

Figura 3.3: Evolução dos níveis de poluentes atmosféricos no período de 1 de janeiro de 1991 a 31 de dezembro de 1993 na cidade de São Paulo. As linhas pontilhadas marcam o nível secundário de qualidade do ar.

A Figura 3.2 mostra a série do número de mortes e as das variáveis metereológicas. A Figura 3.3 mostra as séries dos níveis dos poluentes atmosféricos, juntamente com o *nível secundário de qualidade do ar*, estabelecido pelo IBAMA (não disponível para NO_2). Abaixo deste nível, não se espera efeito nocivo da poluição atmosférica para a saúde ou para o meio ambiente.

Algumas das séries de poluentes atmosféricos apresentam falhas. Pequenas falhas foram substituídas por valores obtidos por meio de interpolação linear. A série de NO_2 apresenta uma falha de cem observações, que foram preenchidas usando um preditor baseado em um modelo ARIMA ajustado às observações anteriores à falha. No entanto, entendemos que uma falha de cem observações é grande demais para ser corrigida de forma razoável e portanto, os resultados relativos a essa covariável devem ser interpretados com certo cuidado.

Descrevemos a seguir todos os passos da análise de forma a ilustrar o uso das técnicas de diagnóstico disponíveis. Um resumo de todos os modelos considerados é apresentado nas Tabelas 3.1 e 3.2.

3.6.3 Análise

Começamos a análise com o mesmo conjunto de covariáveis considerado por Singer e Conceição (1994) e Koyama (1997) que apresentam uma detalhada descrição dos dados. A partir do conjunto de variáveis regressoras consideradas por estes autores, incluímos algumas covariáveis e excluímos outras utilizando os procedimentos de diagnóstico descritos na Seção 3.5. Utilizamos os gráficos de resíduos, em especial os gráficos das funções de autocorrelação e autocorrelação parcial dos resíduos para verificar a adequação do modelo; utilizamos testes de Wald para excluir do modelo uma covariável ou conjunto de covariáveis; utilizamos a função de correlação cruzada dos resíduos com as covariáveis para identificar possíveis efeitos defasados destas covariáveis na variável resposta. Uma vez que o processo latente tende a absorver efeitos não considerados no modelo, utilizamos também o gráfico do processo latente estimado em função do tempo. Este gráfico se mostrou bastante útil na identificação do ciclo anual que acabamos por incorporar ao modelo final.

Modelo 1

Neste primeiro modelo, incluímos fatores construídos a partir das variáveis metereológicas a fim de incorporar possíveis efeitos sazonais. Os valores utilizados para a construção destes fatores correspondem aos primeiro e terceiro quartis da distribuição da variável correspondente.

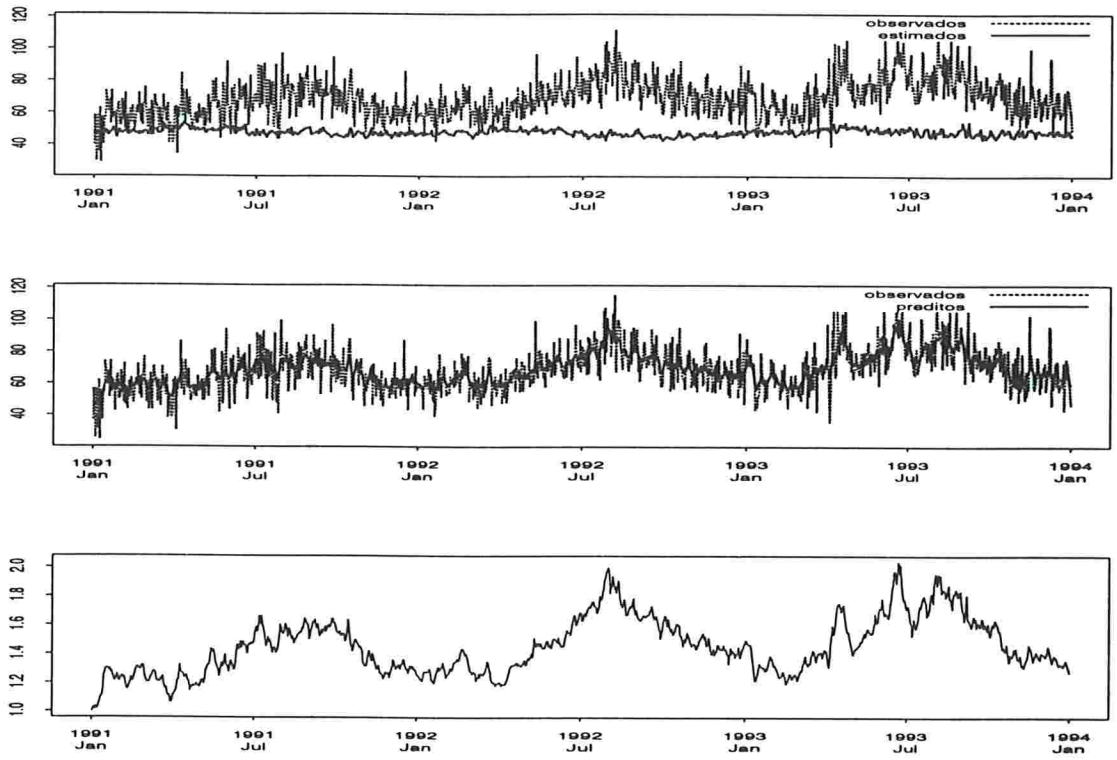

Figura 3.4: Valores observados e valores ajustados, valores observados e valores preditos e processo latente estimado para o modelo 1.

Consideramos as seguintes covariáveis,

- Variáveis Climáticas
  - temperatura (temp);
  - umidade (umid).

- Fatores

  - dias úmidos (umid): 1, se a umidade estiver acima de 75%, 0 caso contrário;
  - dias secos (Dsecos) : 1, a umidade estiver abaixo de 57%, 0 caso contrário;
  - dias quentes (Dquent): 1, se a temperatura estiver acima de 18°C, 0 caso contrário;
  - dias frios (Dfrios): 1, se a temperatura estiver abaixo de 14°C, 0 caso contrário;
  - estações do ano: níveis: verão, primavera (Prim), outono (Out) e inverno (Inv).

- Variáveis relacionadas com a concentração de poluentes atmosféricos, a saber

  - Partículas inaláveis (Part) ;
  - Dióxido de Nitrogênio (NO_2);
  - Enxôfre (SO_2);
  - Ozônio (O_3) ;
  - Monóxido de Carbono (CO).

Consideramos todas as variáveis relacionadas com a poluição como sendo variáveis de longo prazo e todas as outras como sendo de curto prazo, isto é, consideramos o seguinte modelo,

$$\begin{aligned}\log \mathrm{E}(Y_t) &= \alpha_0 + \alpha_1 \mathrm{temp} + \alpha_2 \mathrm{umid} + \alpha_3 \mathrm{Dsecos} + \alpha_4 \mathrm{Dumid} + \alpha_5 \mathrm{Dquent} + \alpha_6 \mathrm{Dfrios} + \\ &\quad \alpha_7 \mathrm{Prim} + \alpha_8 \mathrm{Out} + \alpha_8 \mathrm{Inv} + \beta_1 \mathrm{Part} + \beta_2 \mathrm{NO_2} + \beta_3 \mathrm{SO_2} + \beta_4 \mathrm{O_3} + \beta_5 \mathrm{CO}\end{aligned}$$

Embora nenhum dos efeitos tenha sido individualmente significante (ver Tabela 3.1), o valor da estatística de Wald para o conjunto de variáveis de poluição foi 9.53 com 5 graus de liberdade, correspondendo a um nível descritivo de 0.09, sugerindo que existe um efeito da poluição na mortalidade.

A Figura 3.4 mostra o gráfico da série de valores ajustados, da série de valores preditos (calculados a partir do suavizador) e o processo latente estimado (suavizador de Kalman).

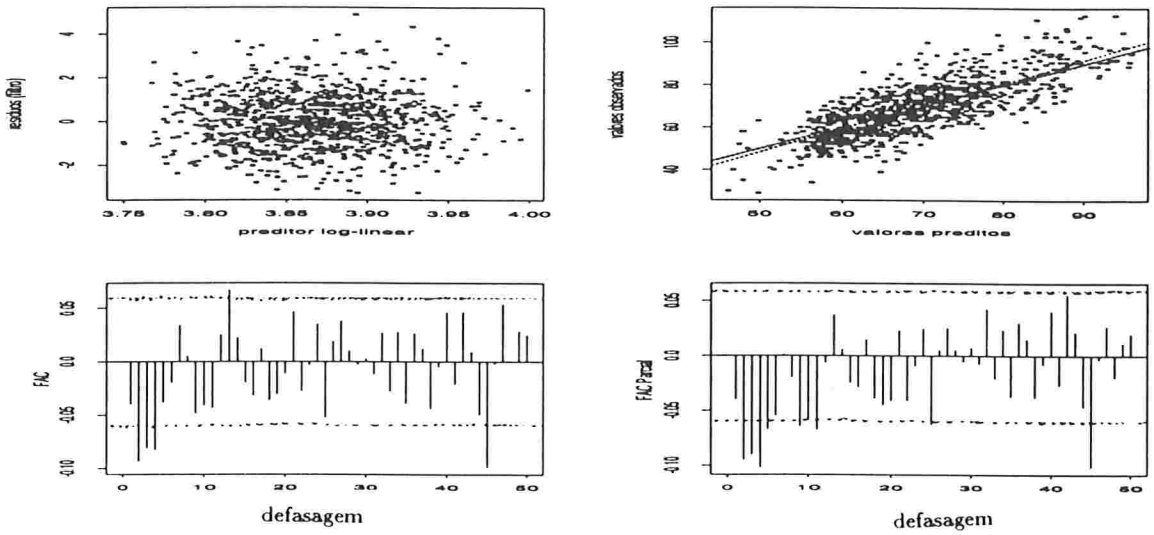

Figura 3.5: Gráficos de resíduos do modelo 1.

A inclusão dos fatores correspondentes às estações do ano não parece ter sido suficiente para explicar a sazonalidade presente nos dados, e esta sazonalidade acabou por ser incorporada pelo processo latente. Desta forma, o modelo subestima o valor esperado do número de mortes embora os valores preditos mostrem um ajuste bem melhor. Além disso, o gráfico da série do processo latente estimado sugere um ciclo anual e uma tendência.

No canto superior esquerdo da Figura 3.5, temos o gráfico de dispersão dos resíduos condicionais em função do preditor log-linear. Este gráfico não mostra nenhum padrão em particular, sugerindo que a forma da função de variância (e portanto, a distribuição da resposta) é adequada. No canto superior direito temos o gráfico dos valores preditos em função dos valores observados, juntamente com a reta de mínimos quadrados e a reta $y = x$. Nos cantos inferiores esquerdo e direito temos, respectivamente, os gráficos da função de autocorrelação (FAC) e da função de autocorrelação parcial dos resíduos padronizados, calculados a partir do filtro. Ambos apresentam valores acima do limite assintótico de 95% para as primeiras defasagens (linha pontilhada nos gráficos), indicando que o modelo

proposto ainda não é adequado aos dados. A fim de tornar mais clara a vizualização da função de autocorrelação, excluímos a primeira defasagem deste gráfico (defasagem 0).

Modelo 2

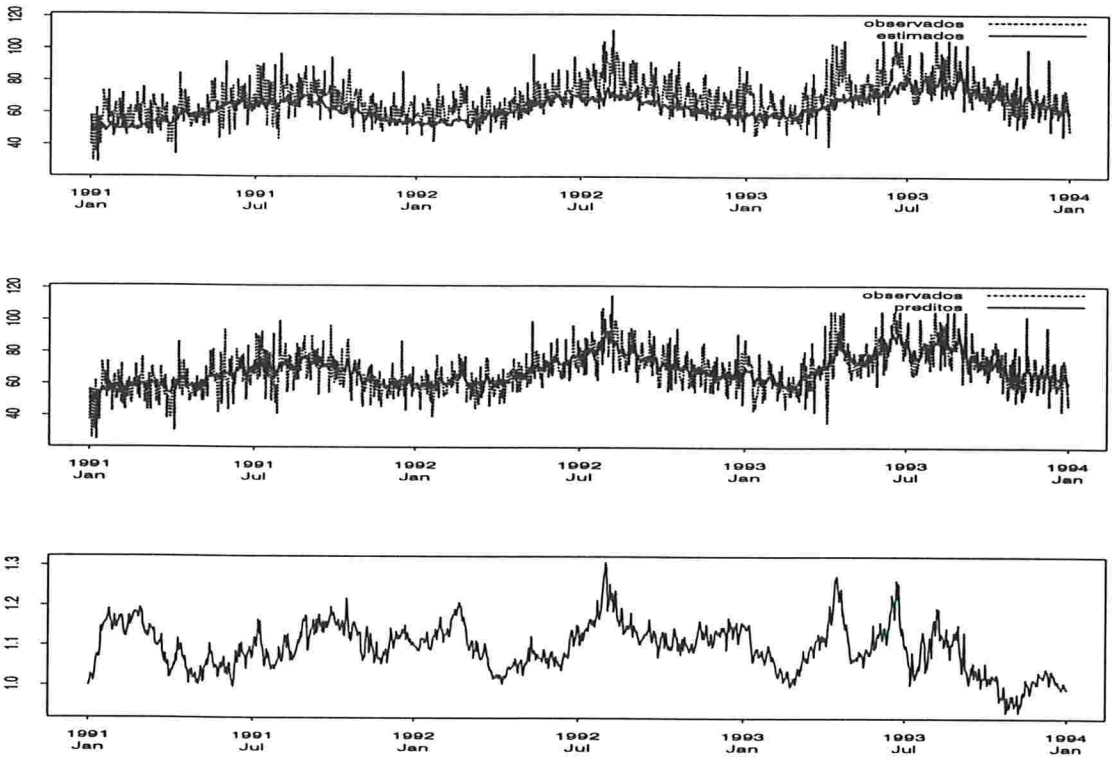

Figura 3.6: Valores observados e estimados, valores observados e preditos e processo latente estimado para o modelo 2.

Ao conjunto de covariáveis considerado no modelo 1, incluímos termos para tendência, t, e para um ciclo anual, $\cos(2\pi t/365)$ e $\text{sen}(2\pi t/365)$ como covariáveis de curto-prazo. Os valores ajustados na Figura 3.6 mostram um ajuste melhor que o do modelo anterior. A Figura 3.7 mostra os gráficos de resíduos para este modelo. Tanto a FAC quanto a FAC parcial para os resíduos baseados no filtro apresentam correlações acima do limite assintótico de 95% apenas a partir do lag 45. Os outros gráficos não apresentam comportamento inesperado.

Para este modelo, os efeitos individuais das concentrações de poluentes foram não significantes a 5%, mas o valor da estatística de Wald para o subvetor de parâmetros correspondentes às 5 variáveis associadas à poluição, foi 16.62 com 5 graus de liberdade (nível descritivo 0.005).

Nenhuma das covariáveis climáticas foi individualmente significante a 5%. No entanto, o efeito de umidade é significante se consideramos o efeito combinado das covariáveis *umid*, *Dumid* e *Dsecos*. A estatística de Wald para o conjunto destas três covariáveis foi 8.3 com 3 graus de liberdade, correspondendo a um nível descritivo de 0.04. Da mesma forma, temos que o efeito de temperatura é significante, se considerarmos em conjunto as covariáveis *temp*, *Dquent* e *Dfrios*. A estatística de Wald para este caso foi 12.8 correspondendo a um nível descritivo de 0.005. Este resultado não é surpreendente uma vez que os fatores *Dsecos* e *Dumid* foram construídos a partir da covariável *umid* e os fatores *Dfrios* e *Dquent* foram construídos a partir da covariável *temp* o que pode ser responsável por multicolinearidade entre estas covariáveis. Um teste de Wald para todos os fatores juntos resultou num valor de 6.6 com 7 graus de liberdade, correspondendo a um nível descritivo de 0.47, o que nos levou a retirar esses fatores do modelo.

Modelo 3

Neste modelo, consideramos as covariáveis utilizadas no modelo anterior, excluindo todos os 7 fatores. Como covariáveis de curto-prazo, consideramos o ciclo anual, temperatura, umidade e tendência; como covariáveis de longo-prazo, todas as covariáveis de poluição.

Tanto temperatura quanto umidade são altamente significantes e o teste de Wald para as variáveis de poluição também é significante (nível descritivo 0.01). O processo latente estimado e os gráficos de resíduos deste modelo não diferem muito dos obtidos no modelo anterior, indicando que a exclusão dos 7 fatores não prejudicou o ajuste.

Tabela 3.1: Coeficientes estimados e erros padrões para os modelos 1 a 3

Efeito	Modelo 1		Modelo 2		Modelo 3	
curto-prazo	**estimativa**	**(e.p)**	**estimativa**	**(e.p.)**	**estimativa**	**(e.p.)**
constante	3.79	(0.13)	4.00	(0.10)	4.003	(0.079)
temp	0.0067	(0.0043)	0.0060	(0.0033)	0.0062	(0.0023)
umid	−0.00081	(0.00085)	−0.00090	(0.00069)	−0.00099	(0.00034)
$\cos(2\pi t/365)$	-	-	−0.123	(0.047)	−0.131	(0.045)
$\text{sen}(2\pi t/365)$	-	-	−0.073	(0.047)	−0.046	(0.045)
Dsecos	−0.00007	(0.017)	−0.0008	(0.014)	-	-
Dumid	−0.003	(0.022)	−0.003	(0.018)	-	-
Dquent	0.014	(0.018)	0.020	(0.014)	-	-
Dfrios	0.015	(0.021)	0.016	(0.016)	-	-
Prim	−0.020	(0.058)	−0.037	(0.036)	-	-
Out	0.057	(0.057)	0.024	(0.036)	-	-
Inv	0.013	(0.067)	−0.019	(0.042)	-	-
tendência	-	-	0.00019	(0.00052)	0.00020	(0.00054)
longo-prazo						
partículas	0.00021	(0.00036)	0.00028	(0.00028)	0.00030	(0.00028)
NO_2	0.000029	(0.000085)	0.000027	(0.000066)	0.000021	(0.000066)
SO_2	0.0003	(0.0010)	0.00041	(0.00079)	0.00036	(0.00078)
O_3	0.00012	(0.00015)	0.00012	(0.00012)	0.00011	(0.00012)
CO	0.0053	(0.0036)	0.0046	(0.0028)	0.0040	(0.0028)
$\hat{\sigma}^2$	0.00203		0.00029		0.00031	

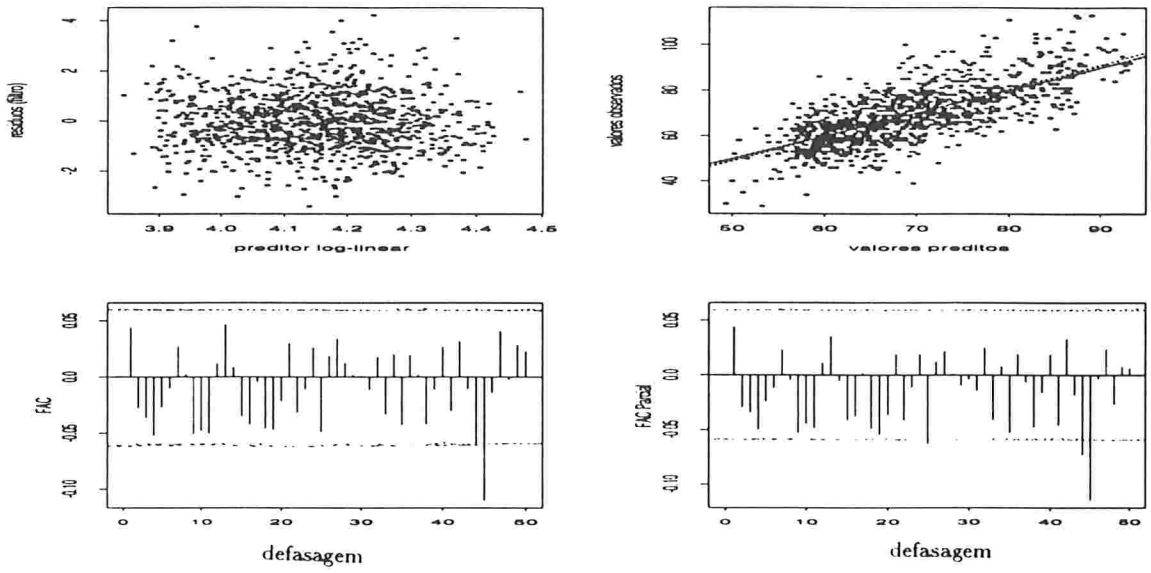

Figura 3.7: Gráficos de resíduos para o modelo 2.

Modelo 4

Como o processo latente tende a absorver efeitos não incluídos no modelo, continuamos a análise buscando possíveis efeitos de valores defasados das covariáveis. Para isso, utilizamos a função de correlação cruzada entre os resíduos padronizados e as covariáveis. Os gráficos (não apresentados aqui), sugerem possíveis efeitos defasados de temperatura e SO_2. Ao conjunto de covariáveis considerado no modelo anterior, incluímos as defasagens de 1, 2 e 3 dias de temperatura (def_1 a def_3) como covariáveis curto-prazo e as defasagens de 1,2 e 3 dias de SO_2 como covariáveis longo-prazo. A defasagem de três dias de SO_2 (def_3) é significante e a estimativa de seu efeito é negativa. A estatística de Wald para as outras covariáveis de poluição foi 3.54 com 4 graus de liberdade, correspondendo a um nível descritivo de 0.47. Novamente, os gráficos de resíduos não sugerem afastamento das hipóteses do modelo (Figura 3.8).

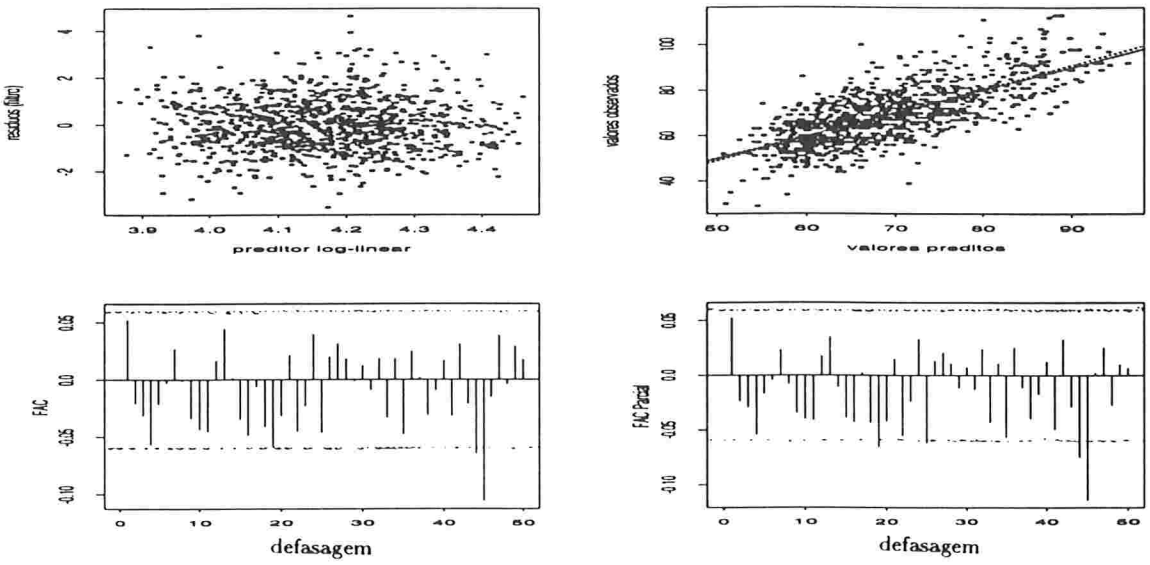

Figura 3.8: Gráficos de resíduos para o modelo 4.

Modelo 5

Consideramos como covariveis de curto-prazo $\cos(2\pi t/365)$, $\text{sen}(2\pi t/365)$, temperatura, defasagens de 1 a 3 dias de temperatura, umidade e tendência; como covariáveis longo-prazo, apenas SO_2 e defasagens de SO_2. Os gráficos de resíduos não apresentam diferenças do modelo anterior, indicando que a exclusão do modelo das demais concentrações de poluentes atmosféricos não acarretou perda de qualidade do ajuste.

A covariável SO_2 e sua defasagem de 3 dias são altamente significantes e seus coeficientes têm sinais opostos. A estatística de Wald para todas as defasagens de SO_2 em conjunto foi 13.813 com 3 graus de liberdade, correspondendo a um nível descritivo de 0.0079. Logo, a covariável SO_2, na ausência das outras covariáveis de poluição, é altamente significante e seu efeito no número de mortes num determinado dia depende não só do seu valor nesse dia como também dos valores nos últimos três dias.

Com o intuito de entender melhor o efeito do poluente SO_2, consideramos a seguinte transformação linear de posto completo desta covariável e suas defasagens,

$F_1(t) = \frac{1}{4}\{SO_2(t) + SO_2(t-1) + SO_2(t-2) + SO_2(t-3)\}$;

$F_2(t) = SO_2(t) - SO_2(t-1)$;

$F_3(t) = \frac{1}{2}\{ SO_2(t) + SO_2(t-1)\} - SO_2(t-2)$;

$F_4(t) = \frac{1}{3}\{ SO_2(t) + SO_2(t-1) + SO_2(t-2)\} - SO_2(t-3)$,

para $t = 4, \ldots, 1095$.

Esta transformação não é ortogonal, mas preserva toda a informação contida nas covariáveis originais além de ser de fácil interpretação. O efeito de F_1 pode ser interpretado como sendo o efeito cumulativo do valor atual e dos valores dos últimos três dias, do nível de SO_2. Variações no nível de SO_2 neste período são medidas por F_2 até F_4. Os efeitos individuais de F_1 a F_4 são mostrados na Tabela 3.2, sob o modelo 5. Apenas F_4 é individualmente significante. O teste de Wald para as covariáveis F_1, F_2 e F_3, resultou num valor de 1.243 com 3 graus de liberdade, correspondendo a um nível descritivo de 0.74, o que nos levou a retirar estas covariáveis do modelo.

Modelo 6 - Modelo final

Consideramos como covariável longo-prazo apenas F_4. Tanto a FAC quanto a FAC parcial para este modelo apresentaram valores acima do limite assintótico de 95% para as defasagens 44 e 45, um efeito também presente nos outros modelos.

Nenhum gráfico de resíduos para este modelo apresentou diferença considerável relativamente aos gráficos do modelo 4, que incluía todas as covariáveis associadas à poluição além das defasagens de SO_2 e temperatura (ver Figura 3.10). Além disso, a função de correlação cruzada entre os resíduos padronizados e todas as covariáveis de poluição não sugerem evidência de correlação residual. Desta forma, o efeito da poluição na mortalidade parece ser bem explicada por SO_2 apenas.

A covariável F_4 é altamente significante e seu coeficiente é positivo. Logo, o efeito da covariável SO_2 pode ser explicado pela diferença entre a média dos últimos dias e o valor de quatro dias atrás. Desta forma, o número esperado de mortes não é afetado se o nível de

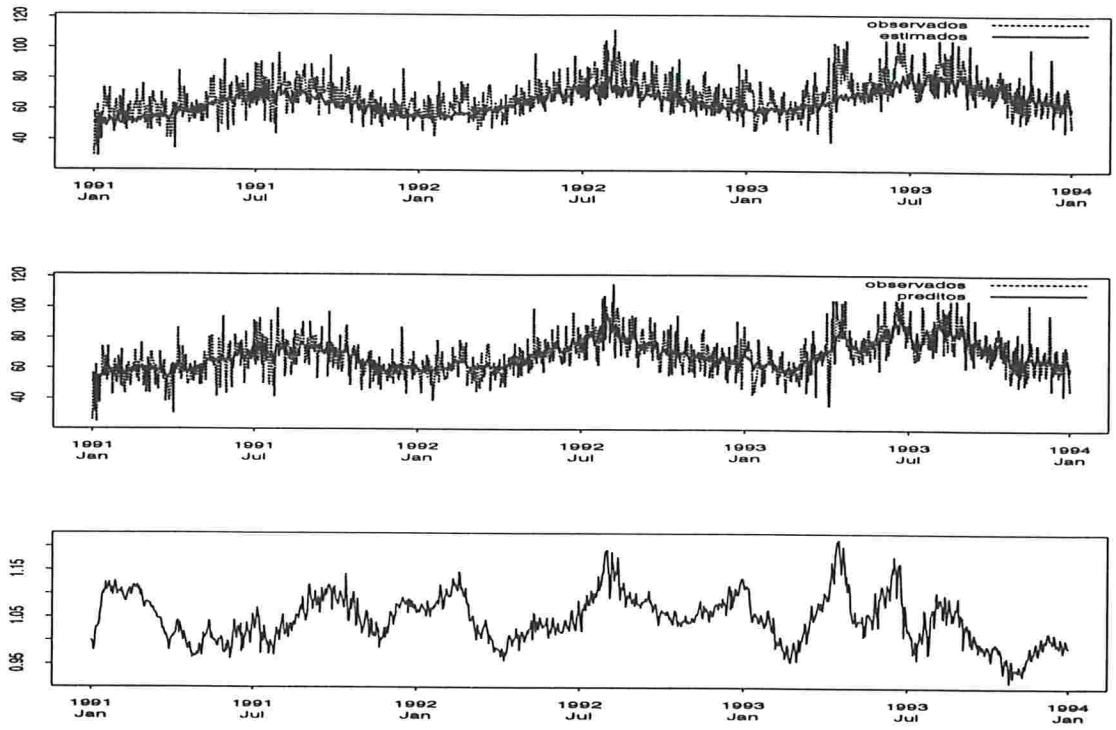

Figura 3.9: Valores observados e estimados, valores observados e preditos e processo latente estimado para o modelo final.

SO_2 permanecer constante, mas aumenta se o nível de SO_2 crescer.

Tabela 3.2: Coeficientes estimados e erros padrões para os modelos 4 ao modelo final.

Efeito	Modelo 4		Modelo 5		Modelo 6	
curto-prazo	**estimativa**	**(e.p.)**	**estimativa**	**(e.p.)**	**estimativa**	**(e.p.)**
constante	4.139	(0.079)	4.135	(0.078)	4.141	(0.078)
$\cos(2\pi t/365)$	− 0.120	(0.040)	− 0.125	(0.040)	− 0.125	(0.040)
$\text{sen}(2\pi t/365)$	−0.041	(0.040)	−0.044	(0.040)	−0.045	(0.041)
temp	0.0061	(0.0025)	0.0069	(0.0025)	0.0073	(0.0024)
def_1-temp	−0.0011	(0.0029)	−0.0004	(0.0028)	−0.0006	(0.0028)
def_2-temp	0.0072	(0.0028)	0.0069	(0.0028)	0.0069	(0.0028)
def_3-temp	− 0.0118	(0.0024)	− 0.0121	(0.0024)	− 0.0120	(0.0024)
umid	− 0.00105	(0.00033)	− 0.00117	(0.00032)	−0.00124	(0.00031)
tendência	0.00020	(0.00048)	0.00018	(0.00048)	0.00018	(0.00048)
longo-prazo						
partículas	0.00027	(0.00028)	-	-	-	-
CO	0.0032	(0.0028)	-	-	-	-
NO_2	7.1×10^{-6}	(6.5×10^{-5})	-	-	-	-
O_3	0.00009	(0.00012)	-	-	-	-
SO_2	0.00017	(0.00079)	0.00107	(0.00056)	-	-
def_1-SO_2	0.00020	(0.00061)	0.00045	(0.00059)	-	-
def_2-SO_2	0.00032	(0.00060)	0.00032	(0.00059)	-	-
def_3-SO_2	− 0.00127	(0.00056)	− 0.00128	(0.00055)	-	-
F_1	-	-	0.00057	(0.00084)	-	-
F_2	-	-	0.00031	(0.00048)	-	-
F_3	-	-	0.00029	(0.00050)	-	-
F_4	-	-	0.00142	(0.00048)	0.00156	(0.00044)
$\hat{\sigma}^2$	0.00024		0.00024		0.00025	

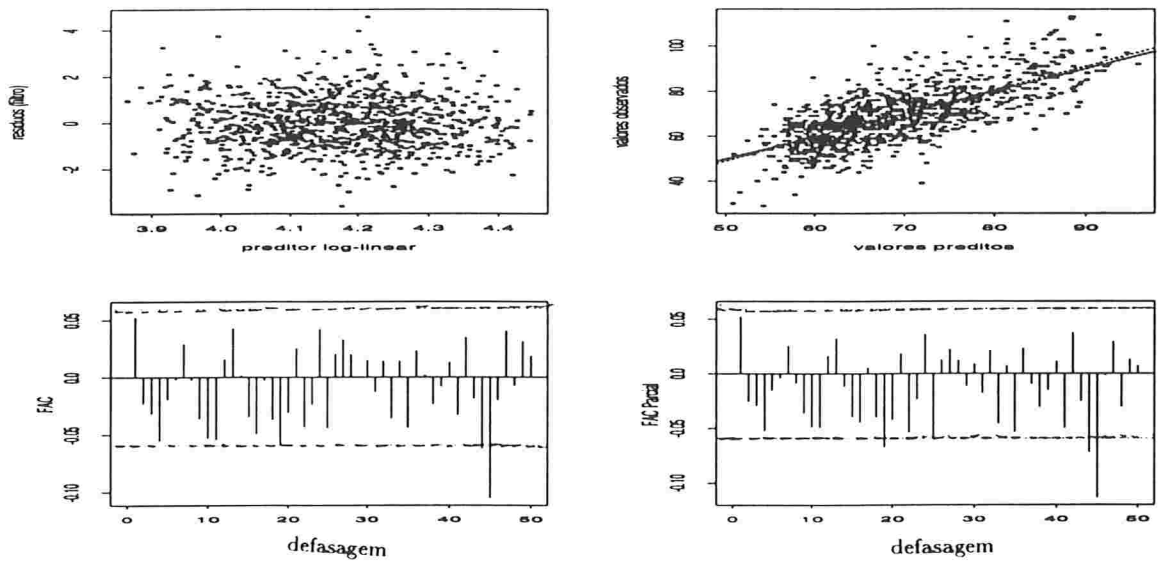

Figura 3.10: Gráficos de resíduos para o modelo final.

Capítulo 4

Modelo retrospectivo

Neste capítulo, apresentamos o modelo retrospectivo e suas propriedades. Na Seção 4.2, descrevemos sua estrutura, na Seção 4.3, obtemos seus momentos e na Seção 4.4 obtemos o filtro e o suavizador de Kalman. Na Seção 4.5, propomos estimadores para parâmetros de regressão e de dispersão por meio de uma técnica de estimação que é uma generalização daquela utilizada no modelo Tweedie-Tweedie. Para a estimação dos parâmetros retrospectivos consideramos na Seção 4.5.3 uma técnica baseada na função de verossimilhança perfilada.

4.1 Introdução

O modelo retrospectivo é um modelo de espaço de estados que generaliza a estrutura de correlação do modelo Tweedie-Tweedie apresentado no Capítulo 3, segundo o qual, dado um processo latente com estrutura de Markov de primeira ordem, as observações são independentes. Esta estrutura de correlação pode ser representada pelo esquema na Figura 4.1. Para o modelo retrospectivo, vamos admitir que o processo latente seja um processo de Markov de ordem $L \geq 1$, por exemplo ou que, dado o processo latente, a distribuição condicional do processo observado dependa de valor presente e de K valores passados do processo latente. O esquema na Figura 4.2 denota a estrutura de correlação de um modelo retrospectivo com

$$
\begin{array}{ccccccccccc}
 & & & \mathbf{Y}_{t-1} & & \mathbf{Y}_{t} & & & & \mathbf{Y}_{t+s} & \\
 & & & \uparrow & & \uparrow & & & & \uparrow & \\
\cdots \rightarrow & \theta_{t-2} & \rightarrow & \theta_{t-1} & \rightarrow & \theta_{t} & \rightarrow & \cdots & \rightarrow & \theta_{t+s} & \cdots
\end{array}
$$

Figura 4.1: Estrutura de correlação do modelo retrospectivo com $K = 0$ e $L = 1$ (modelo Tweedie–Tweedie).

$$
\begin{array}{ccccccccccc}
 & & & \mathbf{Y}_{t-1} & & \mathbf{Y}_{t} & & & & \mathbf{Y}_{t+s} & \\
 & & \nearrow & \uparrow & \nearrow & \uparrow & & & \nearrow & \uparrow & \\
\cdots \rightarrow & \theta_{t-2} & \rightarrow & \theta_{t-1} & \rightarrow & \theta_{t} & \rightarrow & \cdots & \rightarrow & \theta_{t+s} & \cdots
\end{array}
$$

Figura 4.2: Estrutura de correlação do modelo retrospectivo com $K = 1$ e $L = 1$ e um processo latente univariado.

$L = 1$ e $K = 1$. No esquema representado na Figura 4.1, dado θ_t *ou* θ_{t-1}, $\mathbf{Y}_t$ e $\mathbf{Y}_{t-1}$ são condicionalmente independentes; no esquema da Figura 4.2, $\mathbf{Y}_t$ e $\mathbf{Y}_{t-1}$ são condicionalmente independentes, dados θ_t *e* θ_{t-1}.

Para modelar a estrutura de correlação do modelo retrospectivo, vamos considerar, além dos parâmetros de regressão e de dispersão nos moldes daqueles do modelo Tweedie-Tweedie, uma terceira classe de parâmetros que vamos chamar de *parâmetros retrospectivos*. Esses parâmetros são dispostos em vetores de dois tipos: o primeiro, de dimensão $K + 1$, define a forma como o processo observado está associado ao processo latente, e o segundo, de dimensão L, define a ordem do processo de Markov que vamos assumir para o processo latente. Diremos que a *ordem* do modelo é $M = \max\{K + 1, L\}$.

4.2 Modelo

Podemos definir o modelo retrospectivo para h unidades amostrais independentes, como foi feito no Capítulo 3, mas para não carregar a notação, apresentamos nesta seção os detalhes para uma única unidade amostral e obtemos suas propriedades. A extensão para duas ou

mais unidades amostrais será considerada na Seção 4.5.

4.2.1 Definição

Consideremos um vetor de observações $\mathbf{Y}_t$ de dimensão d, avaliado em tempos $t = 1, \ldots, n$, igualmente espaçados, e seja $\mathbf{Y} = (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_n^\top)^\top$ o vetor completo de observações. Como fizemos no modelo Tweedie-Tweedie, chamamos as d componentes de $\mathbf{Y}_t$ de categorias. Consideremos também um processo latente univariado θ_t. Sejam K e L inteiros não-negativos e seja $M = \max\{K+1, L\}$. Definimos o vetor $\boldsymbol{\theta}$ como

$$\boldsymbol{\theta} = (\theta_{-M+1}, \ldots, \theta_0, \theta_1, \ldots, \theta_n)^\top,$$

em que $\theta_{-M+1}, \ldots, \theta_0$ são constantes iniciais que vamos assumir iguais a 1 no caso de uma única unidade amostral. Vamos assumir que as nd componentes de $\mathbf{Y}$ são condicionalmente independentes dado $\boldsymbol{\theta}$, tanto entre categorias quanto no tempo. Sejam $\boldsymbol{\theta}^t$ e $\boldsymbol{\Theta}_t$ sub-vetores de $\boldsymbol{\theta}$ dados por

$$\boldsymbol{\theta}^t = (\theta_{-M+1}, \ldots, \theta_t)^\top \quad \text{e} \quad \boldsymbol{\Theta}_t = (\theta_{t-M+1}, \ldots, \theta_t)^\top,$$

respectivamente. Em particular, no caso de uma única unidade amostral, $\boldsymbol{\theta}^0 = \boldsymbol{\Theta}_0 = (1, \ldots, 1)^\top$.

Sejam $\mathbf{x}_t$ e $\mathbf{z}_t$ vetores de covariáveis que variam no tempo, de dimensões l_1 e l_2 respectivamente, tais que $\mathbf{x}_t$ contém uma covariável constante. Sejam $\boldsymbol{\alpha}_i, i = 1, \ldots, d$, parâmetros de regressão associados a $\mathbf{x}_t$ podendo variar entre as d categorias de $\mathbf{Y}_t$ e $\boldsymbol{\beta}$ um parâmetro de regressão associado a $\mathbf{z}_t$.

Sejam $\boldsymbol{\varphi}_i, i = 1, \ldots, d$, e $\boldsymbol{\psi}$ vetores com $K+1$ e L componentes positivos de soma igual a 1, respectivamente e tais que $\boldsymbol{\varphi}_i, i = 1, \ldots, d$, possam variar entre as d categorias de $\mathbf{Y}_t$. Por conveniência, escrevemos os parâmetros $\boldsymbol{\varphi}_i$ e $\boldsymbol{\psi}$ como vetores M-dimensionais, fazendo seus primeiros $M-K-1$ e $M-L$ componentes, respectivamente, iguais a zero.

O modelo é definido pelo seguinte conjunto de distribuições condicionais para $t = 1, \ldots, n$,

$$Y_{it}|\boldsymbol{\theta}^t \sim \mathrm{Tw}_{p_i}\left(\mu_{it}\frac{\boldsymbol{\varphi}_i^\top \boldsymbol{\Theta}_t}{\boldsymbol{\varphi}_i^\top \boldsymbol{\tau}_t}, \frac{\nu_i^2}{(\boldsymbol{\varphi}_i^\top \boldsymbol{\Theta}_t)^{p_i-1}}\right) \tag{4.1}$$

e

$$\theta_t|\boldsymbol{\theta}^{t-1} \sim \mathrm{Tw}_q\left(\tau_t\frac{\boldsymbol{\psi}^\top\boldsymbol{\Theta}_{t-1}}{\boldsymbol{\psi}^\top\boldsymbol{\tau}_{t-1}}, \frac{\sigma^2}{(\boldsymbol{\psi}^\top\boldsymbol{\Theta}_{t-1})^{q-1}}\right), \tag{4.2}$$

em que $\tau_t = \exp(\mathbf{z}_t^\top\boldsymbol{\beta})$, $\boldsymbol{\tau}_t = (\tau_{t-M+1}, \ldots, \tau_t)^\top$ e $\mu_{it} = \tau_t \exp(\mathbf{x}_t^\top\boldsymbol{\alpha}_i)$. Por conveniência, vamos supor ainda que $\mathbf{z}_0 = \mathbf{0}$. Com isso, obtemos $\tau_0 = 1$ e convencionamos $\boldsymbol{\tau}_0 = (1, \ldots, 1)^\top$. Os parâmetros $\nu_i^2, i = 1, \ldots, d$ e σ^2 são parâmetros de dispersão. Como no Capítulo 3, $p_i, i = 1, \ldots, d$, podem variar entre categorias.

A ordem do modelo retrospectivo para o caso $K = 1$ e $L = 1$ (Figura 4.2), é $M = \max\{1+1, 1\} = 2$ e os parâmetros retrospectivos são representados por $\boldsymbol{\varphi} = (\varphi_1, \varphi_2)^\top$, $\boldsymbol{\psi} = (0, 1)^\top$. Ao considerarmos o sub-vetor $\boldsymbol{\Theta}_t = (\theta_{t-1}, \theta_t)^\top$, a estrutura de correlação acima pode ser representada pelo esquema da Figura 4.3 (ou Figura 2.1).

Os esquemas nas Figuras 4.2 e 4.3 são equivalentes e o segundo representa qualquer modelo retrospectivo. Desta forma, ao *empilhar* o processo latente univariado, conseguimos sempre representar o processo latente por um processo de Markov M-variado, de ordem 1. Este resultado será muito útil na obtenção do filtro e suavizador de Kalman na Seção 4.4.

4.2.2 Propriedades

Da definição da família Tweedie (ver Apêndice A), temos que as médias e variâncias condicionais referentes a (4.1) e (4.2) são,

$$\mathrm{E}(Y_{it}|\boldsymbol{\theta}^t) = \mu_{it}\frac{\boldsymbol{\varphi}_i^\top\boldsymbol{\Theta}_t}{\boldsymbol{\varphi}_i^\top\boldsymbol{\tau}_t}, \quad \mathrm{Var}\left(Y_{it}|\boldsymbol{\theta}^t\right) = \nu_i^2\mu_{it}^{p_i}\frac{\boldsymbol{\varphi}_i^\top\boldsymbol{\Theta}_t}{(\boldsymbol{\varphi}_i^\top\boldsymbol{\tau}_t)^{p_i}}, \tag{4.3}$$

$$\mathrm{E}(\theta_t|\boldsymbol{\theta}^t) = \tau_t\frac{\boldsymbol{\psi}^\top\boldsymbol{\Theta}_{t-1}}{\boldsymbol{\psi}^\top\boldsymbol{\tau}_{t-1}}, \quad \mathrm{Var}\left(\theta_t|\boldsymbol{\theta}^{t-1}\right) = \sigma^2\tau_t^q\frac{\boldsymbol{\psi}^\top\boldsymbol{\Theta}_{t-1}}{(\boldsymbol{\psi}^\top\boldsymbol{\tau}_{t-1})^q}, \tag{4.4}$$

respectivamente.

Tanto as variâncias quanto as médias condicionais são lineares em $\boldsymbol{\theta}$ e esta é a principal razão para a escolha da forma dos parâmetros de dispersão em (4.1) e (4.2) pois vai permitir a utilização dos resultados da Seção 2.2 para obtenção do filtro e suavizador de Kalman para o modelo retrospectivo.

Como no modelo Tweedie-Tweedie, podemos obter as esperanças marginais iterativamente. Temos que $\mathrm{E}(\theta_0) = 1$ e $\mathrm{E}(\theta_t) = \tau_t$ o que nos dá $\mathrm{E}(\Theta_t) = \boldsymbol{\tau}_t$. O modelo retrospectivo é um modelo log-linear, uma vez que o valor esperado do processo observado é

$$\mathrm{E}(Y_{it}) = \mu_{it} = \tau_t \exp\left(\mathbf{x}_t^\top \boldsymbol{\alpha}_i\right) = \exp\left(\mathbf{x}_t^\top \boldsymbol{\alpha}_i + \mathbf{z}_t^\top \boldsymbol{\beta}\right), \tag{4.5}$$

para $t = 1, \ldots, n$. Estes resultados são detalhados nas Proposições 4.1 e 4.2 da Seção 4.3.

O processo latente é um processo de Markov de ordem L. Sua dinâmica pode ser melhor entendida se considerarmos a esperança condicional em (4.4), que consiste de uma média τ_t multiplicada por um fator modulador $(\boldsymbol{\psi}^\top \Theta_{t-1})/(\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1})$ de valor esperado igual a 1. O termo $\boldsymbol{\psi}^\top \Theta_{t-1}$ é uma combinação convexa dos L valores passados de θ_t, de forma que o fator modulador tenderá a manter o processo em seu valor recente médio, relativamente ao valor esperado τ_t. Analogamente, a esperança condicional em (4.3) é dada por um valor esperado μ_{it} multiplicado por um fator modulador de média 1, representando uma média ponderada do valor atual e dos K valores mais recentes do processo latente. Vetores $\boldsymbol{\varphi}_i$ na mesma direção mas com comprimentos distintos, definem o mesmo fator modulador. O mesmo acontece com diferentes vetores $\boldsymbol{\psi}$ e é devido a isso que exigimos que a soma dos componentes destes vetores seja igual a 1. Quando $K > 0$, a distribuição do processo observado depende do valor atual e de K valores passados do processo latente. Dizemos então que essa é a parte *retrospectiva* do modelo, no sentido que o passado recente do processo θ influencia a observação atual. Vamos nos referir ao modelo definido por (4.1)–(4.2) como um *modelo retrospectivo de ordem* M. O modelo retrospectivo de ordem 1 corresponde a $K = 0$ e $L = 1$, com $\boldsymbol{\varphi}_i = \boldsymbol{\psi} = 1$ e se reduz ao modelo Tweedie-Tweedie.

Podemos representar tanto $\boldsymbol{\psi}^\top \Theta_{t-1}$ quanto $\boldsymbol{\varphi}_i^\top \Theta_t$ em termos de polinômios formais no operador *atraso* B, como segue,

$$\begin{aligned}
\boldsymbol{\varphi}_i^\top \Theta_t &= \varphi_i(B)(\theta_t) = \varphi_{i,M} B^0(\theta_t) + \varphi_{i,M-1} B(\theta_t) + \cdots + \varphi_{i,1} B^M(\theta_t), \\
\boldsymbol{\psi}^\top \Theta_{t-1} &= \psi(B)(\theta_t) = \psi_M B(\theta_t) + \psi_{M-1} B^2(\theta_t) + \cdots + \psi_1 B^M(\theta_t),
\end{aligned}$$

em que $B^j \theta_t = \theta_{t-j}$ (ver, por exemplo, Box e Jenkins, 1970). Com o objetivo de simplificar a notação em algumas fórmulas, vamos eventualmente utilizar esta representação polinomial.

$$
\begin{array}{ccccccccc}
 & & \mathbf{Y}_{t-1} & & \mathbf{Y}_t & & & & \mathbf{Y}_{t+s} \\
 & & \uparrow & & \uparrow & & & & \uparrow \\
\cdots \rightarrow & \Theta_{t-2} & \rightarrow & \Theta_{t-1} & \rightarrow & \Theta_t & \rightarrow \cdots \rightarrow & & \Theta_{t+s} \quad \cdots
\end{array}
$$

Figura 4.3: Estrutura de correlação do modelo retrospectivo com $K = 1$, $L = 1$ e um processo latente bivariado.

4.3 Momentos

Nesta seção, vamos investigar a estrutura dos momentos do modelo retrospectivo que é necessária para a obtenção do filtro e do suavizador de Kalman (Seção 4.4) e para a estimação dos parâmetros (Seção 4.5).

Os momentos do processo latente univariado são facilmente obtidos a partir de (4.1)–(4.2) e podem ser expressos em termos do processo latente M-variado Θ_t. Vamos apresentar os resultados sobre estes momentos na forma de proposições.

Proposição 4.1 $\mathrm{E}(\theta_t) = \tau_t$ *e* $\mathrm{E}(\Theta_t) = \boldsymbol{\tau}_t$.

Demonstração. Como θ_t é constante para $t = -M+1, \ldots, 0$, podemos denotar $\mathrm{E}(\theta_t) = \tau_t$ para t neste intervalo. Logo, $\mathrm{E}(\Theta_0) = \boldsymbol{\tau}_0$ e o resultado é obtido por recursão, já que $\mathrm{E}(\theta_t|\boldsymbol{\theta}^{t-1}) = \tau_t(\boldsymbol{\psi}^\top \Theta_{t-1})/(\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1})$. □

Proposição 4.2 $\mathrm{E}(\Theta_t|\boldsymbol{\theta}^{t-1}) = \mathbf{G}_t\Theta_{t-1}$, *em que* $\mathbf{G}_t$ *é uma matriz* $M \times M$, *cujos coeficientes dependem de* $\boldsymbol{\psi}$ *e de* $\boldsymbol{\tau}_t$.

Demonstração. De (4.4), obtemos

$$
\mathrm{E}(\Theta_t|\boldsymbol{\theta}^{t-1}) = \mathrm{E}\left\{ \begin{pmatrix} \theta_{t-M+1} \\ \vdots \\ \theta_t \end{pmatrix} \middle| \boldsymbol{\theta}^{t-1} \right\} = \begin{pmatrix} \theta_{t-M+1} \\ \vdots \\ \theta_{t-1} \\ \tau_t \dfrac{\boldsymbol{\psi}^\top \Theta_{t-1}}{\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1}} \end{pmatrix}.
$$

Seja $c_t = \tau_t/(\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1})$. Denotando por $\mathbf{G}_t$ a matriz

$$\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & \ddots & & \\ 0 & 0 & 0 & \dots & 1 \\ c_t\psi_1 & c_t\psi_2 & & \dots & c_t\psi_M \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{(M-1)\times 1} & \mathbf{I}_{M-1} \\ c_t\boldsymbol{\psi}^\top & \end{pmatrix},$$

obtemos o resultado. □

Proposição 4.3 *A variância do processo latente é dada por*

$$\mathrm{Var}(\theta_t) = \lambda_t + c_t^2 \boldsymbol{\psi}^\top \mathrm{Var}(\boldsymbol{\Theta}_{t-1})\boldsymbol{\psi} \quad \text{e} \quad \mathrm{Var}(\boldsymbol{\Theta}_t) = \boldsymbol{\lambda}_t + \mathbf{G}_t \mathrm{Var}(\boldsymbol{\Theta}_{t-1})\mathbf{G}_t^\top,$$

em que

$$\lambda_t = \sigma^2 \frac{\tau_t^q}{(\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1})^{q-1}} \quad \text{e} \quad \boldsymbol{\lambda}_t = \begin{pmatrix} \mathbf{0}_{M\times M} & \mathbf{0}_{M\times 1} \\ \mathbf{0}_{1\times M} & \lambda_t \end{pmatrix}.$$

Demonstração. De (4.4) e da Proposição 4.1, temos que $\mathrm{E}\{\mathrm{Var}(\theta_t|\boldsymbol{\theta}^{t-1})\} = (\sigma^2\tau_t^q)/(\boldsymbol{\psi}^\top\boldsymbol{\tau}_{t-1})^q$. Também de (4.4), obtemos que $\mathrm{Var}\{\mathrm{E}(\theta_t|\boldsymbol{\theta}^{t-1})\} = \tau_t^2 \left\{\boldsymbol{\psi}^\top \mathrm{Var}(\boldsymbol{\Theta}_{t-1})\boldsymbol{\psi}\right\}/(\boldsymbol{\psi}^\top\boldsymbol{\tau}_{t-1})^2$. Então, pela Propriedade B.1, temos que

$$\mathrm{Var}(\theta_t) = \mathrm{E}\{\mathrm{Var}(\theta_t|\boldsymbol{\theta}^{t-1})\} + \mathrm{Var}\{\mathrm{E}(\theta_t|\boldsymbol{\theta}^{t-1})\} = \lambda_t + c_t^2\boldsymbol{\psi}^\top \mathrm{Var}(\boldsymbol{\Theta}_{t-1})\boldsymbol{\psi}.$$

Para a segunda parte, temos

$$\begin{aligned} \mathrm{Var}(\boldsymbol{\Theta}_t) &= \mathrm{E}\{\mathrm{Var}(\boldsymbol{\Theta}_t|\boldsymbol{\theta}^{t-1})\} + \mathrm{Var}\{\mathrm{E}(\boldsymbol{\Theta}_t|\boldsymbol{\theta}^{t-1})\} \\ &= \mathrm{E}[\mathrm{Var}\{(\theta_{t-M+1},\dots,\theta_t)^\top|\boldsymbol{\theta}^{t-1}\}] + \mathrm{Var}(\mathbf{G}_t\boldsymbol{\Theta}_{t-1}). \end{aligned}$$

Como $(\theta_{t-M+1},\dots,\theta_{t-1})^\top$ é um subvetor de $\boldsymbol{\theta}^{t-1}$, temos que $\mathrm{Var}\{(\theta_{t-M+1},\dots,\theta_{t-1})^\top|\boldsymbol{\theta}^{t-1}\} = \mathbf{0}$ e portanto, $\mathrm{E}\{\mathrm{Var}(\boldsymbol{\Theta}_t|\boldsymbol{\theta}^{t-1})\} = \boldsymbol{\lambda}_t$. □

Proposição 4.4 $\mathrm{Cov}\{\theta_t, \boldsymbol{\Theta}_{t-1}\} = c_t\boldsymbol{\psi}_t^\top \mathrm{Var}(\boldsymbol{\Theta}_{t-1})$.

Demonstração. Da Propriedade B.1, temos que $\mathrm{Cov}\{\theta_t, \Theta_{t-1}\} = \mathrm{Cov}\{\mathrm{E}(\theta_t, \Theta_{t-1}|\boldsymbol{\theta}^{t-1})\}$, pois Θ_{t-1} é um subvetor de $\boldsymbol{\theta}^{t-1}$. Temos então que

$$\mathrm{Cov}\{\mathrm{E}(\theta_t, \Theta_{t-1}|\boldsymbol{\theta}^{t-1})\} = \mathrm{Cov}\{\mathrm{E}(\theta_t|\boldsymbol{\theta}^{t-1}), \Theta_{t-1}\} = \mathrm{Cov}\left\{\tau_t(\psi^\top\Theta_{t-1})/(\psi^\top\boldsymbol{\tau}_{t-1}), \Theta_{t-1}\right\}.$$

□

Proposição 4.5 *Para* $s > 0$, $\mathrm{Cov}(\Theta_t, \Theta_{t+s}) = \mathrm{Cov}(\Theta_t, \Theta_{t+s-1})\mathbf{G}_{t+s}^\top$.

Demonstração. O argumento é semelhante ao utilizado na demonstração da Proposição 4.4. Condicionando em $\boldsymbol{\theta}^{t+s-1}$, obtemos $\mathrm{E}\{\mathrm{Cov}(\Theta_t, \Theta_{t+s}|\boldsymbol{\theta}^{t+s-1}\} = \mathbf{0}$, pois Θ_t é um subvetor de $\boldsymbol{\theta}^{t+s-1}$. Então, $\mathrm{Cov}(\Theta_t, \Theta_s) = \mathrm{Cov}\left\{\mathrm{E}(\Theta_t, \Theta_{t+s}|\boldsymbol{\theta}^{t+s-1})\right\} = \mathrm{Cov}\{\Theta_t, \mathrm{E}(\Theta_{t+s}|\boldsymbol{\theta}^{t+s-1})\} = \mathrm{Cov}(\Theta_t, \mathbf{G}_{t+s}\Theta_{t+s-1})$. □

Proposição 4.6 $\mathrm{E}\left[\mathbf{Y}_t|\boldsymbol{\theta}^t\right] = \mathbf{F}_t\Theta_t$, *em que* $\mathbf{F}_t$ *é uma matriz* $d \times M$, *cujos coeficientes dependem de* φ_i.

Demonstração. De (4.3), temos que $\mathrm{E}(Y_{it}|\boldsymbol{\theta}^t) = \mu_{it}(\varphi_i^\top\Theta_t)/(\varphi_i^\top\boldsymbol{\tau}_t)$. Escrevendo então

$$\mathbf{F}_t = \mathrm{diag}\left\{\frac{\mu_{it}}{\varphi_i^\top\boldsymbol{\tau}_t}\right\}_{d\times d} \times \begin{pmatrix} \varphi_{i,1}^\top \\ \vdots \\ \varphi_{i,d}^\top \end{pmatrix}_{d\times M},$$

o resultado segue. □

A variância marginal do processo observado é composta por dois termos, como mostra a Proposição 4.7.

Proposição 4.7 $\mathrm{Var}[\mathbf{Y}_t] = \boldsymbol{\Lambda}_t + \mathbf{F}_t\mathrm{Var}(\Theta_t)\mathbf{F}_t^\top$, *em que* $\boldsymbol{\Lambda}_t$ *é uma matriz diagonal.*

Demonstração. De (4.3), obtemos

$$\mathrm{Var}(\mathbf{Y}_t|\boldsymbol{\theta}^t) = \mathrm{diag}\left\{\nu_i^2\mu_{it}^{p_i}\frac{\varphi_i^\top\Theta_t}{(\varphi_i^\top\boldsymbol{\tau}_t)^{p_i}}\right\}_{d\times d},$$

pois as observações são condicionalmente independentes dado o processo latente. Da Proposição 4.6, obtemos $\mathrm{Var}\{\mathrm{E}(\mathbf{Y}_t|\boldsymbol{\theta}^t)\} = \mathbf{F}_t\mathrm{Var}(\boldsymbol{\Theta}_t)\mathbf{F}_t^\top$ e temos então que

$$\mathrm{Var}[\mathbf{Y}_t] = \mathrm{E}\{\mathrm{Var}(\mathbf{Y}_t|\boldsymbol{\theta}^t)\} + \mathbf{F}_t\mathrm{Var}(\boldsymbol{\Theta}_t)\mathbf{F}_t^\top = \mathrm{diag}\left\{\frac{\nu_i^2\mu_{it}^{p_i}}{(\boldsymbol{\varphi}_i^\top\boldsymbol{\tau}_t)^{p_i-1}}\right\}_{d\times d} + \mathbf{F}_t\mathrm{Var}(\boldsymbol{\Theta}_t)\mathbf{F}_t^\top.$$

□

O segundo termo na expressão da variância mostra superdispersão relativa ao modelo (4.1) e também que a correlação entre as categorias é positiva. Embora esta seja uma limitação do modelo, correlação positiva entre as categorias ocorre em várias situações reais. Observações separadas por um espaço de tempo $s > 0$ são correlacionadas e sua covariância é dada na Proposição 4.8.

Proposição 4.8 $\mathrm{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t+s}) = \mathbf{F}_t\mathrm{Cov}\left(\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_{t+s}\right)\mathbf{F}_{t+s}^\top$, *para* $s > 0$.

Demonstração. Condicionando em $\boldsymbol{\theta}^n$, temos $\mathrm{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t+s}) = \mathrm{Cov}\{\mathrm{E}(\mathbf{Y}_t, \mathbf{Y}_{t+s}|\boldsymbol{\theta}^n)\} + \mathrm{E}\{\mathrm{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t+s}|\boldsymbol{\theta}^n)\}$. Mas $\mathrm{E}\{\mathrm{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t+s}|\boldsymbol{\theta}^n)\} = \mathbf{0}$, pois $\mathbf{Y}_t$ e $\mathbf{Y}_{t+s}$ são condicionalmente independentes dado $\boldsymbol{\theta}^n$. Usando então a Propriedade B.2, obtemos

$$\mathrm{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t+s}) = \mathrm{Cov}\{\mathrm{E}(\mathbf{Y}_t|\boldsymbol{\theta}^n), \mathrm{E}(\mathbf{Y}_{t+s}|\boldsymbol{\theta}^n)\} = \mathrm{Cov}(\mathbf{F}_t\boldsymbol{\Theta}_t, \mathbf{F}_{t+s}\boldsymbol{\Theta}_{t+s}).$$

□

A demonstração da Proposição 4.9 é análoga à da Proposição 4.8 e será omitida.

Proposição 4.9 $\mathrm{Cov}(\mathbf{Y}_t, \boldsymbol{\Theta}_s) = \mathbf{F}_t\mathrm{Cov}(\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_s)$, *para todo* t, s.

As seguintes proposições serão úteis na obtenção do filtro e suavizador de Kalman.

Proposição 4.10 $\mathrm{E}(\mathbf{Y}_{t+1}, \ldots, \mathbf{Y}_n|\boldsymbol{\Theta}_{t+1})$ *é linear em* $\boldsymbol{\Theta}_{t+1}$.

Demonstração. Como $\boldsymbol{\sigma}(\boldsymbol{\Theta}_{t+1}) \subset \boldsymbol{\sigma}(\boldsymbol{\theta}^n)$, obtemos da Propriedade B.2 que

$$\mathrm{E}(\mathbf{Y}_{t+1}, \ldots, \mathbf{Y}_n|\boldsymbol{\Theta}_{t+1}) = \mathrm{E}\{\mathrm{E}(\mathbf{Y}_{t+1}, \ldots, \mathbf{Y}_n|\boldsymbol{\theta}^n)|\boldsymbol{\Theta}_{t+1}\}.$$

Como as observações são condicionalmente independentes dado o processo latente, obtemos da Proposição 4.6 que

$$\mathrm{E}(\mathbf{Y}_{t+1}, \ldots, \mathbf{Y}_n|\boldsymbol{\theta}^n) = \{\mathrm{E}(\mathbf{Y}_{t+1}|\boldsymbol{\theta}^n), \ldots, \mathrm{E}(\mathbf{Y}_n|\boldsymbol{\theta}^n)\} = (\mathbf{F}_{t+1}\boldsymbol{\Theta}_{t+1}, \ldots, \mathbf{F}_n\boldsymbol{\Theta}_n).$$

Aplicando a Propriedade B.3, obtemos

$$\mathrm{E}(\mathbf{Y}_{t+1},\ldots,\mathbf{Y}_n|\Theta_{t+1}) = \{\mathbf{F}_{t+1}\Theta_{t+1},(\mathbf{F}_{t+2}\Theta_{t+2},\ldots,\mathbf{F}_n\Theta_n|\Theta_{t+1})\}.$$

Agora, basta notar que $\sigma(\Theta_{t+1}) \subset \sigma(\theta^{t+1}) \subset \ldots \subset \sigma(\theta^{n-1})$ e aplicando as Propriedades B.2 e B.3 recursivamente, obtemos,

$$\mathrm{E}(\mathbf{Y}_{t+1},\ldots,\mathbf{Y}_n|\Theta_{t+1}) = (\mathbf{F}_{t+1}\Theta_{t+1},\mathbf{F}_{t+2}\mathbf{G}_{t+2}\Theta_{t+1},\ldots,\mathbf{F}_n\mathbf{G}_n\cdots\mathbf{G}_{t+2}\Theta_{t+1}),$$

e o resultado segue. □

Proposição 4.11 $\mathrm{Cov}(\mathbf{Y}_{t+1},\ldots,\mathbf{Y}_n,\Theta_t|\Theta_{t+1}) = 0.$

Demonstração. Como Θ_{t+1} é um subvetor de θ^n, temos que $\sigma(\Theta_{t+1}) \subset \sigma(\theta^n)$. Como $\mathbf{Y}_{t+1},\ldots,\mathbf{Y}_n$ são condicionalmente independentes dado θ^n, temos que

$$\mathrm{Cov}(\mathbf{Y}_{t+1},\ldots,\mathbf{Y}_n,\Theta_t|\theta^n) = 0.$$

Pela Proposição 4.10, temos que

$$\begin{aligned}\mathrm{E}(\mathbf{Y}_{t+1},\ldots,\mathbf{Y}_n,\Theta_t|\Theta_{t+1}) &= \mathrm{E}\{\mathrm{E}(\mathbf{Y}_{t+1},\ldots,\mathbf{Y}_n,\Theta_t|\theta^n)|\Theta_{t+1}\}\\ &= \mathrm{E}(\mathbf{F}_{t+1}\Theta_{t+1},\mathbf{F}_{t+2}\mathbf{G}_{t+2}\Theta_{t+1},\ldots,\mathbf{F}_n\mathbf{G}_n\cdots\mathbf{G}_{t+2}\Theta_{t+1},\Theta_t|\Theta_{t+1}).\end{aligned}$$

O resultado decorre então das Propriedades B.1 e B.2. □

4.4 O filtro e o suavizador de Kalman

Nesta seção, apresentamos o modelo retrospectivo definido por (4.1)–(4.2) na forma de um modelo de espaço de estados e obtemos o filtro e o suavizador de Kalman.

O filtro de Kalman prediz uma observação futura $\mathbf{Y}_{t+1}$ e o valor atual do processo latente θ_t, baseado nas observações atual $\mathbf{Y}_t$ e passadas, $\mathbf{Y}_1,\ldots,\mathbf{Y}_{t-1}$. O suavizador prediz o processo latente a partir do vetor total de observações, $\mathbf{Y}_1,\ldots,\mathbf{Y}_n$.

4.4.1 Modelo de espaço de estados

O modelo retrospectivo pode ser representado na forma de espaço de estados, como segue. Sejam

$$\varepsilon_{it} = Y_{it} - \mathrm{E}\left(Y_{it}|\boldsymbol{\theta}^t\right) = Y_{it} - \mu_{it}\frac{\boldsymbol{\varphi}_i^\top \boldsymbol{\Theta}_t}{\boldsymbol{\varphi}_i^\top \boldsymbol{\tau}_t} \tag{4.6}$$

e

$$\xi_t = \theta_t - \mathrm{E}\left(\theta_t|\boldsymbol{\theta}^{t-1}\right) = \theta_t - \tau_t\frac{\boldsymbol{\psi}^\top \boldsymbol{\Theta}_{t-1}}{\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1}}, \tag{4.7}$$

para $i = 1, \ldots, d$ e $t = 1, \ldots, n$, as inovações para os processos observado e latente, respectivamente, e sejam $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{dt})^\top$ e $\boldsymbol{\xi}_t = (0, \ldots, 0, \xi_t)^\top$. Então, $\boldsymbol{\xi}_t = \boldsymbol{\Theta}_t - \mathrm{E}(\boldsymbol{\Theta}_t|\boldsymbol{\theta}^{t-1})$ e, das Proposições 4.2 e 4.6, obtemos a seguinte representação para o modelo retrospectivo,

$$\mathbf{Y}_t = \mathbf{F}_t\boldsymbol{\Theta}_t + \boldsymbol{\varepsilon}_t, \tag{4.8}$$

$$\boldsymbol{\Theta}_t = \mathbf{G}_t\boldsymbol{\Theta}_{t-1} + \boldsymbol{\xi}_t. \tag{4.9}$$

Desta forma, ao empilhar o processo latente, obtemos um processo de Markov multivariado de primeira ordem. As definições das inovações apresentadas acima garantem que elas têm esperança zero. Além disso, elas são mutuamente não correlacionadas, o que demonstramos, juntamente com outras propriedades, nas proposições que seguem. Estes resultados fazem desta representação do modelo retrospectivo uma extensão dos modelos de espaço de estados tradicionais.

Proposição 4.12 $\mathrm{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_s) = \mathbf{0}$, *para* $s \neq t$.

Demonstração. Sem perda de generalidade, podemos supor $s > t$. Então, $\mathrm{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_s) = \mathrm{Cov}(\mathbf{Y}_t - \mathbf{F}_t\boldsymbol{\Theta}_t, \mathbf{Y}_s - \mathbf{F}_s\boldsymbol{\Theta}_s) = \mathrm{Cov}(\mathbf{Y}_t, \mathbf{Y}_s) - \mathbf{F}_t\mathrm{Cov}(\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_s)\mathbf{F}_s^\top = \mathbf{0}$, pela Proposição 4.8. □

Proposição 4.13 $\mathrm{Cov}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_s) = \mathbf{0}$, *para* $s \neq t$.

Demonstração. Para $s > t$, temos $\mathrm{Cov}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_s) = \mathrm{Cov}(\boldsymbol{\Theta}_t - \mathbf{G}_t\boldsymbol{\Theta}_{t-1}, \boldsymbol{\Theta}_s - \mathbf{G}_s\boldsymbol{\Theta}_{s-1}) = \mathrm{Cov}(\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_s) - \mathbf{G}_t\mathrm{Cov}(\boldsymbol{\Theta}_{t-1}, \boldsymbol{\Theta}_s) - \mathrm{Cov}(\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_{s-1})\mathbf{G}_s^\top + \mathbf{G}_t\mathrm{Cov}(\boldsymbol{\Theta}_{t-1}, \boldsymbol{\Theta}_{s-1})\mathbf{G}_s^\top$. Pela Proposição 4.5, temos que o primeiro e o terceiro termos do lado direito da igualdade acima se anulam. Pela mesma proposição, temos também que o segundo e quarto termos se anulam e o resultado segue. □

Proposição 4.14 $\mathrm{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\xi}_s) = \mathbf{0}$, *para todo* s, t.

Demonstração.

Temos que $\mathrm{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\xi}_s) = \mathrm{Cov}(\mathbf{Y}_t - \mathbf{F}_t\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_s - \mathbf{G}_s\boldsymbol{\Theta}_{s-1}) = \mathrm{Cov}(\mathbf{Y}_t, \boldsymbol{\Theta}_s) - \mathbf{F}_t\mathrm{Cov}(\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_s) - \mathrm{Cov}(\mathbf{Y}_t, \boldsymbol{\Theta}_{s-1})\mathbf{G}_s^\top - \mathbf{F}_t\mathrm{Cov}(\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_{s-1})\mathbf{G}_s^\top$. O resultado segue então da Proposição 4.9, que nos dá $\mathrm{Cov}(\mathbf{Y}_t, \boldsymbol{\Theta}_s) = \mathbf{F}_t\mathrm{Cov}(\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_s)$ e $\mathrm{Cov}(\mathbf{Y}_t, \boldsymbol{\Theta}_{s-1}) = \mathbf{F}_t\mathrm{Cov}(\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_{s-1})$. □

Proposição 4.15 $\mathrm{Var}(\boldsymbol{\xi}_t) = \boldsymbol{\lambda}_t$, *em que* $\boldsymbol{\lambda}_t$ *é definida na Proposição 4.3.*

Demonstração. Temos $\mathrm{Var}(\boldsymbol{\xi}_t) = \mathrm{Var}(\boldsymbol{\Theta}_t - \mathbf{G}_t\boldsymbol{\Theta}_{t-1}) = \mathrm{Var}(\boldsymbol{\Theta}_t) + \mathbf{G}_t\mathrm{Var}(\boldsymbol{\Theta}_{t-1})\mathbf{G}_t^\top - \mathrm{Cov}(\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_{t-1})\mathbf{G}_t^\top - \mathbf{G}_t\mathrm{Cov}(\boldsymbol{\Theta}_{t-1}, \boldsymbol{\Theta}_t)$. Pela Proposição 4.5, temos que $\mathrm{Cov}(\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_{t-1}) = \mathbf{G}_t\mathrm{Var}(\boldsymbol{\Theta}_{t-1})$. O resultado segue então da Proposição 4.3. □

Proposição 4.16 $\mathrm{Var}(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Lambda}_t$, *em que* $\boldsymbol{\Lambda}_t$ *é a matriz diagonal definida na Proposição 4.7.*

Demonstração. $\mathrm{Var}(\boldsymbol{\varepsilon}_t) = \mathrm{Var}(\mathbf{Y}_t - \mathbf{F}_t\boldsymbol{\Theta}_t) = \mathrm{Var}(\mathbf{Y}_t) - \mathrm{Cov}(\mathbf{Y}_t, \boldsymbol{\Theta}_t)\mathbf{F}_t^\top - \mathbf{F}_t\mathrm{Cov}(\boldsymbol{\Theta}_t, \mathbf{Y}_t) + \mathbf{F}_t\mathrm{Var}(\boldsymbol{\Theta}_t)\mathbf{F}_t^\top$. Pela Proposição 4.9, temos que $\mathrm{Cov}(\mathbf{Y}_t, \boldsymbol{\Theta}_t) = \mathbf{F}_t\mathrm{Var}(\boldsymbol{\Theta}_t)$. O resultado segue então da Proposição 4.7. □

Proposição 4.17 $\mathrm{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\Theta}_t) = \mathbf{0}$.

Demonstração. $\mathrm{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\Theta}_t) = \mathrm{Cov}(\mathbf{Y}_t - \mathbf{F}_t\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_t) = \mathrm{Cov}(\mathbf{Y}_t, \boldsymbol{\Theta}_t) - \mathbf{F}_t\mathrm{Var}(\boldsymbol{\Theta}_t) = \mathbf{0}$, pela Proposição 4.9. □

Proposição 4.18 $\mathrm{Cov}(\boldsymbol{\xi}_t, \boldsymbol{\Theta}_{t-1}) = \mathbf{0}$.

Demonstração.

$$\mathrm{Cov}(\boldsymbol{\xi}_t, \boldsymbol{\Theta}_{t-1}) = \mathrm{Cov}(\boldsymbol{\Theta}_t - \mathbf{G}_t\boldsymbol{\Theta}_{t-1}, \boldsymbol{\Theta}_{t-1}) = \mathrm{Cov}(\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_{t-1})\mathbf{G}_t^\top - \mathbf{G}_t\mathrm{Var}(\boldsymbol{\Theta}_{t-1}) = \mathbf{0},$$

pela Proposição 4.5. □

Proposição 4.19 $\mathrm{Cov}(\boldsymbol{\xi}_t, \mathbf{Y}^{t-1}) = \mathbf{0}$, *em que* $\mathbf{Y}^t$ *denota o vetor* $(\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_t^\top)^\top$.

Demonstração. É suficiente mostrar que $\mathrm{Cov}(\boldsymbol{\xi}_t, \mathbf{Y}_s) = \mathbf{0}$, para $s < t$. Pelas Proposições 2.1 e 2.2,

$$\mathrm{Cov}(\boldsymbol{\xi}_t, \mathbf{Y}_s) = \mathrm{Cov}(\boldsymbol{\Theta}_t - \mathbf{G}_t\boldsymbol{\Theta}_{t-1}, \mathbf{Y}_s) = \mathrm{Cov}\{\mathrm{E}(\boldsymbol{\Theta}_t - \mathbf{G}_t\boldsymbol{\Theta}_{t-1}|\boldsymbol{\theta}^n), \mathrm{E}(\mathbf{Y}_s|\boldsymbol{\theta}^n)\} =$$

$$\mathrm{Cov}(\boldsymbol{\Theta}_t - \mathbf{G}_t\boldsymbol{\Theta}_{t-1}, \mathbf{F}_s\boldsymbol{\Theta}_s) = \mathrm{Cov}(\boldsymbol{\Theta}_t, \boldsymbol{\Theta}_s)\mathbf{F}_s^\top - \mathbf{G}_t\mathrm{Cov}(\boldsymbol{\Theta}_{t-1}, \boldsymbol{\Theta}_s)\mathbf{F}_s^\top,$$

que é igual a zero pela Proposição 4.5, pois $s < t$. □

As Proposições 4.10 a 4.19 correspondem às condições (2.6) a (2.15) do Capítulo 2 (não na mesma ordem). Desta forma, podemos utilizar o filtro e o suavizador de Kalman como obtidos naquele capítulo, isto é, por meio das Proposições 2.11, 2.12 e 2.14, respectivamente.

4.5 Estimação dos parâmetros

Nesta seção, consideramos a estimação dos parâmetros definidos na Seção 4.2, para o caso de h séries ou unidades amostrais independentes, correspondendo a $N = n_1 + \cdots + n_h$ vetores de observações. Como no Capítulo 3, vamos denotar por Y_{ijt} a observação para a categoria i, unidade amostral j e tempo t, $i = 1, \ldots, d$, $j = 1, \ldots h$ e $t = 1, \ldots, n_j$. O vetor de observações e o processo latente no tempo t para a série j são denotados $\mathbf{Y}_{jt}$ e $\boldsymbol{\Theta}_{jt}$, respectivamente, e as correspondentes séries de covariáveis são denotadas $\mathbf{z}_{jt}$ e $\mathbf{x}_{jt}$. Assumimos que a cada série j está associado um terceiro vetor de dimensão l_3 de covariáveis $\mathbf{w}_j$ que são constantes no

tempo e que representam características específicas da j-ésima série de observações. Sejam $g_j, j = 1, \ldots, h$, constantes iniciais e $\Theta_{j0} = (g_j, \ldots, g_j, \theta_{j0})^\top$ um vetor de dimensão $M+1$ que inicia o processo latente. Assumimos que, dado Θ_{j0}, a distribuição condicional do vetor de dados para o j-ésimo indivíduo é representada pelo modelo (4.1)–(4.2), com parâmetros $\boldsymbol{\alpha}_{ij} \in \mathbb{R}^{l_1}, \boldsymbol{\varphi}_{ij} \in \mathbb{R}^M, \boldsymbol{\beta}_j \in \mathbb{R}^{l_2}$ e $\boldsymbol{\psi}_j \in \mathbb{R}^M$ podendo depender de j mas com σ^2 e ν_i^2 independentes de j. Assumimos que

$$\theta_{j0} \sim \mathrm{Tw}_r(g_j, \omega^2), \tag{4.10}$$

e que

$$g_j = \exp\left(\mathbf{w}_j^\top \boldsymbol{\gamma}\right), \tag{4.11}$$

em que $\boldsymbol{\gamma}$ é um terceiro parâmetro de regressão e ω^2 é um parâmetro de dispersão que reflete a variabilidade entre as unidades amostrais. Para este modelo, a esperança marginal das observações é dada por

$$\mathrm{E}(Y_{ijt}) = \exp\left(\mathbf{x}_{jt}^\top \boldsymbol{\alpha}_{ij} + \mathbf{z}_{jt}^\top \boldsymbol{\beta}_j + \mathbf{w}_j^\top \boldsymbol{\gamma}\right).$$

Para encontrar a variância de Θ_{jt}, usamos a expressão $\mathrm{Var}(\Theta_{jt}) = \boldsymbol{\lambda}_{jt} + \mathbf{G}_{jt}\mathrm{Var}(\Theta_{j,t-1})\mathbf{G}_{jt}^\top$, recursivamente, iniciando com

$$\mathrm{Var}(\Theta_{j0}) = \boldsymbol{\lambda}_0 = \begin{pmatrix} \mathbf{0}_{M\times M} & \mathbf{0}_{M\times 1} \\ \mathbf{0}_{1\times M} & \omega^2 g_j^r \end{pmatrix},$$

em que $\omega^2 g_j^r$ é a variância de θ_{j0}, obtida de (4.10). Se interpretarmos $\mathrm{Tw}_r(g, 0)$ como sendo a distribuição degenerada em g e considerarmos $\omega^2 = 0$ em (4.10), obtemos um caso particular do modelo geral (4.1)–(4.2) no qual Θ_{j0} é constante e igual a $(g_j, \ldots, g_j)$. O modelo Tweedie-Tweedie do Capítulo 3 é um caso particular deste modelo, com $\boldsymbol{\varphi}_{ji} = \boldsymbol{\psi}_j = 1$, para $i = 1, \ldots, d$ e $j = 1, \ldots, h$.

Os parâmetros desconhecidos do modelo definido por (4.1)–(4.2) para h unidades amostrais, são $\boldsymbol{\alpha}_{ij}$, $\boldsymbol{\beta}_j$, $\boldsymbol{\varphi}_{ij}^+$, $\boldsymbol{\psi}_j^+$, $\boldsymbol{\gamma}$, ν_i^2, σ^2 e ω^2, em que $^+$ denota os subvetores de dimensões K e $(L-1)$ de componentes desconhecidos em $\boldsymbol{\varphi}_{ij}$ e $\boldsymbol{\psi}_j$, respectivamente. O vetor de dados é definido por

$$\mathbf{Y} = \left(\mathbf{Y}_{11}^\top, \ldots, \mathbf{Y}_{1n_1}^\top, \ldots, \mathbf{Y}_{h1}^\top, \ldots, \mathbf{Y}_{hn_h}^\top\right)^\top.$$

Definimos também os vetores de parâmetros de regressão

$$\boldsymbol{\alpha} = \left(\boldsymbol{\alpha}_{11}^\top, \ldots, \boldsymbol{\alpha}_{1d}^\top, \ldots, \boldsymbol{\alpha}_{h1}^\top, \ldots, \boldsymbol{\alpha}_{hd}^\top\right)^\top, \quad \boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_h^\top\right)^\top \quad \text{e} \quad \boldsymbol{\gamma},$$

os vetores de parâmetros retrospectivos

$$\boldsymbol{\varphi} = \left(\boldsymbol{\varphi}_{11}^\top, \ldots, \boldsymbol{\varphi}_{1d}^\top, \ldots, \boldsymbol{\varphi}_{h1}^\top, \ldots, \boldsymbol{\varphi}_{hd}^\top\right)^\top \quad \text{e} \quad \boldsymbol{\psi} = \left(\boldsymbol{\psi}_1,^\top \ldots, \boldsymbol{\psi}_h^\top\right)^\top,$$

e os vetores de valores suavizados

$$\mathbf{m}^* = (m_{10}^*, \ldots, m_{1n_1}^*, \ldots, m_{h0}^*, \ldots, m_{hn_h}^*)^\top.$$

Vamos denotar por $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top, \boldsymbol{\eta}_3^\top)^\top$ o vetor completo de parâmetros, em que $\boldsymbol{\eta}_1 = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$, $\boldsymbol{\eta}_2 = (\nu_1^2, \ldots, \nu_d^2, \sigma^2, \omega^2)^\top$ e $\boldsymbol{\eta}_3 = (\boldsymbol{\varphi}^\top, \boldsymbol{\psi}^\top)^\top$. Para estimar os parâmetros $\boldsymbol{\eta}_1$ e $\boldsymbol{\eta}_2$, vamos utilizar um procedimento análogo ao descrito no Capítulo 3 para o modelo Tweedie-Tweedie, ou seja, vamos utilizar o algoritmo escores de Newton para resolver equações de estimação de Kalman para os parâmetros de regressão e vamos estimar os parâmetros de dispersão por meio de estimadores ajustados de Pearson. Para $\boldsymbol{\eta}_3$, vamos utilizar um procedimento gráfico baseado na função de verossimilhança perfilada.

Com o objetivo de simplificar a notação, vamos representar o modelo retrospectivo para h séries (ou unidades amostrais) como

$$Y_{ijt}|\boldsymbol{\theta}_j^t \sim Tw_{p_{ij}}\left(\mu_{ijt}^*, \nu_{ijt}^{*2}\right), \qquad \theta_{jt}|\boldsymbol{\theta}_j^{t-1} \sim Tw_{q_j}\left(\mu_{\theta_{jt}}^*, \sigma_t^{*2}\right), \qquad \theta_{j0} \sim Tw_r\left(g_j, \omega\right),$$

em que $\mu_{ijt}^* = \mu_{ijt}(\boldsymbol{\varphi}_i^\top \boldsymbol{\Theta}_{jt})/(\boldsymbol{\varphi}_i^\top \boldsymbol{\tau}_{jt})$, $\nu_{ijt}^{*2} = \nu_i^2/(\boldsymbol{\varphi}_i^\top \boldsymbol{\Theta}_{jt})^{p_i-1}$, $\mu_{\theta_{jt}}^* = \tau_{jt}(\boldsymbol{\psi}^\top \boldsymbol{\Theta}_{j,t-1})/(\boldsymbol{\psi}^\top \boldsymbol{\tau}_{j,t-1})$ e $\sigma_{jt}^{*2} = \sigma^2/(\boldsymbol{\psi}^\top \boldsymbol{\Theta}_{j,t-1})^{q-1}$.

O logaritmo da função de verossimilhança para o vetor completo $(\mathbf{Y}^\top, \boldsymbol{\theta}^\top)^\top$ que seria obtido caso $\boldsymbol{\theta}$ pudesse ser observado, é dado por

$$l(\boldsymbol{\eta}; \mathbf{y}, \boldsymbol{\theta}) = l_1(\boldsymbol{\eta}; \mathbf{y}|\boldsymbol{\theta}) + l_2\{\boldsymbol{\eta}; (\theta_1, \ldots, \theta_n)^\top|\boldsymbol{\theta}_0\} + l_3(\boldsymbol{\eta}; \boldsymbol{\theta}_0), \tag{4.12}$$

em que

$$l_1(\boldsymbol{\eta}; \mathbf{y}|\boldsymbol{\theta}) \;=\; \sum_{j=1}^{h}\sum_{t=1}^{n_j}\sum_{i=1}^{d} c_{p_{ij}}^*(y_{ijt}, \nu_{ij}^2, \varphi_{ij})$$

$$
\begin{aligned}
& + \sum_{j=1}^{h}\sum_{t=1}^{n_j}\sum_{i=1}^{d} \frac{1}{\nu_{ijt}^{2*}} \left\{ Y_{ijt}\mathcal{T}_{p_{ij}}^{-1}(\mu_{ijt}^{*}) - \kappa_{p_{ij}}\left(\mathcal{T}_{p_{ij}}^{-1}(\mu_{ijt}^{*})\right)\right\}, \\
l_2\{\boldsymbol{\eta}; (\theta_1,\ldots,\theta_n)^\top|\boldsymbol{\theta}_0\} &= \sum_{j=1}^{h}\sum_{t=1}^{n_j} c_{q_j}^{*}(\theta_{jt},\sigma^2,\boldsymbol{\psi}_j) \\
& + \sum_{j=1}^{h}\sum_{t=1}^{n_j} \frac{1}{\sigma_t^{2*}} \left\{ \theta_{jt}\mathcal{T}_{q_j}^{-1}(\mu_{\theta_{jt}}^{*}) - \kappa_{q_j}\left(\mathcal{T}_{q_j}^{-1}(\mu_{\theta_{jt}}^{*})\right)\right\}, \\
l_3(\boldsymbol{\eta};\boldsymbol{\theta}_0) &= \sum_{j=1}^{h} c_r^{*}(\theta_{j0},\omega) + \sum_{j=1}^{h}\frac{1}{\omega^2}\left\{\theta_{j0}\mathcal{T}_r^{-1}(g_j) - \kappa_r\left(\mathcal{T}_r^{-1}(g_j)\right)\right\}.
\end{aligned}
$$

Nas expressões acima, $\mathcal{T}$ e κ representam as funções média e cumulante apropriadas (ver Apêndice A ou Jørgensen (1997)).

4.5.1 Parâmetros de regressão

Para obter as funções de estimação de Kalman para os parâmetros de regressão, que denotamos por $\mathcal{F}_1(\boldsymbol{\eta})$, substituímos $\boldsymbol{\theta}$ pelo suavizador de Kalman na função escore obtida da expressão (4.12). O estimador do parâmetro $\boldsymbol{\eta}_1$ é então definido como sendo a solução da equação de estimação $\mathcal{F}_1(\boldsymbol{\eta}) = \mathbf{0}$. As funções escore para os parâmetros de regressão são obtidas facilmente e os cálculos serão omitidos.

O vetor $\mathcal{F}_1(\boldsymbol{\eta})$ tem componentes definidas por

$$
\mathcal{F}_{\alpha_{ij}}(\boldsymbol{\eta}) = \sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\varphi}_{ij}^\top\boldsymbol{\tau}_{jt}\right)^{p_i-1}}{\mu_{ijt}^{p_i-1}} \left(Y_{ijt} - \mu_{ijt}\frac{\boldsymbol{\varphi}_{ij}^\top\mathbf{m}_{jt}^{*}}{\boldsymbol{\varphi}_{ij}^\top\boldsymbol{\tau}_{jt}}\right)\mathbf{x}_{tj},
$$

para $i = 1,\ldots,d$ e $j = 1,\ldots,h$,

$$
\begin{aligned}
\mathcal{F}_{\beta_j}(\boldsymbol{\eta}) &= \sum_{t=1}^{n_j}\sum_{i=1}^{d} \frac{\left(\boldsymbol{\varphi}_{ij}^\top\boldsymbol{\tau}_{jt}\right)^{p_i-2}}{\nu_i^2\mu_{ijt}^{p_i-1}} \left(Y_{ijt} - \mu_{ijt}\frac{\boldsymbol{\varphi}_{ij}^\top\mathbf{m}_{jt}^{*}}{\boldsymbol{\varphi}_{ij}^\top\boldsymbol{\tau}_{jt}}\right)\{\varphi_{ij}(B)(\tau_{jt})\mathbf{z}_{jt} - \varphi_{ij}(B)(\tau_{jt}\mathbf{z}_{jt})\} \\
&+ \sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\psi}_j^\top\boldsymbol{\tau}_{j,t-1}\right)^{q-2}}{\sigma^2\tau_{jt}^{q-1}} \left(m_{jt}^{*} - \tau_{jt}\frac{\boldsymbol{\psi}_j^\top\mathbf{m}_{j,t-1}^{*}}{\boldsymbol{\psi}_j^\top\boldsymbol{\tau}_{j,t-1}}\right)\{\psi_j(B)(\tau_{jt})\mathbf{z}_{jt} - \psi_j(B)(\tau_{jt}\mathbf{z}_{jt})\},
\end{aligned}
$$

para $j = 1, \ldots, h$, e

$$\mathcal{F}_\gamma(\eta) = \sum_{j=1}^{h} \frac{1}{g_j^{r-1}} \left(m_{j0}^* - g_j\right) \mathbf{w}_j.$$

Para utilizar os resultados referentes a funções de estimação regulares, devemos verificar se as condições *i)* a *iv)* da Definição 1.1 são válidas. A condição *i)* é satisfeita, pois ambos os fatores $\left(\boldsymbol{\varphi}_{ij}^\top \mathbf{m}_{jt}^* / \boldsymbol{\varphi}_{ij}^\top \boldsymbol{\tau}_{jt}\right)$ e $\left(\boldsymbol{\psi}_j^\top \mathbf{m}_{j,t-1}^* / \boldsymbol{\psi}_j^\top \boldsymbol{\tau}_{j,t-1}\right)$ que aparecem nas expressões acima têm valor esperado igual a 1, o que implica que a função de estimação é não viciada. Embora todas as derivadas envolvidas tenham sido calculadas sem problemas, as outras condições não foram verificadas diretamente. Convém notar que a condição *iv)* é necessária para que o algoritmo de escore de Newton possa ser utilizado e para que possamos avaliar erros padrões dos estimadores dos parâmetros de regressão. Não tivemos problemas nas aplicações, obtendo sempre matrizes de sensibilidade não singulares. Assumimos então que as funções de estimação de Kalman são regulares e portanto os erros assintóticos dos estimadores podem ser calculados a partir da matriz de informação de Godambe para o conjunto completo de funções de estimação.

A matriz de sensibilidade

Para obter o bloco da matriz de sensibidade correspondente aos parâmetros de regressão, $\mathbf{S}_{\eta_1}$, é necessário primeiramente obter as derivadas do suavizador de Kalman com respeito a estes parâmetros, o que pode ser feito iterativamente. Os detalhes dos cálculos destas derivadas e suas esperanças são apresentados no Apêndice B.3.

Para obter o bloco da matriz de sensibilidade referente ao parâmetro $\boldsymbol{\alpha}$, notamos primeiramente que $\mathrm{E}(m_{ijt}^*) = \tau_{ijt}$, que implica em $\mathrm{E}\left(Y_{ijt} - \mu_{ijt}\{\boldsymbol{\varphi}_{ij}^\top \mathbf{m}_{jt}^* / \boldsymbol{\varphi}_{ij}^\top \boldsymbol{\tau}_{jt}\}\right) = 0$. Logo,

$$\begin{aligned}
\mathrm{E}\left(\frac{\partial \mathcal{F}_{\alpha_{ij}}(\boldsymbol{\eta})}{\partial \boldsymbol{\alpha}_{kj}^\top}\right) &= \sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\varphi}_{ij}^\top \boldsymbol{\tau}_{jt}\right)^{p_i-1}}{\mu_{ijt}^{p_i-1}} \mathrm{E}\left(\frac{\partial}{\partial \boldsymbol{\alpha}_{kj}^\top}\left\{Y_{ijt} - \mu_{ijt}\frac{\boldsymbol{\varphi}_{ij}^\top \mathbf{m}_{jt}^*}{\boldsymbol{\varphi}_{ij}^\top \boldsymbol{\tau}_{jt}}\right\}\right) \mathbf{x}_{jt} \\
&= -\sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\varphi}_{ij}^\top \boldsymbol{\tau}_{jt}\right)^{p_i-1}}{\mu_{ijt}^{p_i-1}} \mathrm{E}\left(\frac{\partial}{\partial \boldsymbol{\alpha}_{kj}^\top} \mu_{ijt} \frac{\boldsymbol{\varphi}_{ij}^\top \mathbf{m}_{jt}^*}{\boldsymbol{\varphi}_{ij}^\top \boldsymbol{\tau}_{jt}}\right) \mathbf{x}_{jt}.
\end{aligned}$$

O parâmetro $\boldsymbol{\tau}_{jt}$ não depende de $\boldsymbol{\alpha}$ e μ_{ijt} só depende de $\boldsymbol{\alpha}_{kj}$ se $i = k$. Logo, fazendo $\mathrm{S}_{\alpha_{ij},\alpha_{kj}} = \mathrm{E}\left(\partial\mathcal{F}(\boldsymbol{\alpha}_{ij})/\partial\boldsymbol{\alpha}_{kj}^{\top}\right)$, obtemos

$$\mathrm{S}_{\alpha_{ij},\alpha_{kj}} = -\sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{jt}\right)^{p_i-2}}{\mu_{ijt}^{p_i-2}} \left\{\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\mathcal{D}}_{\alpha_{kj1}}^{*}\mathbf{x}_{jt}, \ldots, \boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\mathcal{D}}_{\alpha_{kjl_1}}^{*}\mathbf{x}_{jt}\right\}, \quad i \neq k = 1, \ldots, d,\ j = 1, \ldots, h,$$

e

$$\mathrm{S}_{\alpha_{ij},\alpha_{ij}} = -\sum_{t=1}^{n} \frac{\left(\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{jt}\right)^{p_i-1}}{\mu_{ijt}^{p_i-2}} \left\{\left(x_{jt1} + \frac{\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\mathcal{D}}_{\alpha_{kj1}}^{*}(t)}{\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{jt}}\right)\mathbf{x}_{jt}, \ldots, \left(x_{jtl_1} + \frac{\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\mathcal{D}}_{\alpha_{kjl_1}}^{*}(t)}{\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{jt}}\right)\mathbf{x}_{jt}\right\},$$

em que $\boldsymbol{\mathcal{D}}_{\alpha_{ijk}}(t)$ e $\boldsymbol{\mathcal{D}}_{\alpha_{ij}}^{*}(t)$ denotam as esperanças das derivadas do filtro e do suavizador de Kalman, respectivamente, com respeito ao parâmetro $\alpha_{ijk}, i = 1, \ldots d, j = 1, \ldots, d$ e $k = 1, \ldots, l_1$.

Utilizando uma notação similar, obtemos

$$\mathrm{S}_{\alpha_{ij},\beta_{jk}} = -\sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{jt}\right)^{p_i-1}}{\mu_{ijt}^{p_i-2}} \left\{z_{jtk} + \frac{\varphi_{ij}(B)\left(\mathcal{D}_{\beta_{jk}}^{*}(t) - \tau_{jt}z_{jtk}\right)}{\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{jt}}\right\}\mathbf{x}_{jt},$$

para $i = 1, \ldots, d,\ k = 1, \ldots, l_2,\ j = 1, \ldots, h$, em que $\mathcal{D}_{\beta_{jk}}(t)$ e $\mathcal{D}_{\beta_{jk}}^{*}(t)$ denotam os M-ésimos componentes de $\boldsymbol{\mathcal{D}}_{\beta_{jk}}$ e $\boldsymbol{\mathcal{D}}_{\beta_{jk}}^{*}$, respectivamente. Obtemos também

$$\mathrm{S}_{\alpha_{ij},\gamma_k} = -\sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{jt}\right)^{p_i-2}}{\mu_{ijt}^{p_i-2}} \boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\mathcal{D}}_{\gamma_k}^{*}(t)\mathbf{x}_{jt}.$$

Cálculos análogos fornecem

$$\begin{aligned}
\mathrm{S}_{\beta_j,\alpha_{ijk}} &= -\sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{jt}\right)^{p_i-2}}{\nu_i^2\mu_{ijt}^{p_i-2}} \left(x_{jtk} + \frac{\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\mathcal{D}}_{\alpha_{ijk}}^{*}(t)}{\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{jt}}\right)\left\{\varphi_{ij}(B)(\tau_{jt})\mathbf{z}_{jt} - \varphi_{ij}(B)(\tau_{jt}\mathbf{z}_{jt})\right\} \\
&+ \sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\psi}_j^{\top}\boldsymbol{\tau}_{jt-1}\right)^{q-2}}{\sigma^2\tau_{jt}^{q-1}} \left(\mathcal{D}_{\alpha_{ijk}}^{*}(t) - \tau_{jt}\frac{\boldsymbol{\psi}_j^{\top}\boldsymbol{\mathcal{D}}_{\alpha_{ijk}}^{*}(t-1)}{\boldsymbol{\psi}_j^{\top}\boldsymbol{\tau}_{j,t-1}}\right) \\
&\qquad \times \left\{\psi_j(B)(\tau_{jt})\mathbf{z}_{jt} - \psi_j(B)(\tau_{jt}\mathbf{z}_{jt})\right\},
\end{aligned}$$

$i = 1, \ldots, d,\ k = 1, \ldots, l_1, j = 1, \ldots, h,$

$$\mathrm{S}_{\beta_j,\beta_{jk}} = -\sum_{t=1}^{n_j}\sum_{i=1}^{d} \frac{\left(\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{jt}\right)^{p_i-2}}{\nu_i^2\mu_{ijt}^{p_i-2}} \left\{z_{jtk} + \frac{\varphi_{ij}(B)(\mathcal{D}_{\beta_{jk}}^{*}(t) - \tau_{jt}z_{jtk})}{\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{jt}}\right\}$$

$$\times \left\{\varphi_{ij}(B)(\tau_{jt})\mathbf{z}_{jt} - \varphi_{ij}(B)(\tau_{jt}\mathbf{z}_{jt})\right\}$$

$$+ \sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\psi}_j^\top \boldsymbol{\tau}_{jt}\right)^{q-2}}{\sigma^2 \tau_{jt}^{q-1}} \left[\mathcal{D}^*_{\beta_{jk}}(t) - \tau_{jt}\left\{z_{jtk} + \frac{\psi_j(B)\left(\mathcal{D}^*_{\beta_{jk}}(t) - \tau_{jt} z_{jtk}\right)}{\boldsymbol{\psi}_j^\top \boldsymbol{\tau}_{j,t-1}}\right\}\right]$$

$$\times \left\{\psi_j(B)(\tau_{jt})\mathbf{z}_{jt} - \psi_j(B)(\tau_{jt}\mathbf{z}_{jt})\right\}, \; k = i, \ldots, l_2, \; j = 1, \ldots, h,$$

e

$$\begin{aligned}
\mathbf{S}_{\beta_j,\gamma_k} &= -\sum_{t=1}^{n_j}\sum_{i=1}^{d} \frac{\left(\boldsymbol{\varphi}_{ij}^\top \boldsymbol{\tau}_{jt}\right)^{p_i-3}}{\nu_i^2 \mu_{ijt}^{p_i-2}} \left(\boldsymbol{\varphi}_{ij}^\top \boldsymbol{\mathcal{D}}^*_{\gamma_k}(t)\right) \left\{\varphi_{ij}(B)(\tau_{jt})\mathbf{z}_{jt} - \varphi_{ij}(B)(\tau_{jt}\mathbf{z}_{jt})\right\} \\
&+ \sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\psi}_j^\top \boldsymbol{\tau}_{jt-1}\right)^{q-2}}{\sigma^2 \tau_{jt}^{q-1}} \left(\mathcal{D}^*_{\gamma_k}(t) - \tau_{jt} \frac{\boldsymbol{\psi}_j^\top \boldsymbol{\mathcal{D}}^*_{\gamma_k}(t-1)}{\boldsymbol{\psi}_j^\top \boldsymbol{\tau}_{j,t-1}}\right) \\
&\times \left\{\psi_j(B)(\tau_{jt})\mathbf{z}_{jt} - \psi_j(B)(\tau_{jt}\mathbf{z}_{jt})\right\}, j = 1, \ldots, h, \; k = 1, \ldots, l_3.
\end{aligned}$$

Finalmente, obtemos

$$\mathbf{S}_\gamma = \sum_{j=1}^{h} \frac{1}{g_j^{r-1}} \left\{(\mathcal{D}^*_{\gamma_1}(0) - g_j w_{j1}), \ldots, (\mathcal{D}^*_{\gamma_{l_3}}(0) - g_j w_{jl_3})\right\} \mathbf{w}_j,$$

$$\mathbf{S}_{\gamma,\alpha_{ijk}} = \frac{1}{g_j^{r-1}} \mathcal{D}^*_{\alpha_{ijk}}(0)\mathbf{w}_j, \quad j = 1, \ldots, h, \; , i = 1, \ldots, d, k = 1, \ldots, l_1,$$

e

$$\mathbf{S}_{\gamma,\beta_{jk}} = \frac{1}{g_j^{r-1}} \mathcal{D}^*_{\beta_{jk}}(0)\mathbf{w}_j, \quad j = 1, \ldots, h, \; k = 1, \ldots, l_2.$$

Para obter os blocos $\mathbf{S}_{\eta_1,\eta_2}$ da matriz sensibilidade, usamos o fato que $\boldsymbol{\mathcal{D}}^*_{\varphi_{ij}} = \mathbf{0}$. Como a função de estimação $\boldsymbol{\mathcal{F}}_{\alpha_{ij}}(\boldsymbol{\eta})$ depende de $\boldsymbol{\varphi}_{kj}$ apenas por meio de $\mathbf{m}_j^*$ para $i \neq k$, temos $\mathbf{S}_{\alpha_{ij}\varphi_{kj}} = \mathbf{0}$ neste caso. Agora,

$$\mathbf{S}_{\alpha_{ij},\varphi_{ij}} = -\sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\varphi}_{ij}^\top \boldsymbol{\tau}_{jt}\right)^{p_i-1}}{\mu_{ijt}^{p_i-2}} \mathrm{E}\left\{\frac{\mathbf{m}_{jt}^* \boldsymbol{\varphi}_{ij}^\top \boldsymbol{\tau}_{jt} - \boldsymbol{\tau}_{jt} \boldsymbol{\varphi}_{ij}^\top \mathbf{m}_{jt}^*}{\left(\boldsymbol{\varphi}_{ij}^\top \boldsymbol{\tau}_{jt}\right)^2}\right\} \mathbf{x}_t = \mathbf{0},$$

$i = 1, \ldots, d$, $j = 1, \ldots, h$, pois $\mathrm{E}(\mathbf{m}_{jt}^*) = \boldsymbol{\tau}_{jt}$. Temos também que $\boldsymbol{\mathcal{F}}_\alpha(\boldsymbol{\eta})$ depende de $\boldsymbol{\psi}_j$ apenas por meio de $\mathbf{m}_j^*$. Logo, $\mathbf{S}_{\alpha\psi} = \mathbf{0}$. Um resultado análogo se aplica a $\mathbf{S}_{\beta_j\varphi_{ij}}$ para $i = 1, \ldots, d$ e também a $\mathbf{S}_{\eta_1,\eta_3}$. Logo, $\mathbf{S}_{\beta_j\varphi_{ij}} = \mathbf{0}$ e $\mathbf{S}_{\eta_1,\eta_3} = \mathbf{0}$. Obtemos então a seguinte

estrutura para a matriz de sensibilidade do modelo retrospectivo, considerando a função de estimação $\mathcal{F}_{\eta}(\eta)$,

$$\mathrm{S} = \begin{pmatrix} \mathrm{S}_{\eta_1\eta_1} & 0 & 0 \\ \mathrm{S}_{\eta_2\eta_1} & \mathrm{S}_{\eta_2\eta_2} & \mathrm{S}_{\eta_2\eta_3} \\ \mathrm{S}_{\eta_3\eta_1} & \mathrm{S}_{\eta_3\eta_2} & \mathrm{S}_{\eta_3\eta_3} \end{pmatrix} = \begin{pmatrix} \mathrm{S}_1 & 0 \\ \mathrm{S}_{2,3.1} & \mathrm{S}_{2,3} \end{pmatrix}, \tag{4.13}$$

em que $\mathrm{S}_1 = \mathrm{S}_{\eta_1\eta_1}$. Para o modelo Tweedie-Tweedie, esta forma triangular da matriz de sensibilidade foi obtida por Knudsen (1998). A estrutura da matriz de sensibilidade dada em (4.13) permite que os erros padrão dos estimadores de η_1 possam ser calculados como se os parâmetros η_2 e η_3 fossem conhecidos. Uma simplificação no algoritmo de escores de Newton também é obtida graças a esta estrutura.

O algoritmo de escores de Newton

O algoritmo de escores de Newton utiliza a matriz de sensibilidade. Como só calculamos o bloco correspondente aos parâmetros de regressão, só utilizaremos o algoritmo para este grupo de parâmetros. Os outros blocos dependem dos terceiro e quarto momentos do filtro de Kalman, que não calculamos.

O algoritmo de escores de Newton fornece, no passo s, as seguintes atualizações dos parâmetros de regressão,

$$\eta_1^{(s+1)} = \eta_1^{(s)} - \mathrm{S}_{\eta_1}^{-1}\left(\eta^{(s)}\right)\mathcal{F}_1\left(\eta^{(s)}\right), \tag{4.14}$$

em que S_{η_1} é o bloco da matriz sensibilidade correspondente a η_1, isto é,

$$\mathrm{S}_{\eta_1} = \mathrm{E}_{\eta}\left(\frac{\partial \mathcal{F}_{\eta_1}(\eta)}{\partial \eta_1^{\top}}\right).$$

Como veremos a seguir, este algoritmo tem uma forma equivalente a de um algoritmo de escores de Newton para o parâmetro η, que seria obtido caso tivéssemos calculado a matriz de sensibilidade completa.

Consideremos o algoritmo para os dois subconjuntos de η, η_1 e $\eta_{2,3} = (\eta_2^{\top}, \eta_3^{\top})^{\top}$, que forneceria os seguintes valores atualizados no passo s,

$$\eta_1^{(s+1)} = \eta_1^{(s)} - \mathrm{S}_1^{-1}(\eta^{(s)})\mathcal{F}_{\eta_1}(\eta^{(s)}),$$

$$\eta_{2,3}^{(s+1)} = \eta_{2,3}^{(s)} - \mathrm{S}_{2,3}^{-1}(\eta^{(s)})\mathcal{F}_{\eta_{2,3}}(\eta^{(s)}).$$

Agora, considerando um algoritmo de escores de Newton para o vetor completo $\eta = (\eta_1^\top, \eta_2^\top, \eta_3^\top)^\top$, teríamos no passo s os seguintes valores atualizados dos parâmetros

$$\eta^{(s+1)} = \eta^{(s)} - \mathrm{S}_\eta^{-1}(\eta^{(s)})\mathcal{F}_\eta(\eta^{(s)}), \tag{4.15}$$

em que S_η e $\mathcal{F}_\eta$ denotam a matriz sensibilidade e a função de estimação para η, respectivamente. A estrutura da matriz de sensibilidade em (4.13) implica então que

$$\mathrm{S}^{-1} = \begin{pmatrix} \mathrm{S}_1^{-1} & 0 \\ \mathrm{S}^{2,3.1} & \mathrm{S}_{2,3}^{-1} \end{pmatrix},$$

em que $\mathrm{S}^{2,3.1} = -\mathrm{S}_2^{-1}\mathrm{S}_{2,3.1}\mathrm{S}_1^{-1}$. Então, o algoritmo de escores de Newton para o vetor de parâmetros η pode ser expresso como

$$\begin{aligned}
\begin{pmatrix} \eta_1 \\ \eta_{2,3} \end{pmatrix}^{(s+1)} &= \begin{pmatrix} \eta_1 \\ \eta_{2,3} \end{pmatrix}^{(s)} - \begin{pmatrix} \mathrm{S}_1^{-1} & 0 \\ \mathrm{S}^{2,3.1} & \mathrm{S}_{2,3}^{-1} \end{pmatrix}^{(s)} \begin{pmatrix} \mathcal{F}_{\eta_1}(\eta^{(s)}) \\ \mathcal{F}_{\eta_{2,3}}(\eta^{(s)}) \end{pmatrix} \\
&= \begin{pmatrix} \eta_1^{(s+1)} - \mathrm{S}_1^{-1}(\eta^{(s)})\mathcal{F}_{\eta_1}(\eta^{(s)}) \\ \eta_{2,3}^{(s+1)} - \mathrm{S}_{2,3}^{-1}(\eta^{(s)})\mathcal{F}_{\eta_{2,3}}(\eta^{(s)}) - \mathrm{S}^{2,3.1}\mathcal{F}_{\eta_1}(\eta^{(s)}) \end{pmatrix}.
\end{aligned}$$

Logo, a forma combinada do algoritmo em (4.15) para os parâmetros de regressão é equivalente ao algoritmo em (4.14) que utilizamos. O termo $\mathrm{S}^{2,3.1}\mathcal{F}_{\eta_1}(\eta^{(s)})$ na forma combinada do algoritmo em (4.15) para $\eta_{2,3}$ pode ser escrito como $\mathrm{S}_{2,3}^{-1}\mathrm{S}_{2,3.1}(\eta_1^{(s+1)} - \eta_1^{(s)})$. Logo, próximo a uma solução, a forma do algoritmo em (4.14) aproxima-se da forma combinada.

A matriz de variabilidade

Por definição, a matriz de variabilidade para uma dada função de estimação $\mathcal{F}$ é $\mathrm{V} = \mathrm{E}\{\mathcal{F}\mathcal{F}^\top\}$. O bloco da matriz de variabilidade correspondente a $\mathcal{F}_1(\eta)$ é

$$\mathrm{V}_1 = \begin{pmatrix} \mathrm{V}_\alpha & \mathrm{V}_{\alpha,\beta} & \mathrm{V}_{\alpha,\gamma} \\ \mathrm{V}_{\beta,\alpha} & \mathrm{V}_\beta & \mathrm{V}_{\beta,\gamma} \\ \mathrm{V}_{\gamma,\alpha} & \mathrm{V}_{\gamma,\beta} & \mathrm{V}_\gamma \end{pmatrix},$$

e ela pode ser calculada a partir dos resultados sobre as covariâncias dos resíduos obtidos na Seção 4.6 (ver detalhes na Seção B.4). Seja $\mathbf{J}$ a matriz de informação de Godambe, dada por $\mathbf{S}^\top\mathbf{V}^{-1}\mathbf{S}$. Então, a matriz de covariância assintótica dos estimadores é dada por $\mathbf{J}^{-1} = \mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1^\top}$. Devido à particular estrutura da matriz de sensibilidade em (4.13), obtemos que

$$\mathbf{J}^{-1} = \begin{pmatrix} \mathbf{S}_1^{-1}\mathbf{V}_1\mathbf{S}_1^{-1^\top} & \mathbf{J}^{1.2,3} \\ \mathbf{J}^{2,3.1} & \mathbf{J}^{2,3} \end{pmatrix}.$$

4.5.2 Parâmetros de dispersão

A cada passo do algoritmo de escores de Newton para os parâmetros de regressão, os parâmetros de dispersão são atualizados. Para isso, utilizamos estimadores de Pearson modificados de forma a obter funções de estimação não viciadas.

As funções de estimação para os parâmetros de dispersão são obtidas a partir das variâncias das inovações. Da Proposição 4.3 temos que $\mathrm{Var}(\xi_{jt}) = \lambda_{jt} = \sigma^2\tau_{jt}^q/(\boldsymbol{\psi}^\top\boldsymbol{\tau}_{j,t-1})$ e da Proposição 4.16, temos que $\mathrm{Var}(\varepsilon_{ijt}) = (\nu_i^2\mu_{ijt}^{p_i}/(\boldsymbol{\varphi}_i^\top\boldsymbol{\tau}_{jt})^{p_i-1}$. Temos então as seguintes expressões para os parâmetros ν_i^2, σ^2 e ω^2,

$$\nu_i^2 = \mathrm{Var}(\varepsilon_{ijt})\frac{(\boldsymbol{\varphi}_i^\top\boldsymbol{\tau}_{jt})^{p_i-1}}{\nu_i^2\mu_{ijt}^{p_i}}, \quad \sigma^2 = \mathrm{Var}(\xi_{jt})\frac{(\boldsymbol{\psi}^\top\boldsymbol{\tau}_{j,t-1})^{q-1}}{\tau_{jt}^q} \quad \text{e} \quad \omega^2 = \frac{\mathrm{Var}(\theta_{jo})}{g_j^r}.$$

Os parâmetros de dispersão são então definidos a partir dos resíduos que podem ser expressos como sendo as inovações *preditas* (ver Seção 4.6). Os resíduos calculados a partir do suavizador de Kalman são não viciados e as equações de estimação para os parâmetros ν_i^2, σ^2 e ω^2 são obtidas corrigindo-se o vício das somas $\sum\hat{\xi}_{jt}^{*2}$, $\sum\hat{\varepsilon}_{ijt}^{*2}$ e $\sum(m_{j0}^* - g_j)^2$. As atualizações são dadas por

$$\hat{\nu}_i^2 = \frac{1}{N}\sum_{j=1}^{h}\sum_{t=1}^{n_j}\frac{\left(\boldsymbol{\varphi}_{ij}^\top\boldsymbol{\tau}_{jt}\right)^{p_i-1}}{\mu_{ijt}^{p_i}}\hat{\varepsilon}_{ijt}^{*2} + \frac{1}{N}\sum_{j=1}^{h}\sum_{t=1}^{n_j}\frac{\left(\boldsymbol{\varphi}_{ij}^\top\boldsymbol{\tau}_{jt}\right)_{ij}^{p_i-3}}{\mu_{ijt}^{p_i-1}}\boldsymbol{\varphi}_{ij}^\top\mathbf{C}_{jt}^*\boldsymbol{\varphi}_{ij},$$

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{N}\sum_{j=1}^{h}\sum_{t=1}^{n_j}\frac{\left(\boldsymbol{\psi}_j^\top\boldsymbol{\tau}_{jt-1}\right)^{q-1}}{\tau_{jt}^q}\hat{\xi}_{jt}^{*2} \\ &+ \frac{1}{N}\sum_{j=1}^{h}\sum_{t=1}^{n_j}\frac{\left(\boldsymbol{\psi}_j^\top\boldsymbol{\tau}_{jt-1}\right)^{q-1}}{\tau_{jt}^q}\left(\mathbf{a}^\top\mathbf{C}_{jt}^*\mathbf{a} - \mathbf{a}^\top\mathbf{C}_{jt,t-1}^*c_{jt}\boldsymbol{\psi}_j - c_{jt}\boldsymbol{\psi}^\top\mathbf{C}_{jt-1,t}^*\mathbf{a} + c_{jt}^2\boldsymbol{\psi}_j^\top\mathbf{C}_{jt-1}^*\boldsymbol{\psi}_j\right),\end{aligned}$$

em que $N = \sum_{j=1}^{h} n_j$, e

$$\hat{\omega}^2 = \frac{1}{h}\sum_{j=1}^{h}\frac{(m_{j0}^* - g_j)^2}{g_j^r} + \frac{1}{h}\sum_{j=1}^{h}\frac{\mathrm{C}_{j0}^*}{g_j^r},$$

com $c_{jt} = \tau_{jt}/(\boldsymbol{\psi}_j^\top \boldsymbol{\tau}_{j,t-1})$ e $\mathrm{C}_{t,t-1}^* = \mathrm{C}_{t-1}\mathrm{G}_t^\top \mathrm{B}_t^{-1}\mathrm{C}_t^*$.

4.5.3 Parâmetros retrospectivos

Nesta seção, vamos discutir a estimação dos parâmetros retrospectivos $\boldsymbol{\varphi}_i, i = 1, \ldots, d$ e $\boldsymbol{\psi}$. A estimação dos parâmetros retrospectivos é dificultada por uma série de fatores. Por exemplo, a função escore, que mostramos abaixo para uma única unidade amostral, não tem uma forma única para todos os elementos da família Tweedie. Mesmo nos casos mais simples como das distribuições normal e Poisson, a função escore para estes parâmetros é não linear no processo latente, o que torna a utilização das funções de estimação de Kalman pouco eficiente. Esta não linearidade não permite, por exemplo, que se verifique se as funções obtidas com a substituição do processo latente pelo suavizador de Kalman nas funções escore são viciadas ou não. Uma outra dificuldade surge da restrição imposta aos parâmetros (as componentes dos parâmetros retrospectivos somam 1), sem a qual o modelo se torna não identificável.

Nesta seção, propomos um procedimento para estimação dos parâmetros retrospectivos que se baseia na função de verossimilhança perfilada (ver, por exemplo, Cordeiro (1992)). Na Seção 4.8, discutimos brevemente uma série de outras alternativas que consideramos inicialmente.

A função escore para os parâmetros retrospectivos

No que segue, por simplicidade, nos restringimos ao caso de uma única unidade amostral. Neste caso, a função escore para os parâmetros retrospectivos é dada por

$$\begin{aligned}\mathrm{U}(\varphi_{ik}) &= \sum_{t=1}^{n}\frac{\partial}{\partial\varphi_{ik}}\log\left(c_{p_i}(y_{it},\nu_i^2,\boldsymbol{\varphi}_i)\right) + \sum_{t=1}^{n}\left\{y_{it}\mathcal{T}_p(\mu_{it}^*) - \kappa_p(\mu_{it}^*)\right\}\frac{\partial}{\partial\varphi_{ik}}(\nu_{it}^*)^{-1} \\ &+ \sum_{t=1}^{n}\frac{\left(\boldsymbol{\varphi}_i^\top\boldsymbol{\tau}_t\right)^{p_i-1}}{\nu_i^2\mu_{it}^{p_1-1}}(y_{it}-\mu_{it}^*)\left\{\frac{\theta_{t-K+k}}{\boldsymbol{\varphi}_i^\top\boldsymbol{\Theta}_t} - \frac{\tau_{t-K+k,}}{\boldsymbol{\varphi}_i^\top\boldsymbol{\tau}_t}\right\},\end{aligned} \tag{4.16}$$

que escreveremos como

$$\mathrm{U}_{1\varphi}(\varphi_{ik}) + \mathrm{U}_{2\varphi}(\varphi_{ik}), \quad k = 0, \ldots, K,$$

em que

$$\mathrm{U}_{2\varphi}(\varphi_{ik}) = \sum_{t=1}^{n} \frac{\left(\boldsymbol{\varphi}_i^\top \boldsymbol{\tau}_t\right)^{p_i - 1}}{\nu_i^2 \mu_{it}^{p_1 - 1}} (y_{it} - \mu_{it}^*) \left\{ \frac{\theta_{t-K+k}}{\boldsymbol{\varphi}_i^\top \boldsymbol{\Theta}_t} - \frac{\tau_{t-K+k,}}{\boldsymbol{\varphi}_i^\top \boldsymbol{\tau}_t} \right\},$$

e por

$$\begin{aligned} \mathrm{U}(\psi_k) &= \sum_{t=1}^{n} \frac{\partial}{\partial \psi_k} \log\left(c_q(\theta_t, \sigma^2, \boldsymbol{\psi})\right) + \sum_{t=1}^{n} \left\{ \theta_t \mathcal{T}_q(\mu_{\theta_t}^*) - \kappa_q(\mu_{\theta_t}^*) \right\} \frac{\partial}{\partial \psi_j} (\sigma_t^*)^{-1} \\ &+ \sum_{t=1}^{n} \frac{\left(\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1}\right)^{q-1}}{\sigma^2 \tau_t^{q-1}} (\theta_t - \mu_{\theta_t}^*) \left\{ \frac{\theta_{t-L-1+k}}{\boldsymbol{\psi}^\top \boldsymbol{\Theta}_{t-1}} - \frac{\tau_{t-L-1+k}}{\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1}} \right\} \end{aligned} \tag{4.17}$$

que escreveremos como

$$\mathrm{U}_{1\psi}(\psi_k) + \mathrm{U}_{2\psi}(\psi_k), \quad k = 1, \ldots, L$$

em que

$$\mathrm{U}_{2\psi}(\psi_k) = \sum_{t=1}^{n} \frac{\left(\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1}\right)^{q-1}}{\sigma^2 \tau_t^{q-1}} (\theta_t - \mu_{\theta_t}^*) \left\{ \frac{\theta_{t-L-1+k}}{\boldsymbol{\psi}^\top \boldsymbol{\Theta}_{t-1}} - \frac{\tau_{t-L-1+k}}{\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1}} \right\}.$$

$\mathrm{U}_{1\varphi}(\varphi_{ik})$ e $\mathrm{U}_{1\psi}(\psi_k)$ são obtidas por diferença.

Substituir o processo latente pelo suavizador de Kalman nas equações acima produz, tanto para $\boldsymbol{\varphi}_i$ quanto para $\boldsymbol{\psi}$, funções de estimação difíceis de serem trabalhadas. Uma alternativa seria considerar apenas as funções $\mathrm{U}_{2\varphi}$ e $\mathrm{U}_{2\psi}$ que são lineares nos parâmetros retrospectivos, corrigidas a fim de obter funções não viciadas. Este procedimento foi tentado mas não produziu bons resultados (ver Seção 4.8). Com o objetivo de evitar as dificuldades acarretadas pelas derivadas em relação aos parâmetros retrospectivos, consideramos então a função de verossimilhança do modelo retrospectivo, substituindo o processo latente pelo suavizador de Kalman.

Sejam $\mathcal{L}_1$ e $\mathcal{L}_2$ as funções definidas por

$$\begin{aligned} \mathcal{L}_1(\boldsymbol{\eta}_3; \mathbf{y}, \hat{\boldsymbol{\eta}}_{1,2}) &= l_1\{(\hat{\boldsymbol{\eta}}_{1,2}^\top, \boldsymbol{\eta}_3^\top)^\top; \mathbf{y}|\mathbf{m}^*)\}, \\ \mathcal{L}_2(\boldsymbol{\eta}_3; \mathbf{y}, \hat{\boldsymbol{\eta}}_{1,2}) &= l_2\{(\hat{\boldsymbol{\eta}}_{1,2}^\top, \boldsymbol{\eta}_3^\top)^\top; (m_1^*, \ldots, m_n^*)|\mathbf{m}_0^*\} \end{aligned}$$

e seja $\mathcal{L}(\eta_3, \mathrm{y}, \hat{\eta}_{1,2}) = \mathcal{L}_1(\eta_3, \mathrm{y}, \hat{\eta}_{1,2}) + \mathcal{L}_2(\eta_3, \mathrm{y}, \hat{\eta}_{1,2})$, de modo que

$$\mathcal{L}(\eta_3; \mathrm{y}, \hat{\eta}_{1,2}) = l\{(\hat{\eta}_{1,2}^\top, \eta_3^\top)^\top; \mathrm{y}, \mathrm{m}^*\},$$

em que l, l_1 e l_2 são dadas por (4.12) para uma única unidade amostral, $\eta_{1,2}$ denota o vetor $(\eta_1, \eta_2)^\top$ e $\hat{\eta}_{1,2}$ é obtido a partir dos procedimentos de estimação de η_1 e η_2 descritos na Seção 4.5.

Para obter o estimador de η_3, vamos considerar inicialmente três possibilidades, definindo $\hat{\eta}_3$ como sendo o valor de η_3 que maximiza uma das três funções $\mathcal{L}, \mathcal{L}_1$ ou $\mathcal{L}_2$. Embora o estimador obtido a partir de $\mathcal{L}$ pareça ser o mais natural por se basear na função de verossimilhança completa, o estimador obtido a partir de $\mathcal{L}_1$ apresentou melhores resultados. Um dos prováveis motivos para este fato é que $\mathcal{L}_1$ é a parte da verossimilhança correspondente às observações $\mathbf{Y}$, fornecendo, desta forma, o estimador que corresponde ao valor do parâmetro que melhor se ajusta aos dados.

Para obter $\hat{\eta}_3$ baseado na função $\mathcal{L}$, utilizamos o gráfico desta função obtido a partir de uma partição do espaço paramétrico. Fixamos um valor de η_3 em cada elemento da partição e, a partir dele, encontramos o estimador $\hat{\eta}_{1,2}$ e calculamos o valor da função $\mathcal{L}$. O valor $\hat{\eta}_3$ é aquele que fornece o maior valor de $\mathcal{L}$ entre os calculados. Este procedimento aproximado se fez necessário devido ao fato de que a obtenção de $\hat{\eta}_{1,2}$ para η_3 fixado é um processo iterativo computacionalmente dispendioso devido ao cálculo do suavizador de Kalman. O procedimento gráfico permite que a inclusão das restrições nos parâmetros seja feita de forma simples por meio do espaço paramétrico considerado. O método gráfico se mostrou razoavelmente eficiente nas simulações discutidas no Seção 4.7 e descritas abaixo para um caso particular.

Exemplo 2 Consideremos o modelo retrospectivo de ordem 2, com $K = 0$ e $L = 2$. Temos então $\boldsymbol{\varphi} = (0, 1)^\top$ e $\boldsymbol{\psi} = (\psi_1, \psi_2)^\top$, com $\psi_1 + \psi_2 = 1$ e $\psi_1, \psi_2 > 0$. Consideramos o coeficiente de θ_{t-1} no produto $\boldsymbol{\psi}^\top \boldsymbol{\Theta}_{t-1}$, ψ_2, como sendo o parâmetro desconhecido, podendo variar no intervalo $[0, 1]$.

O procedimento de estimação de ψ_2 consiste em considerar uma partição do espaço paramétrico, que neste caso é o intervalo $[0, 1]$, representada por um conjunto de valores

eqüiespaçados, digamos, $\psi_{2,1}, \ldots, \psi_{2,a}$, e para cada um destes valores, calcular a função $\mathcal{L}(\boldsymbol{\eta}_3(\boldsymbol{\psi}_j); \mathbf{y}, \hat{\boldsymbol{\eta}}_{1,2}(\boldsymbol{\psi}_j))$, em que $\boldsymbol{\psi}_j$ denota o vetor $(1 - \psi_{2,j}, \psi_{2,j})^\top, j = 1, \ldots, a$. Obtemos então a valores para esta função com os quais construímos o gráfico a partir do qual determinamos o valor de ψ_2 que forneceu o maior valor de $\mathcal{L}$. O gráfico fornece uma idéia da forma da função no intervalo $[0, 1]$ e pode indicar, por exemplo, a necessidade de se investigar uma partição mais refinada. A princípio, este procedimento pode ser realizado recursivamente e uma boa aproximação do gráfico de $\mathcal{L}$ pode ser obtido em uma vizinhança do máximo. Nas simulações da Seção 4.7, consideramos apenas uma partição com 11 valores de ψ_2.

Este método pode ser aplicado a qualquer membro da família Tweedie, desde que a função de verossimilhança seja explicitada. A busca do máximo não precisa ser feita graficamente e portanto o método pode ser aplicado para modelos de ordem mais alta. Neste caso no entanto, o espaço paramétrico não será mais um subconjunto de $\mathbb{R}$ e uma partição que permita uma aproximação satisfatória do gráfico de $\mathcal{L}$ deve envolver um número muito grande de valores dos parâmetros a serem considerados, tornando o processo de estimação extremamente lento. Algumas características da função $\mathcal{L}$ são comentadas no Seção 4.7.

Este procedimento é análogo ao de encontrar o estimador de um parâmetro na presença de parâmetros de perturbação por meio da função de verossimilhança perfilada. Denotando por

$$l^*(\boldsymbol{\eta}_3; \mathbf{y}, \boldsymbol{\theta}) = \sup_{\boldsymbol{\eta}_{1,2}} l(\boldsymbol{\eta}_3, \boldsymbol{\eta}_{1,2}; \mathbf{y}, \boldsymbol{\theta}),$$

o logaritmo da função de verossimilhança perfilada para o modelo retrospectivo, o estimador de $\boldsymbol{\eta}_3$ seria o valor que maximiza l^*. A função $\mathcal{L}$ difere de l^* devido à substituição de $\boldsymbol{\theta}$ pelo suavizador de Kalman e ao fato de que os estimadores de $\boldsymbol{\eta}_{1,2}$ que consideramos não são estimadores de máxima verossimilhança.

4.6 Resíduos no modelo retrospectivo

Nesta seção, apresentamos alguns resultados sobre os resíduos do modelo retrospectivo que podem ser definidos de forma análoga aos resíduos do modelo Tweedie-Tweedie do Capítulo 3.

Consideramos o modelo retrospectivo na forma de modelo de espaço de estados a partir do qual podemos definir resíduos para os processos observado e latente. Vamos considerar resíduos baseados no filtro de Kalman, no suavizador de Kalman e também em um terceiro preditor linear de Θ_t dado apenas o passado das observações, $\mathbf{Y}^{t-1}$. Denotaremos a parte linear deste preditor por $\mathbf{m}_t^+$. Os resíduos baseados em um particular preditor linear de Θ_t, digamos $\mathbf{m}_{\Theta_t|,\mathbf{Y}}$, são definidos como sendo as inovações preditas obtidas a partir de (4.8)–(4.9). Podemos obter também *valores preditos* a partir destes preditores lineares, que serão dados pela média condicional de $\mathbf{Y}_t$ estimada a partir do preditor em questão. Chamaremos de *valores estimados* os componentes da média marginal estimada de $\mathbf{Y}_t$.

Todos os resíduos têm média zero. Para obter sua estrutura de covariância, utilizamos os resultados sobre preditores lineares do Capítulo 2. Todas as covariâncias envolvendo um preditor linear de $\boldsymbol{\theta}$ podem ser expressas em termos das covariâncias de $\boldsymbol{\theta}$ e da matriz de erros quadráticos médios associada a este preditor. Na Tabela 4.6, apresentamos um resumo dos resultados e os detalhes dos cálculos são apresentados no Apêndice B.2.

Os gráficos dos resíduos padronizados são utilizados para checar a estrutura do modelo retrospectivo da mesma forma que no modelo Tweedie-Tweedie (ver Seção 3.5 para detalhes sobre estes gráficos). Tanto os resíduos construídos a partir de $\mathbf{m}^+$ quanto os obtidos a partir do filtro, são não correlacionados e portanto, úteis para a construção destes gráficos.

Vamos então definir os seguintes resíduos,

- utilizando $\mathbf{m}^+$: $\hat{\varepsilon}_t^+ = \mathbf{Y}_t - \mathbf{F}_t\mathbf{m}_t^+$, e $\hat{\boldsymbol{\xi}}_t^+ = \mathbf{m}_t^+ - \mathbf{G}_t\mathbf{m}_{t-1}^+$,
- utilizando $\mathbf{m}$: $\hat{\varepsilon}_t = \mathbf{Y}_t - \mathbf{F}_t\mathbf{m}_t$, e $\hat{\boldsymbol{\xi}}_t = \mathbf{m}_t - \mathbf{G}_t\mathbf{m}_{t-1}$, (filtro)
- utilizando $\mathbf{m}^*$: $\hat{\varepsilon}_t^* = \mathbf{Y}_t - \mathbf{F}_t\mathbf{m}_t^*$, e $\hat{\boldsymbol{\xi}}_t^* = \mathbf{m}_t^* - \mathbf{G}_t\mathbf{m}_{t-1}^*$, (suavizador)

em que $\mathbf{m}_t^+ = \mathbf{G}_t\mathbf{m}_{t-1}$ (ver Seção 4.4).

4.7 Simulações

Nesta seção descrevemos os resultados obtidos utilizando o método descrito na Seção 4.5.3 em algumas amostras simuladas com o modelo retrospectivo de ordem 2. Simulações do modelo

Tabela 4.1: Resíduos condicionais baseados em preditores lineares

Tipo	Fonte	Definição	Variância	Covariância
$\Theta_t \vert \mathbf{Y}^{t-1}$	$\mathbf{Y}$	$\hat{\varepsilon}_t^+ = \mathbf{Y}_t - \mathbf{F}_t \mathbf{m}_t^+$	$\mathbf{Q}_t$	$\mathrm{Cov}(\hat{\varepsilon}_t^+, \hat{\varepsilon}_s^+) = \mathbf{0}, s \neq t,$ $\mathrm{Cov}(\hat{\varepsilon}_t^+, \hat{\xi}_s^+) = \begin{cases} \mathbf{F}_t \mathbf{B}_t \mathbf{G}_s^\top, & \text{s=t+1}, \\ \mathbf{0}, & \text{caso contrário} \end{cases}$
	$\boldsymbol{\theta}$	$\hat{\boldsymbol{\xi}}_t^+ = \mathbf{m}_t^+ - \mathbf{G}_t \mathbf{m}_{t-1}^+$	$\mathbf{G}_t \mathbf{B}_{t-1} \mathbf{F}_{t-1} \mathbf{Q}_{t-1}^{-1} \mathbf{F}_{t-1}^\top \mathbf{B}_{t-1}^\top \mathbf{G}_t^\top$	$\mathrm{Cov}(\hat{\boldsymbol{\xi}}_t^+, \hat{\boldsymbol{\xi}}_s^+) = \mathbf{0}, s \neq t.$
$\Theta_t \vert \mathbf{Y}^t$	$\mathbf{Y}$	$\hat{\varepsilon}_t = \mathbf{Y}_t - \mathbf{F}_t \mathbf{m}_t$	$\boldsymbol{\Lambda}_t - \mathbf{F}_t \mathbf{C}_t \mathbf{F}_t^\top$	$\mathrm{Cov}(\hat{\varepsilon}_t, \hat{\varepsilon}_s) = \mathbf{0}$ $\mathrm{Cov}(\hat{\varepsilon}_t, \hat{\boldsymbol{\xi}}_s) = \begin{cases} \mathbf{F}_t \mathbf{C}_t, & \text{s=t}, \\ \mathbf{0}, & \text{caso contrário} \end{cases}$
	$\boldsymbol{\theta}$	$\hat{\boldsymbol{\xi}}_t = \mathbf{m}_t - \mathbf{G}_t \mathbf{m}_{t-1}$	$\mathbf{B}_t \mathbf{F}_t \mathbf{Q}_t^{-1} \mathbf{F}_t^\top \mathbf{B}_t^\top$	$\mathrm{Cov}(\hat{\boldsymbol{\xi}}_t, \hat{\boldsymbol{\xi}}_s) = \mathbf{0}, s \neq t.$
$\Theta_t \vert \mathbf{Y}^n$	$\mathbf{Y}$	$\hat{\varepsilon}_t^* = \mathbf{Y}_t - \mathbf{F}_t \mathbf{m}_t^*$	$\boldsymbol{\Lambda}_t - \mathbf{F}_t \mathbf{C}_t^* \mathbf{F}_t^\top$	$-\mathbf{F}_t \mathbf{C}_s^* \mathbf{F}_s^\top, \ s > t,$ $\mathrm{Cov}(\hat{\varepsilon}_t^*, \hat{\boldsymbol{\xi}}_s^*) = \mathbf{F}_s \mathbf{C}_{s,t}^* - \mathbf{F}_s \mathbf{C}_{s,t-1}^* \mathbf{G}_t^\top$
	$\boldsymbol{\theta}$	$\hat{\boldsymbol{\xi}}_t^* = \mathbf{m}_t^* - \mathbf{G}_t \mathbf{m}_{t-1}^*$	veja abaixo	

$$\mathrm{Var}(\hat{\boldsymbol{\xi}}_t^*) = \boldsymbol{\lambda}_t - \mathbf{C}_t^* - \mathbf{G}_t \mathbf{C}_{t-1}^* \mathbf{G}_t^\top + \mathbf{C}_{t,t-1}^* \mathbf{G}_t^\top + \mathbf{G}_t \mathbf{C}_{t-1,t}^*$$

$$\mathrm{Cov}(\hat{\boldsymbol{\xi}}_t^*, \hat{\boldsymbol{\xi}}_s^*) = -\mathbf{C}_{t,s}^* + \mathbf{C}_{t,s-1}^* \mathbf{G}_s^\top - \mathbf{G}_t \mathbf{C}_{t-1,s-1}^* \mathbf{G}_s^\top + \mathbf{G}_t \mathbf{C}_{t-1,s}^*$$

retrospectivo com o objetivo de investigar as propriedades dos estimadores são praticamente inviáveis de serem conduzidas dada a natureza do processo de estimação que consideramos. No entanto, simulações de algumas amostras para casos particulares servem para dar uma idéia da estabilidade do método.

Simulamos algumas amostras, todas de tamanho 200, com valores fixos dos parâmetros $\boldsymbol{\eta}_1$ (regressão) e $\boldsymbol{\eta}_2$ (dispersão) para dois modelos: normal-gama e Poisson-gama, considerando $L = 2$. As covariáveis foram também simuladas e são as mesmas para todas as amostras consideradas.

O modelo normal-gama considerado nas simulações ($h = 1,\ d = 1$) é dado por:

$$Y_t \sim \mathrm{N}\left(\mu_t \frac{\boldsymbol{\varphi}^\top \boldsymbol{\Theta}_t}{\boldsymbol{\varphi}^\top \boldsymbol{\tau}_t}, \nu^2 \boldsymbol{\varphi}^\top \boldsymbol{\Theta}_t\right), \quad \theta_t \sim \mathrm{Ga}\left(\tau_t \frac{\boldsymbol{\psi}^\top \boldsymbol{\Theta}_{t-1}}{\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1}}, \frac{\sigma^2}{\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1}}\right),$$

em que $N(\mu, \sigma^2)$ denota a distribuição normal com média μ e variância σ^2. Temos então o seguinte logaritmo da função de verossimilhança,

$$\begin{aligned} l(\mathbf{y}, \boldsymbol{\eta}) \;=\; & -\frac{n}{2}\log(2\pi\nu^2) - \frac{1}{2}\sum_{t=1}^{n}\left\{\log(\boldsymbol{\varphi}^\top \boldsymbol{\Theta}_t) + \frac{1}{\nu^2 \boldsymbol{\varphi}^\top \boldsymbol{\Theta}_t}\left(y_t - \frac{\mu_t \boldsymbol{\varphi}^\top \boldsymbol{\Theta}_t}{\boldsymbol{\varphi}^\top \boldsymbol{\tau}_t}\right)^2\right\} + \\ & \sum_{t=1}^{n}\left[\left(\frac{\boldsymbol{\psi}^\top \boldsymbol{\Theta}_{t-1}}{\sigma^2}\right)\log\left\{\frac{\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1}}{\sigma^2 \tau_t}\right\} - \log\left\{\Gamma\left(\frac{\boldsymbol{\psi}^\top \boldsymbol{\Theta}_{t-1}}{\sigma^2}\right)\right\} + \right. \\ & \left.\left(\frac{\boldsymbol{\psi}^\top \boldsymbol{\Theta}_{t-1}}{\sigma^2} - 1\right)\log(\theta_t) - \frac{\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1}}{\sigma^2 \tau_t}\theta_t\right]. \end{aligned} \tag{4.18}$$

O modelo Poisson-gama é dado por

$$Y_t \sim \mathrm{Po}\left(\mu_t \frac{\boldsymbol{\varphi}^\top \boldsymbol{\Theta}_t}{\boldsymbol{\varphi}^\top \boldsymbol{\tau}_t}\right), \quad \theta_t \sim \mathrm{Ga}\left(\tau_t \frac{\boldsymbol{\psi}^\top \boldsymbol{\Theta}_{t-1}}{\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1}}, \frac{\sigma^2}{\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1}}\right),$$

com o logaritmo da função de verossimilhança dado por

$$\begin{aligned} l(\mathbf{y}, \boldsymbol{\eta}) \;=\; & \sum_{t=1}^{n}\left\{-\mu_t \frac{\boldsymbol{\varphi}^\top \boldsymbol{\Theta}_t}{\boldsymbol{\varphi}^\top \boldsymbol{\tau}_t} + y_t \log\left(\frac{\mu_t \boldsymbol{\varphi}^\top \boldsymbol{\Theta}_t}{\boldsymbol{\varphi}^\top \boldsymbol{\tau}_t}\right) - \log(y_t!)\right\} + \\ & \sum_{t=1}^{n}\left[\left(\frac{\boldsymbol{\psi}^\top \boldsymbol{\Theta}_{t-1}}{\sigma^2}\right)\log\left\{\frac{\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1}}{\sigma^2 \tau_t}\right\} - \log\left\{\Gamma\left(\frac{\boldsymbol{\psi}^\top \boldsymbol{\Theta}_{t-1}}{\sigma^2}\right)\right\} + \right. \\ & \left.\left(\frac{\boldsymbol{\psi}^\top \boldsymbol{\Theta}_{t-1}}{\sigma^2} - 1\right)\log(\theta_t) - \frac{\boldsymbol{\psi}^\top \boldsymbol{\tau}_{t-1}}{\sigma^2 \tau_t}\theta_t\right]. \end{aligned} \tag{4.19}$$

Tabela 4.2: Algumas amostras simuladas, modelo normal-gama, $\varphi = (0,1)^\top$, $\sigma^2 = 0.01$

Amostra	parâmetros na simulação		ajuste		
	$\psi^\top$	ν^2	$\hat{\psi}^{(a)\top}$	$\hat{\psi}^{(b)\top}$	$\hat{\psi}^{(c)\top}$
1	$(0.7, 0.3)$	10	$(0.0, 1.0)$	$(0.8, 0.2)$	$(0.0, 1.0)$
2	$(0.7, 0.3)$	10	$(0.0, 1.0)$	$(0.7, 0.3)$	$(0.0, 1.0)$
3	$(0.7, 0.3)$	10	$(0.6, 0.4)$	$(0.6, 0.4)$	$(0.0, 1.0)$
4	$(0.7, 0.3)$	10	$(0.0, 1.0)$	$(0.6, 0.4)$	$(0.0, 1.0)$
5	$(0.3, 0.7)$	10	$(0.0, 1.0)$	$(0.5, 0.5)$	$(0.0, 1.0)$
6	$(0.3, 0.7)$	10	$(0.0, 1.0)$	$(0.5, 0.5)$	$(0.0, 1.0)$
7	$(0.3, 0.7)$	10	$(0.0, 1.0)$	$(0.5, 0.5)$	$(0.0, 1.0)$
8	$(0.3, 0.7)$	1	$(0.4, 0.6)$	$(0.4, 0.6)$	$(0.0, 1.0)$
9	$(0.7, 0.3)$	1	$(0.6, 0.4)$	$(0.6, 0.4)$	$(0.0, 1.0)$
10	$(0.0, 1.0)$	1	$(0.3, 0.7)$	$(0.3, 0.7)$	$(0.0, 1.0)$
11	$(0.0, 1.0)$	1	$(0.0, 1.0)$	$(0.0, 1.0)$	$(0.0, 1.0)$
12	$(1.0, 0.0)$	1	$(0.8, 0.2)$	$(0.8, 0.2)$	$(0.0, 1.0)$
13	$(1.0, 0.0)$	1	$(0.9, 0.1)$	$(0.9, 0.1)$	$(0.0, 1.0)$

A função $\mathcal{L}(\eta_3; \mathbf{y})$ é obtida substituindo Θ_t por $\mathbf{m}_t^*$, o suavizador de Kalman, nas expressões (4.18) e (4.19). Vale notar que a função $\mathcal{L}_1$ não depende funcionalmente de ψ.

Consideramos onze valores distintos de ψ, fazendo o coeficiente de θ_{t-1} na expressão $\psi^\top \Theta_{t-1}$ variar variar entre 0 e 1 e para cada um desses valores de ψ obtivemos os estimadores de η_1 e η_2. Denotamos por $\hat{\psi}^{(a)}$, $\hat{\psi}^{(b)}$ e $\hat{\psi}^{(c)}$ os valores que maximizam as funções $\mathcal{L}$, $\mathcal{L}_1$ e $\mathcal{L}_2$, respectivamente. Estes valores foram obtidos diretamente dos gráficos de forma aproximada.

Na Tabela 4.2 apresentamos um resumo dos resultados obtidos para algumas das amostras simuladas. Algumas outras amostras foram geradas a partir do mesmo processo latente que gerou a amostra 8 com valores distintos de φ. A Tabela 4.3 apresenta os resultados obtidos nestas amostras. A Figura 4.4 mostra os gráficos de $\mathcal{L}_1$ em função do valor de ψ_2 utilizado no

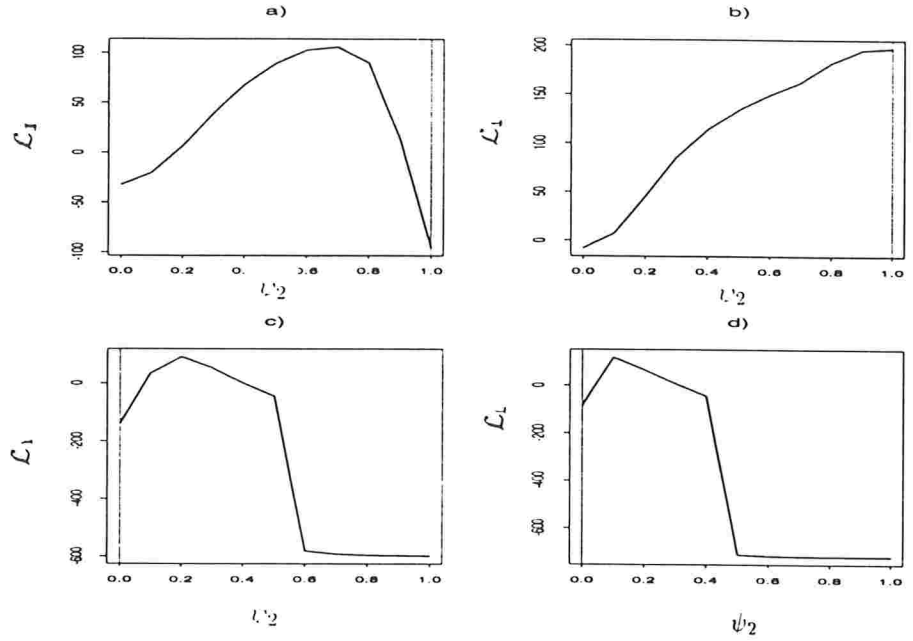

Figura 4.4: Gráficos de $\mathcal{L}_1$ para amostras simuladas com $\psi = (0.0, 1.0)^\top$ (a, b) e com $\psi = (1.0, 0.0)^\top$ (c, d). A linha vertical marca o valor de ψ_2 utilizado na simulação da amostra.

ajuste, para amostras simuladas com $\psi = (0.0, 1.0)^\top(\psi_2 = 1.0)$ e $\psi = (1.0, 0.0)^\top(\psi_2 = 0.0)$. As Figuras 4.5 e 4.6 mostram os gráficos de $\mathcal{L}_1$ em função de ψ_2 para amostras geradas com $\psi = (0.7, 0.3)(\psi_2 = 0.3)$ e $\psi = (0.3, 0.7)$ $(\psi_2 = 0.7)$, respectivamente. A linha vertical nestas três figuras marca o valor de ψ_2 utilizado na simulação.

Uma amostra Poisson foi gerada a partir do mesmo $\boldsymbol{\theta}$ e os mesmos parâmetros utilizados na amostra 8 (amostra 8_d). Para as amostras 8 e 8_d foram ajustados os modelos Poisson-gama e normal-gama. Com excessão da amostra 8, todas resultaram em estimativas na fronteira $(0.0, 1.0)^\top$.

Para obter os estimadores, construímos gráficos das funções $\mathcal{L}, \mathcal{L}_1$ e $\mathcal{L}_2$ da maneira que descrevemos abaixo para $\mathcal{L}$.

- para cada um dos onze valores de ψ, foi feito o ajuste até a convergência, estimando o parâmetro $\eta_{1,2}(\psi_j)$.

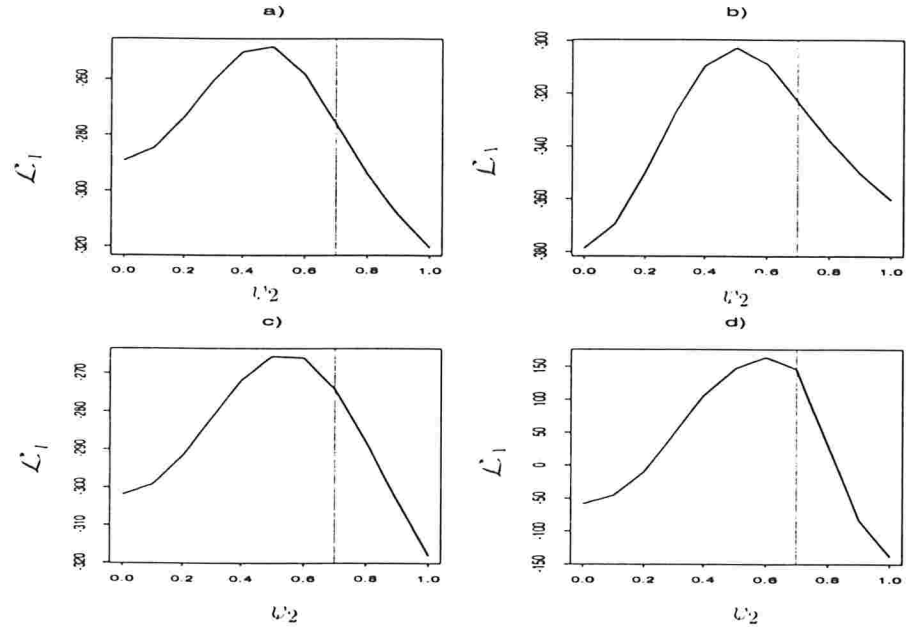

Figura 4.5: Gráficos de $\mathcal{L}_1$ em função de ψ_2 para amostras simuladas com $\psi = (0.3, 0.7)^{\top}$

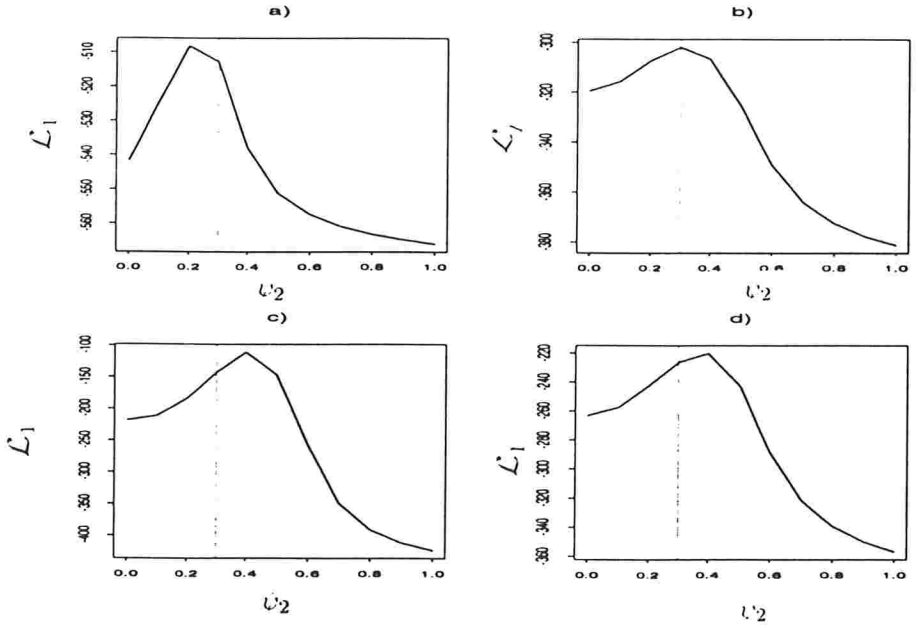

Figura 4.6: Gráficos de $\mathcal{L}_1$ em função de ψ_2 para amostras simuladas com $\psi = (0.7, 0.3)^{\top}$

Tabela 4.3: Algumas amostras simuladas, modelo normal-gama, $\nu_i^2 = 1$, $\sigma^2 = 0.01$

Amostra	parâmetros na simulação		ajuste			
	$\varphi^\top$	$\psi^\top$	$\varphi^\top$	$\hat{\psi}^{(a)\top}$	$\hat{\psi}^{(b)\top}$	$\hat{\psi}^{(c)\top}$
8	$(0.0, 1.0)$	$(0.3, 0.7)$	$(0.0, 1.0)$	$(0.4, 0.6)$	$(0.4, 0.6)$	$(0.0, 1.0)$
8^a	$(0.3, 0.7)$	$(0.3, 0.7)$	$(0.5, 0.5)$	$(0.0, 1.0)$	$(0.0, 1.0)$	$(0.0, 1.0)$
8^b	$(0.3, 0.7)$	$(0.3, 0.7)$	$(0.3, 0.7)$	$(0.5, 0.5)$	$(0.5, 0.5)$	$(0.0, 1.0)$
8^c	$(0.3, 0.7)$	$(0.3, 0.7)$	$(0.0, 1.0)$	$(0.2, 0.8)$	$(0.2, 0.8)$	$(0.0, 1.0)$

- para cada um dos ajustes, construímos o gráfico da função $\mathcal{L}(\eta_3(\psi), \mathbf{y}, \hat{\eta}_{1,2}(\psi_j))$.
- o gráfico de $\mathcal{L}$ foi obtido unindo-se os pontos $\mathcal{L}\left\{\eta_3(\psi_j), \mathbf{y}, \eta_{1,2}(\psi_j)\right\}$.

Os gráficos para cada um dos ajustes dão uma idéia de como estas funções se comportam em relação ao parâmetro de interesse. Para cada amostra simulada com o modelo normal-gama, além dos gráficos após a convergência, foram feitos gráficos com a função $\mathcal{L}$ após a iteração 1, após a iteração 5 e após a iteração 50. O resultado para a amostra 12 é mostrado na Figura C.15.

Embora as simulações descritas acima não tenham sido extensivas, alguns resultados se repetiram sistematicamente, sugerindo alguns pontos que descrevemos a seguir.

- Os gráficos obtidos nas simulações sugerem que procurar o máximo da função $\mathcal{L}$ em cada passo do algoritmo pode ser de fato bastante ineficiente. Para grande parte das amostras geradas, as funções $\mathcal{L}$ e $\mathcal{L}_2$ são monótonas atingindo o máximo em uma das fronteiras. Este comportamento é ainda mais evidente nas primeiras iterações, sugerindo que qualquer algoritmo de maximização procurando o máximo destas funções em cada iteração, vai *caminhar* na direção da fronteira de onde não sairia, uma vez que as funções para valores na fronteira também são monótonas. A função $\mathcal{L}_1\left\{\eta_3(\psi), \mathbf{y}, \hat{\eta}_{1,2}(\psi_j)\right\}$ é constante e, portanto, também não se presta a um procedimento iterativo.

- A monotonicidade das funções $\mathcal{L}$ e $\mathcal{L}_2$ observada nas simulações sugere que funções de estimação baseadas na função escore não devem fornecer bons resultados, uma vez que as derivadas em geral não se anulam no intervalo de definição dos parâmetros.

- Parece haver algum tipo de *confundimento*. Para algumas amostras (não mostradas aqui), utilizamos um procedimento análogo para estimar o parâmetro φ. Quando consideramos o processo latente como um processo de Markov de ordem 1 ($L = 1$), não conseguimos encontrar diferenças entre ajustes para valores distintos de φ. Quando $L = 2$, em todas as amostras investigadas, os gráficos de resíduos diferem bastante, com diferentes pares de valores de φ e ψ produzindo ajustes distintos.

- Nas amostras simuladas com $\nu^2 = 1$, a função $\mathcal{L}$ parece fornecer mais informação sobre o parâmetro ψ que a função $\mathcal{L}_2$, levando a estimadores mais próximos dos valores utilizados na simulação. Notemos que a função $\mathcal{L}_1$ só depende de ψ por meio dos outros parâmetros e do processo latente, uma vez que não depende funcionalmente de ψ. No caso das amostras simuladas com $\nu^2 = 10$, tanto a função $\mathcal{L}$ quanto a $\mathcal{L}_2$ só levaram a estimadores no ponto $\psi = 1$, que corresponde à fronteira do espaço paramétrico, o intervalo [0,1], mas a função $\mathcal{L}_1$ forneceu estimadores mais próximos dos valores utilizados na simulação.

- A curvatura observada no gráfico da função $\mathcal{L}$ não é reproduzida pelas curvas intermediárias para nenhuma das amostras geradas.

- Para outras amostras geradas (não mostradas aqui) foi feito um estudo semelhante para a função de estimação de Kalman para o parâmetro φ. Os gráficos sugerem que, em geral, estas funções não se anulam no intervalo [0,1].

- O melhor ajuste, no sentido de *melhor* gráfico de resíduos, nem sempre coincidiu com o ajuste para algum dos valores $\hat{\psi}^{(a)}, \hat{\psi}^{(b)}$ ou $\hat{\psi}^{(c)}$.

- As funções $\mathcal{L}$ e $\mathcal{L}_1$ crescem com as iterações mas o mesmo não acontece com a função $\mathcal{L}_2$.

- Para o caso $L = 2$, consideramos também a função de verossimilhança incluindo uma distribuição a priori para ψ, a Dirichlet. Fazendo variar os parâmetros da priori, consegue-se uma pequena curvatura na função $\mathcal{L}$, especialmente próximo às fronteiras. Este procedimento, no entanto, parece um tanto forçado uma vez que não temos, de fato, nenhuma informação a priori sobre o parâmetro. Além disso, a curvatura obtida não parece suficiente para resolver o problema, apenas elimina a fronteira das possíveis soluções.

- O comportamento da curva $\mathcal{L}$ parece só se definir na convergência (ver Figura C.15). O número de iterações necessárias para se obter a convergência variou com a amostra e com o modelo. Este processo de estimação é bastante demorado, embora viável para o caso $L = 2$. Para $n = 200$, caso destas amostras, gastou-se de 30 minutos a 1 hora e 30 minutos para ajustar os 11 valores distintos de ψ, com φ fixo.

As Figuras C.1–C.14 mostram os gráficos das funções $\mathcal{L}$ (canto superior esquerdo), $\mathcal{L}_2$ (canto superior direito), $\mathcal{L}_1$ (canto inferior esquerdo) e os valores de $\hat{\sigma}^2$, para as amostras da Tabela 4.2. O gráfico obtido com os valores de $\hat{\sigma}^2$ são, para algumas destas amostras, bastante semelhantes ao gráfico de $\mathcal{L}_1$. Gráficos semelhantes foram feitos para $\hat{\nu}^2$ e mostraram um resultado *oposto*, no sentido que valores altos de $\hat{\nu}^2$ correspondem a valores baixos de $\hat{\sigma}^2$ em um comportamento semelhante ao que ocorre com a razão das variâncias Σ/W em modelos de espaço de estados univariados tradicionais (ver West e Harrison (1997), por exemplo).

Os resultados destas simulações sugerem que apesar de $\mathcal{L}_1$ não depender funcionalmente de ψ, ela é a melhor função para se obter os estimadores deste parâmetro. Este resultado não é de todo estranho, pois $\mathcal{L}_1$ é a função que expressa a qualidade do ajuste aos dados observados.

A Figura C.5 mostra as funções de autocorrelação parcial dos resíduos obtidos com os ajustes referentes à amostra 3, utilizando valores de $\psi^\top = (0.0, 1.0), (0.6, 0.4), (0.7, 0.3)$ e $(1.0, 0.0)$. O valor $(0.6, 0.4)^\top$ corresponde ao máximo da função $\mathcal{L}_1$ e o valor $(0.7, 0.3)^\top$ corresponde ao valor de ψ utilizado na simulação. Embora estes gráficos não sejam muito

distintos entre si, existe uma evidente melhora com relação ao modelo com $\psi^\top = (0.0, 1.0)$, que corresponde ao modelo Tweedie-Tweedie.

De um modo geral, no caso das outras amostras, os resíduos obtidos com o valor de $\hat{\psi}^{(b)}$ representam alguma melhora com relação ao modelo Tweedie-Tweedie, com $\psi^\top = (0.0, 1.0)$, embora em alguns casos exista ainda uma correlação residual não explicada pelo modelo. Ver, por exemplo, a Figura C.16, correspondendo aos resíduos para a amostra 12, com valores de $\psi^\top = (0.0, 1.0), (0.8, 0.2), (0.9, 0.1)$ e (1.0,0.0), respectivamente. O valor $(0.8, 0.2)^\top$ corresponde a $\hat{\psi}^{(b)}$ e o valor utilizado na simulação desta amostra foi $(1.0, 0.0)^\top$. Neste caso, o ajuste correspondente ao valor $\psi^\top = (1.0, 0.0)$ é superior ao ajuste correspondente a $(0.8, 0.2)^\top$ que, por sua vez, é superior ao ajuste com $(0.0, 1.0)^\top$.

4.8 Outros métodos de estimação

Além do processo de estimação descrito na Seção 4.5.3, consideramos uma série de alternativas que descrevemos brevemente a seguir. Estas alternativas foram avaliadas a partir de uma série de amostras simuladas.

Funções de estimação de Kalman

Consideramos como função de estimação uma modificação das funções escore (4.16)–(4.17), considerando apenas $\mathbf{U}_{2\varphi}$ e $\mathbf{U}_{2\psi}$ e substituindo algumas das expressões em $\boldsymbol{\theta}$ por seus valores esperados e $\boldsymbol{\theta}$ pelo processo latente. As funções de estimação obtidas tiveram o vício facilmente corrigido. Na implementação desta alternativa, no entanto, não conseguimos uma forma eficiente de introduzir as restrições sobre os parâmetros no processo de estimação.

Nas simulações, obtivemos estimadores satisfazendo as restrições para amostras da distribuição normal com ν^2 pequeno. Nesses casos e para as amostras simuladas com parâmetros afastados da fronteira, os estimadores ficaram próximos dos valores verdadeiros.

Subrotina S-plus *nlminb*: minimização com restrições

Consideramos os modelos Poisson-gama e normal-gama, com $K = 1$ e $L = 1$. Para algumas amostras simuladas, utilizamos a rotina *nlminb* do S-plus, que admite restrições nos parâmetros, para obter um estimador de φ baseado na função de verossimilhança. O procedimento foi considerar apenas a verossimilhança do processo observado, substituindo o processo latente pelo suavizador de Kalman. Para quase a totalidade das amostras, geradas com diferentes valores de φ, o valor do estimador situou-se na fronteira da região, o que equivale a considerar o modelo com $K = 0$ ($\hat{\varphi} = (0, 1)^\top$).

Reparametrização

Consideramos a seguinte reparametrização,

$$\varphi_j = \frac{e^{\delta_j}}{1 + e^{\delta_1} + \cdots + e^{\delta_K}},$$

na qual os parâmetros δ_j podem variar livremente. Essa reparametrização garante que os componentes de φ situem-se no intervalo desejado, isto é, $0 < \varphi_j < 1$, $j = 0, \ldots, K$, $\sum_{j=0}^{K} \varphi_j = 1$ e equivale a considerar $\delta_j = \log(\varphi_j/\varphi_0)$, para $j = 1, \ldots, K$ e $\delta_0 = 0$. Para o modelo com $K > 0, L = 1$ e $d = 1$, tentamos então estimar δ_j a partir das funções de estimação de Kalman, modificadas para incluir as derivadas em relação aos novos parâmetros. Novamente, as estimativas situaram-se na fronteira para todas as amostras testadas. Vale notar que as derivadas em relação aos novos parâmetros se anulam nas fronteiras para os casos testados ($K = 1$). Para algumas das amostras simuladas, foram ajustados diferentes modelos com o valor de φ variando em cada um dos ajustes, mas mantendo esse valor fixo no procedimento. Embora os ajustes não tenham sido idênticos, todos os modelos foram equivalentes, considerando-se os gráficos de adequação de ajuste disponíveis.

4.9 Transferência entre tarefas sincronizatórias

Na análise de dados considerada na Seção 3.6, mostramos como a modelagem por meio de espaço de estados juntamente com as técnicas de diagnóstico apresentadas naquele capítulo, são úteis na determinação de um conjunto de covariáveis adequado para o problema em questão. Estas técnicas, disponíveis também para o modelo retrospectivo, podem e devem ser utilizadas na prática. Nesta seção, no entanto, vamos considerar um conjunto de dados reais com o objetivo de ilustrar as possibilidades de utilização do modelo retrospectivo, dando ênfase agora à modelagem da correlação serial presente nos dados. Para isso, vamos considerar um subconjunto dos dados referentes a um estudo de transferência entre tarefas sincronizatórias apresentado em Lima e Sañudo (1996) que descrevemos brevemente a seguir.

Em alguns esportes, o desempenho de um atleta depende de seu grau de coordenação e da sincronização do movimento executado com relação à chegada de um estímulo. A sincronização entre o estímulo e a ação motora decorrente deste estímulo é chamada tarefa sincronizatória. Um exemplo é a sincronização entre o salto do goleiro e o chute de um pênalti.

O objetivo do estudo relatado em Lima e Sañudo (1996), foi avaliar o processo de aprendizagem, o desempenho em dois tipos de tarefas sincronizatórias realizadas sequencialmente e a transferência de uma tarefa para outra. Foram considerados dois grupos de voluntários, cada um deles iniciando o experimento com um dos dois tipos de tarefas, que foram classificadas como tarefa simples e tarefa complexa. O experimento foi dividido em 4 etapas, sendo que nas três primeiras o indivíduo realizava uma das tarefas, transferindo para a outra tarefa na última etapa. Estas etapas foram designadas por *pré-teste* (50 repetições, sem conhecimento do resultado pelo voluntário), *prática* (400 repetições, conhecimento do resultado em tentativas alternadas), *pós-teste* (50 tentativas, sem conhecimento do resultado) e *transferência* (50 tentativas com a segunda tarefa, sem conhecimento do resultado). Durante todas estas etapas, foram registradas os tempos entre a emissão de um estímulo e a reação motora do indivíduo.

A variável resposta que vamos considerar neste exemplo é o *erro variável*, utilizado em

estudos envolvendo aprendizagem motora para medir a consistência das respostas dos indivíduos. O erro variável é definido como sendo o desvio padrão das diferenças entre o tempo de emissão do estímulo e o tempo de reação de um indivíduo nas várias tentativas. Para obter os valores desta variável neste experimento, as tentativas foram colocadas em grupos de 10, sendo que na etapa da prática foram deprezadas tanto as primeiras quanto as últimas 20 tentativas.

Vamos denotar por $d_{it}, i = 1, \ldots, 10, t = 1, \ldots, n$ as diferenças entre o tempo de recebimento do estímulo e o tempo de reação do indivíduo na i-ésima tentativa, em que o tempo t neste caso se refere à sequência de realização das tentativas. O erro variável no tempo t é dado então por

$$Y_t = \sqrt{\sum_{i=1}^{10} \frac{(d_{it} - \bar{d}_t)^2}{10}},$$

em que $\bar{d}_t = \sum_{i=1}^{10} d_{it}/10$.

Uma análise preliminar com os dados completos indicou que o modelo Tweedie-Tweedie era suficiente para explicar a correlação serial presente nas observações. Como nosso objetivo aqui é ilustrar a utilização do modelo retrospectivo, vamos considerar apenas um dos indivíduos que participaram do experimento, para o qual o modelo Tweedie-Tweedie não forneceu um bom ajuste. A Figura 4.7 mostra o gráfico dos valores observados para o indivíduo 26, que iniciou o experimento com a tarefa simples. As linhas verticais separam as quatro etapas do experimento.

Para o processo latente, vamos assumir uma distribuição gama ($q = 2$) e para o processo observado vamos considerar duas alternativas: a distribuição normal ($p = 0$) e a distribuição gama ($p = 2$). Para a componente sistemática, vamos considerar dois modelos. O primeiro, corresponde a 4 segmentos de reta, um para cada uma das etapas, todos eles com o mesmo coeficiente angular. Denotando por $D_{it}, i = 1, \ldots, 4$, as variáveis indicadoras das 4 etapas do experimento no tempo t, temos

Modelo 1

$$\log \mathrm{E}(Y_t) = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + \beta t.$$

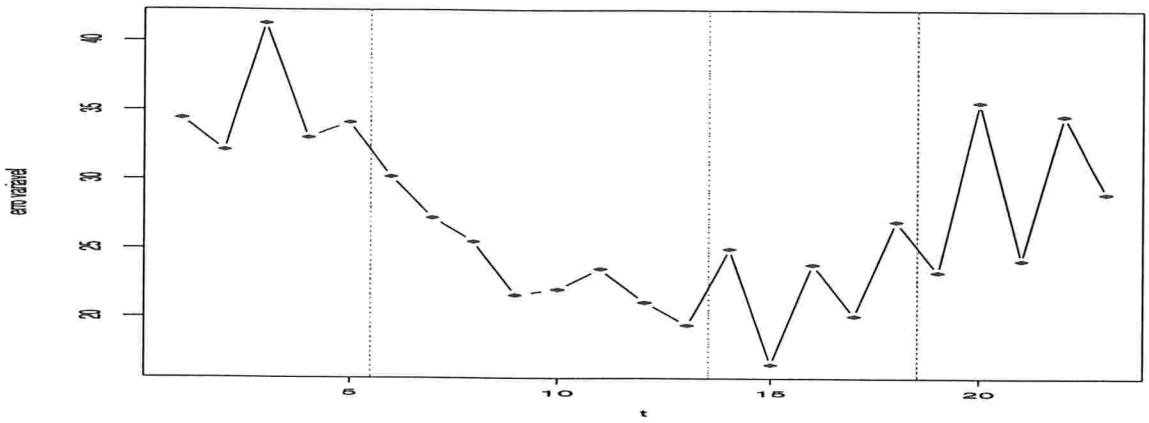

Figura 4.7: Valores observados para o indivíduo 26 (erro variável nas tentativas).

Valores iguais dos parâmetros α_i indicam, sob este modelo, não haver efeito de etapa na resposta.

O segundo modelo, que considera um segmento de reta com diferentes coeficientes para cada etapa do experimento, é dado por

Modelo 2

$$\log \mathrm{E}(Y_t) = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 D_{4t} + \beta_1 D_{1t} t + \beta_2 D_{2t} t + \beta_3 D_{3t} t + \beta_4 D_{4t} t.$$

Nos dois modelos, temos que $t = 1, \ldots, 5$, corresponde ao pré–teste, $t = 6, \ldots, 13$, à prática, $t = 14, \ldots, 18$, ao pós–teste, e $t = 19, \ldots, 23$ à transferência.

Considerando então $K = 0$ e $L = 2$, temos $\boldsymbol{\varphi} = (0.0, 1.0)^\top$ e $\boldsymbol{\psi} = (1 - \psi_2, \psi_2)$. Para estimar o parâmetro ψ_2, vamos utilizar o procedimento descrito na Seção 4.7, considerando $\psi_2 = 0.0, 0.1, \ldots, 1.0$. Fixando cado um destes valores, obtemos os estimadores dos outros parâmetros e construímos o gráfico de $\mathcal{L}_1$. A Figura 4.8 mostra os gráficos de $\mathcal{L}_1$ e de $\hat{\sigma}^2$ em função de ψ_2 para o modelo 1, considerando a distribuição normal para a variável resposta. Pelo gráfico, vemos que o valor máximo de $\mathcal{L}_1$ é obtido para $\psi_2 = 0.3$, ou seja, $\hat{\boldsymbol{\psi}}^{(b)\top} = (\hat{\psi}_1, \hat{\psi}_2)^\top = (0.7, 0.3)$. Na Figura 4.9, temos os gráficos das funções de autocorrelação

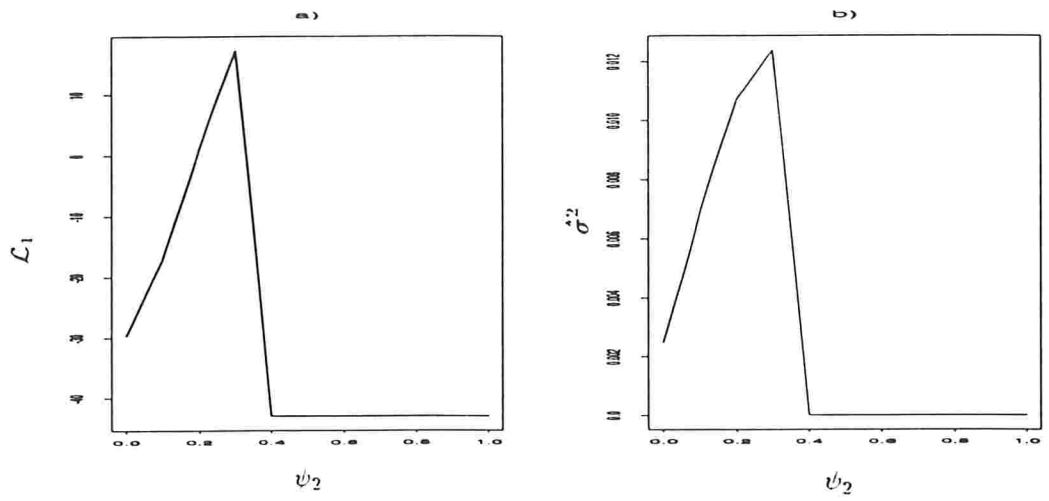

Figura 4.8: Gráficos de $\mathcal{L}_1$ (a) e $\hat{\sigma}^2$ (b) em função de ψ_2, considerando resposta normal e o modelo 1.

e autocorrelação parcial dos resíduos dos ajustes feitos com $\boldsymbol{\psi}^\top = (0.0, 1.0)$ (modelo Tweedie-Tweedie) e (0.7,0.3). Os gráficos que correspondem à $\boldsymbol{\psi}^\top = (0.7, 0.3)$, mostram um padrão de ruído branco, uma evidente melhora em relação ao modelo com $\boldsymbol{\psi}^\top = (0.0, 1.0)$, mostrados no lado esquerdo da figura. Para valores do parâmetro ψ_2 maiores que 0.3, o valor de $\hat{\sigma}^2$ foi muito pequeno e, nestes casos, o número de iterações foi sempre igual a 1000, que foi o limite superior considerado na rotina. Isto indica que, na verdade, o processo de estimação não convergiu. Investigando os processos latentes estimados para estes valores de ψ_2, notamos que eram praticamente iguais ao valor esperado estimado, o que corresponde ao modelo de independência. Devido a isso, ajustamos novamente o modelo 1, agora considerando o valor de σ^2 fixo para todos os ajustes e igual ao valor estimado anteriormente com $\boldsymbol{\psi}^\top = (0.7, 0.3)$. Com isso, forçamos um processo latente não constante.

Os resultados deste novo ajuste são mostrados nas Figuras 4.10 e 4.11. Embora a forma da curva $\mathcal{L}_1$ tenha se modificado bastante, o estimador de $\boldsymbol{\psi}$ não se alterou e as funções de autocorrelação e autocorrelação parciais dos resíduos nestes ajustes são bastante semelhantes àquelas dos ajustes anteriores (ver Figura 4.11).

Ajustamos então o modelo 2, ainda considerando a resposta com distribuição normal ($p = 0$). As Figuras 4.12 e 4.13 mostram os resultados destes ajustes. O estimador de $\boldsymbol{\psi}$

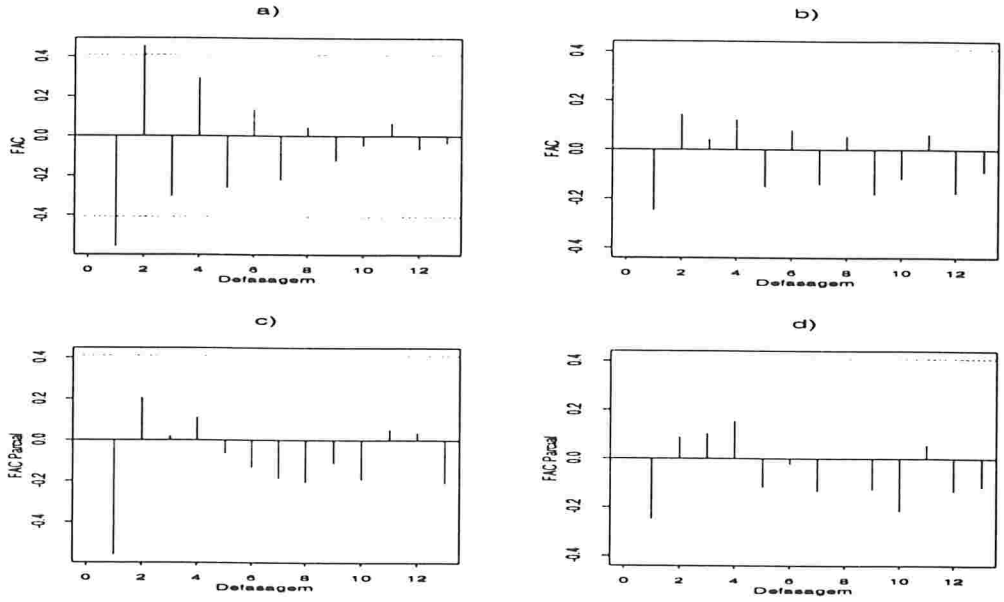

Figura 4.9: Funções de autocorrelação e autocorrelação parcial dos resíduos obtidos com ajustes do modelo 1 para $\psi_2 = 1.0$ (a, c) e 0.3 (b, d), considerando resposta normal.

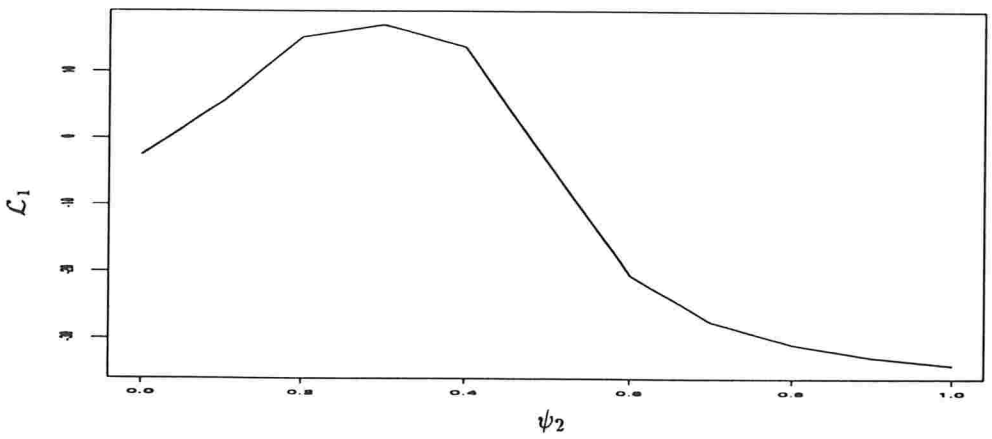

Figura 4.10: Gráficos de $\mathcal{L}_1$ em função de ψ_2, considerando resposta normal e o modelo 1 com $\sigma^2 = 0.0124$ em todos os ajustes.

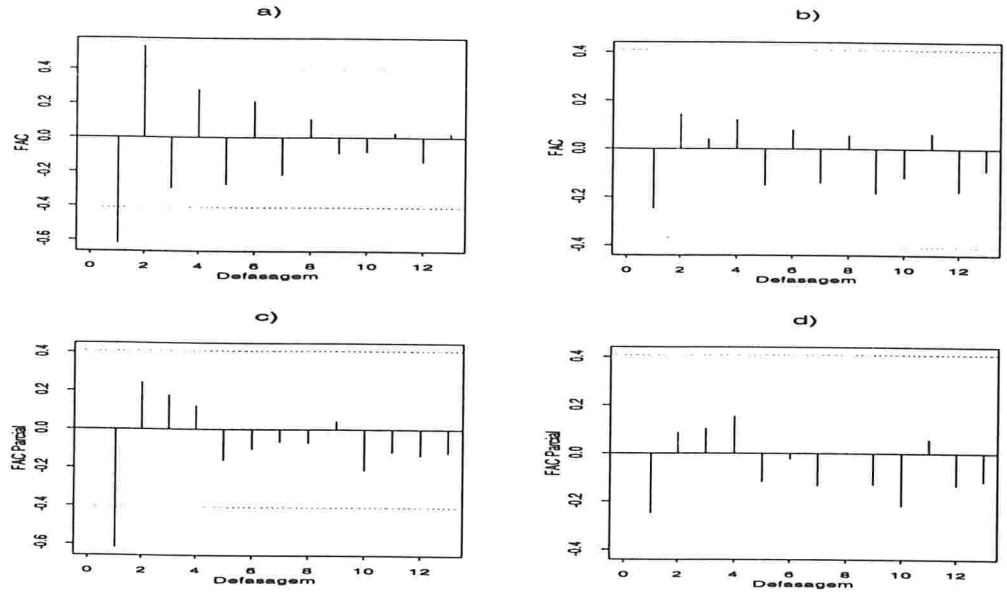

Figura 4.11: Funções de autocorrelação e autocorrelação parcial dos resíduos obtidos com ajustes do modelo 1 com $\sigma^2 = 0.0124$ fixo, para $\psi_2 = 1.0$ (lado esquerdo), 0.3 (lado direito), considerando resposta normal.

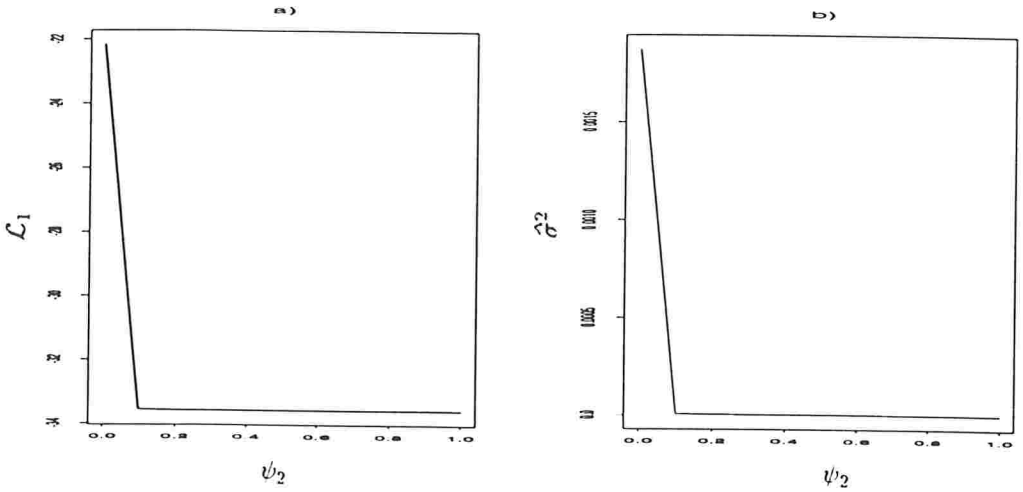

Figura 4.12: Gráficos de $\mathcal{L}_1$ (a) e de $\hat{\sigma}^2$ (b) em função de ψ_2, considerando resposta normal e o modelo 2.

está na fronteira $(1.0, 0.0)(\psi_2 = 0.0)$ e a função de autocorrelação e autocorrelação parcial dos resíduos obtidos com o ajuste feito com este valor apresentam padrão de ruído branco. Neste caso, ocorreu novamente o comportamento observado no caso anterior, com σ^2 indo para zero para valores de ψ_2 menores que 1.

Como o ajuste para o modelo 2 foi melhor que o ajuste do modelo 1 (ver Figura 4.14 a) e b)), ajustamos novamente o modelo 2 considerando agora uma distribuição gama para as observações ($p = 2$). A Figura 4.15 mostra os gráficos de $\mathcal{L}_1$ e $\hat{\sigma}^2$ para este modelo e a Figura 4.16 mostra gráficos das funções de autocorrelação dos resíduos para alguns valores de ψ_2. O valor que maximiza a função $\mathcal{L}_1$ agora é $\hat{\psi}_2 = 0.3$ mas as funções de autocorrelação e de autocorrelação parcial dos resíduos neste caso ainda não apresentam padrão de ruído branco, o que acontece, no entanto, para $\psi_2 = 0.2$.

A análise poderia prosseguir, procurando um modelo mais adequado para a componente sistemática, por exemplo, mas os resultados referentes a um único indivíduo neste caso não são de interesse prático. Os resultados desta seção, no entanto, mostram que o modelo retrospectivo pode ser de fato utilizado em dados reais, uma vez que se conseguiu, para uma mesma componente sistemática, um ajuste melhor com o modelo retrospectivo do que o obtido com o modelo Tweedie-Tweedie ($\boldsymbol{\psi}^\top = (0.0, 1.0)$).

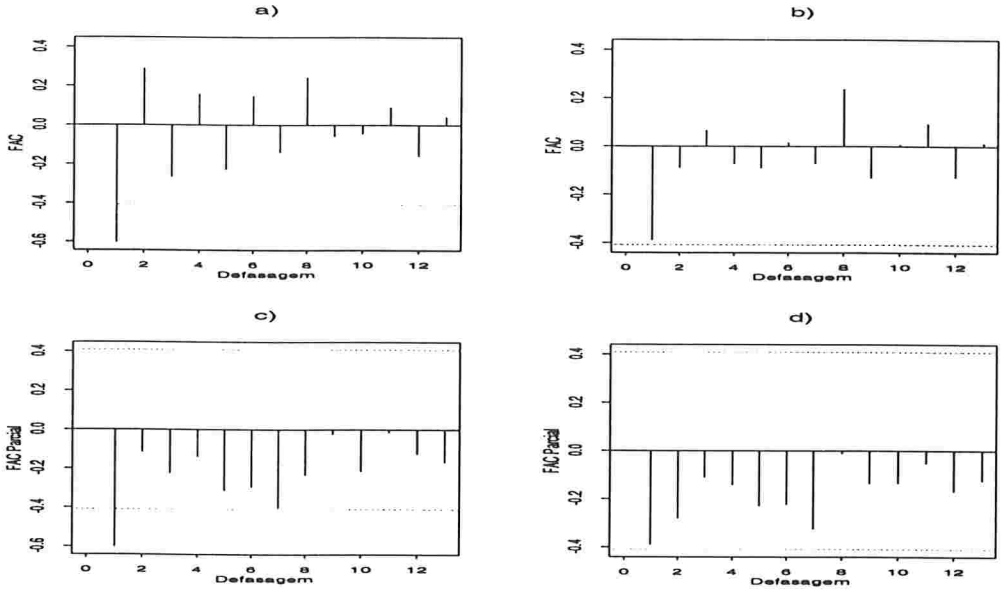

Figura 4.13: Funções de autocorrelação e autocorrelação parcial dos resíduos obtidos com ajustes do modelo 2 para $\psi_2 = 1.0$ (a, c) e 0.3 (b, d) e resposta normal.

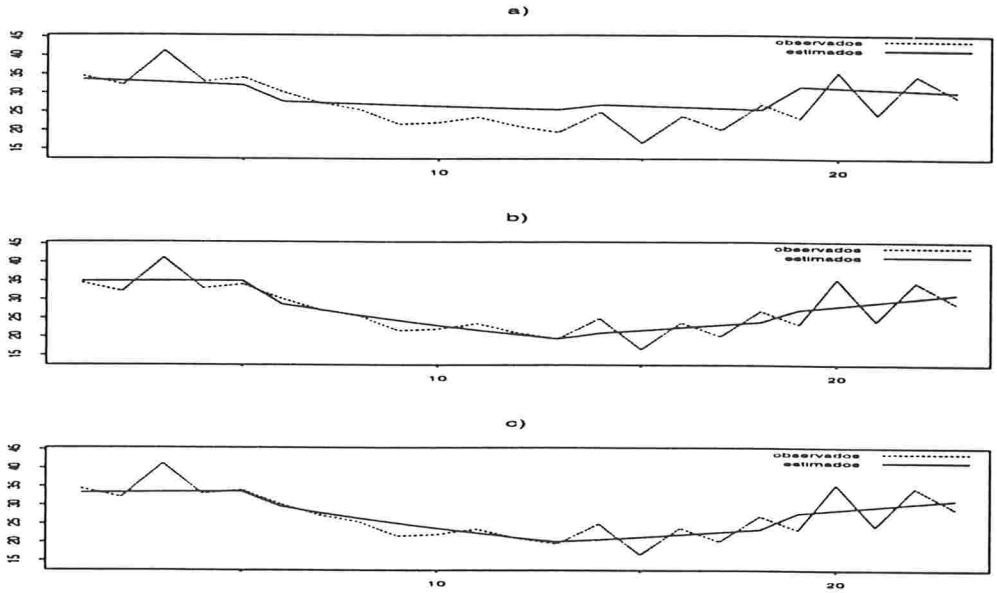

Figura 4.14: Valores observados e estimados com ajustes do modelo 1 com resposta normal (a), do modelo 2 com resposta normal (b) e do modelo 2 com resposta gama (c). Os valores observados são apresentados em linha pontilhada e os estimados em linha cheia.

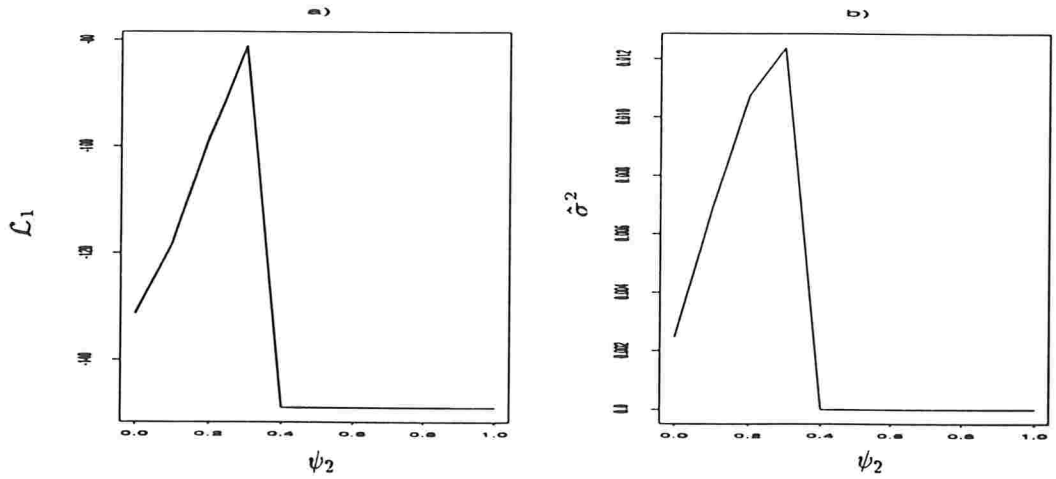

Figura 4.15: Gráficos de $\mathcal{L}_1$ e $\hat{\sigma}^2$ em função de ψ_2, considerando resposta gama e o modelo 2.

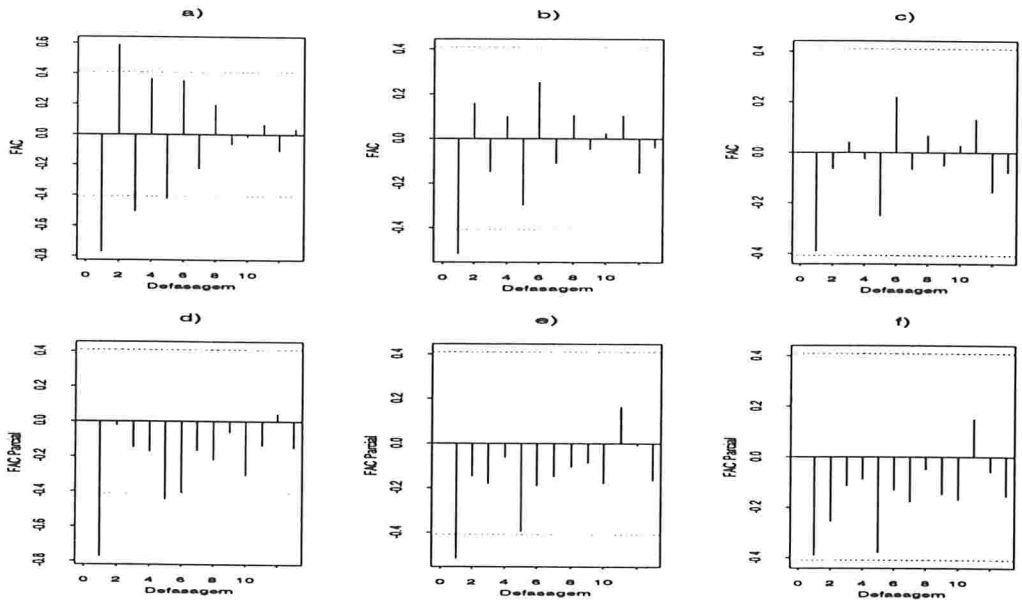

Figura 4.16: Funções de autocorrelação (no alto) e autocorrelação parcial (embaixo) dos resíduos obtidos com ajustes do modelo 2 para $\psi_2 = 1.0$ (a, d), $\psi_2 = 0.3$ (b, e) e $\psi_2 = 0.2$ (c, f), resposta gama.

Capítulo 5

Conclusões e estudos futuros

A utilização do modelo retrospectivo em análise de dados reais deve levar em conta, naturalmente, as características do modelo e sua adequação ao conjunto de dados em questão. No entanto, um dos aspectos do modelo retrospectivo que consideramos relevante, é justamente a possibilidade que ele oferece de investigar vários aspectos e características dos dados, tanto no que diz respeito à escolha de uma componente sistemática adequada, quanto no que diz respeito à estrutura de correlação considerada. A utilização das técnicas de diagnóstico que descrevemos aqui em conjunto com o processo latente, se mostraram meios eficientes na determinação de um modelo adequado, podendo funcionar, inclusive, como um interessante procedimento de análise exploratória de dados. Outro ponto importante é a simplicidade de interpretação dos parâmetros, uma vez que uma análise de dados utilizando o modelo retrospectivo é conduzida de forma semelhante à de uma análise de regressão tradicional.

Alguns aspectos do modelo retrospectivo merecem uma investigação mais profunda. Por exemplo,

- a proposta que fizemos de estimação dos parâmetros retrospectivos não foi investigada para modelos com ordem maior que 2. Para investigar modelos de ordem mais alta, será necessário primeiramente melhorar o procedimento de estimação, uma vez que, neste caso, o procedimento gráfico se tornará ineficiente;

- estudos de simulação mais exaustivos podem ser conduzidos com o objetivo de investigar propriedades dos estimadores propostos;

- estudo semelhante ao conduzido aqui pode ser feito para outras distribuições da família Tweedie;

- pode-se investigar o desempenho da aproximação ponto de sela (ver Jørgensen (1997)) no procedimento de estimação dos parâmetros retrospectivos; uma vantagem seria a possibilidade de tratamento unificado de todos os membros da família Tweedie, o que permitiria o desenvolvimento de rotinas mais simples.

- extensões do modelo retrospectivos poderiam ser estudadas de forma a incluir, por exemplo, funções de ligação diferentes da função logaritmo; a utilização das técnicas de estimação descritas aqui ficariam condicionadas, no entanto, à preservação de algumas propriedades do modelo, principalmente no que diz respeito aos seus momentos, devido ao uso do filtro e suavizador de Kalman que desenvolvemos no Capítulo 2.

Apêndice A

Família Tweedie

Neste apêndice, apresentamos um breve resumo das definições e propriedades básicas da família Tweedie dos modelos de dispersão. Para detalhes, ver Jørgensen (1997).

Definição A.1 *Sejam Ω e C intervalos tais que $\Omega \subseteq C \subseteq \mathbb{R}$, sendo Ω aberto. Uma relação $d : C \times \Omega \to \mathbb{R}$ é denominada desvio unitário se satisfaz*

$$d(y; y) = 0 \quad \forall\, y \in \Omega,$$

$$d(y; \mu) > 0 \quad \forall\, y \neq \mu.$$

Um desvio unitário d é dito *regular* se é duas vezes continuamente diferenciável com respeito a (y, μ) em $\Omega \times \Omega$ e satisfaz

$$\frac{\partial^2}{\partial \mu^2} d(\mu; \mu) > 0, \quad \forall\, \mu \in \Omega.$$

Definição A.2 *Uma função de variância unitária, $V : \Omega \to \mathbb{R}_+$ de um desvio unitário regular é definida por*

$$V(\mu) = \frac{2}{\dfrac{\partial^2}{\partial \mu^2} d(\mu; \mu)}. \tag{A.1}$$

Definição A.3 *Uma família de distribuições é um modelo de dispersão reprodutivo,* $\mathrm{Y} \sim \mathrm{MD}(\mu, \sigma^2)$ *com parâmetro de posição* μ *e parâmetro de dispersão* σ^2 *se as densidades de probabilidade dos elementos dessa família puderem ser escritas na forma (padrão):*

$$p(y; \mu, \sigma^2) \;=\; a(y; \sigma^2) exp\left\{-\frac{1}{2\sigma^2} d(y; \mu)\right\}, \quad y \in C,$$

em que $a \geq 0$, d *é um desvio unitário em* $C \times \Omega$, $\mu \in \Omega$ *e* $\sigma^2 > 0$.

C é o *suporte convexo* da família, isto é, o menor intervalo contendo S, onde S é a união de todos os suportes da família. A média e a variância de Y são dadas por μ e $\sigma^2 \mathrm{V}(\mu)$, respectivamente, onde V é a função de variância unitária como em (A.1). Se $d(y; \mu) = yf(\mu) + g(\mu) + h(y)$, dizemos que a família de distribuições é um *modelo de dispersão exponencial reprodutivo*, denotado por $Y \sim \mathrm{ED}(\mu, \sigma^2)$.

Definição A.4 *Se* $Y \sim \mathrm{ED}(\mu, \sigma^2)$ *é um modelo exponencial de dispersão reprodutivo, a família obtida pela transformação* $Z = Y/\sigma^2$ *(transformação dual) é um modelo exponencial de dispersão aditivo, denotado por* $Z \sim \mathrm{ED}^*(\theta, \lambda)$. *A função de densidade de probabilidade de* Z *tem a seguinte forma,*

$$p(y; \theta, \lambda) \;=\; c^*(y; \lambda) \exp\left[\theta y - \lambda \kappa(\theta)\right], \; y \in \mathbb{R},$$

em que $\lambda = 1/\sigma^2$ *e* κ *é denominada função cumulante. O parâmetro* θ *definido por* $\mathcal{T}(\theta) = \kappa'(\theta)$ *é denominado parâmetro canônico.*

Um modelo exponencial de dispersão aditivo satisfaz as seguintes propriedades,

- A média e a variância de Z são

$$\mathrm{E}(Z) = \mu = \lambda \mathcal{T}(\theta) \quad \text{e } \mathrm{Var}(Z) = \lambda V(\mu/\lambda),$$

  respectivamente, em que $V(\mu) = \kappa''(\tau^{-1}(\mu))$ é a função de variância unitária em $\Omega = \mathcal{T}(\mathrm{int}(\Theta))$.

- (Fórmula da convolução)

$$\mathrm{ED}^*(\theta, \lambda_1) * \mathrm{ED}^*(\theta, \lambda_2) = \mathrm{ED}^*(\theta, \lambda_1 + \lambda_2).$$

Definição A.5 *A família Tweedie, $Tw_p(\mu, \sigma^2)$, é a classe de modelos exponenciais de dispersão reprodutivos cuja função de variância unitária é dada por*

$$V(\mu) = \mu^p, \quad \text{para} \quad p \in (-\infty, 0] \cup [1, \infty).$$

A função de densidade de probabilidade da família Tweedie tem a seguinte forma,

$$f(y; \mu, \sigma^2, p) = c_p(y; \sigma^2) \exp\left[\frac{1}{\sigma^2}\left\{y\mathcal{T}_p^{-1}(\mu) - \kappa_p\left(\mathcal{T}_p^{-1}(\mu)\right)\right\}\right],$$

em que $\kappa_p(\theta) = \alpha^{-1}(\alpha - 1)\{\theta/(\alpha - 1)\}^\alpha$, $\mathcal{T}_p(\theta) = \{\theta/(\alpha - 1)\}^{\alpha-1}$ e $\alpha = (p - 2)/(p - 1)$. Tem-se que $\mathrm{Var}(Y) = \sigma^2\mu^p$. A família Tweedie pode ser caracterizada como sendo o único modelo exponencial de dispersão que satisfaz a seguinte propriedade,

$$c\mathrm{Tw}_p(\mu, \sigma^2) \sim \mathrm{Tw}_p(c\mu, c^{2-p}\sigma^2) \quad \forall\, c > 0.$$

Devido a isso, tem-se que a família Tweedie tem uma forma aditiva obtida pela transformação dual.

Para $\lambda = 1/\sigma^2$, tem-se $\mathrm{Tw}_p^*(\theta, \lambda) = \mathrm{Tw}_p\{\lambda\mathcal{T}_p(\theta), \lambda^{1-p}\}$. Para $p \leq 0$ e $p > 2$, as distribuições da família Tweedie correspondem a *distribuições estáveis extremas*. Uma distribuição F é dita *estável* se para $X_1, X_2, \ldots, X_n$ independentes e identicamente distribuídas com distribuição F, existem constantes b_n e c_n tais que $b_n + c_n(X_1 + \cdots + X_n) \sim F$. Para a família Tweedie, as constantes c_n são dadas por $c_n = n^{-1/\alpha}$ para $0 < \alpha \leq 2$ (α é chamado *índice* de F). A classe de distribuições estáveis tem também outro parâmetro, β, com domínio $[-1, 1]$. Uma distribuição estável é dita *extrema* se $\beta = +1$ ou -1.

Na Tabela A.1, apresentamos algumas distribuições da família Tweedie.

Tabela A.1: Família Tweedie

Distribuição	p	S	Ω	Θ
Estável extrema	$p < 0$	$\mathbb{R}$	$\mathbb{R}_+$	$\mathbb{R}_0$
Normal	$p = 0$	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$
(Não existe)	$0 < p < 1$	-	-	-
Poisson	$p = 1$	$\mathbf{N}_0$	$\mathbb{R}_+$	$\mathbb{R}$
Poisson Composta	$1 < p < 2$	$\mathbb{R}_0$	$\mathbb{R}_+$	$\mathbb{R}_-$
Gama	$p = 2$	$\mathbb{R}_+$	$\mathbb{R}_+$	$\mathbb{R}_-$
Estável positiva	$2 < p < 3$	$\mathbb{R}_+$	$\mathbb{R}_+$	$-\mathbb{R}_0$
Normal Inversa	$p = 3$	$\mathbb{R}_+$	$\mathbb{R}_+$	$-\mathbb{R}_0$
Estável positiva	$p > 3$	$\mathbb{R}_+$	$\mathbb{R}_+$	$-\mathbb{R}_0$

[a] Notação: S= Suporte, $\mathbb{R}_0 = [0, \infty)$, $\mathbf{N}_0 = \mathbf{N} \cup \{0\}$.

Apêndice B

Cálculos

Neste apêndice, apresentamos algumas propriedades utilizados nos capítulos anteriores e detalhes de alguns resultados do Capítulo 4.

B.1 Esperança condicional

Na obtenção das propriedades dos preditores lineares e dos momentos do modelo retrospectivo, utilizamos uma série de propriedades conhecidas da esperança e da variância condicionais, que listamos a seguir. A notação $\mathrm{E}[\mathbf{X}|\mathbf{Y}]$ significa $\mathrm{E}[\mathbf{X}|\boldsymbol{\sigma}(\mathbf{Y})]$, em que $\boldsymbol{\sigma}(\mathbf{Y})$ denota a σ-álgebra gerada por $\mathbf{Y}$.

Propriedade B.1 *Sejam* $\mathbf{X}$, $\mathbf{Y}$ *e* $\mathbf{Z}$ *vetores aleatórios definidos no mesmo espaço de probabilidades. Então,*

$$\mathrm{E}[\mathbf{X}] = \mathrm{E}\left[\mathrm{E}[\mathbf{X}|\mathbf{Y}]\right]$$

e

$$\mathrm{Cov}(\mathbf{X},\mathbf{Y}) = \mathrm{E}\left[\mathrm{Cov}(\mathbf{X},\mathbf{Y}|\mathbf{Z})\right] + \mathrm{Cov}\left(\mathrm{E}[\mathbf{X}|\mathbf{Z}], \mathrm{E}[\mathbf{Y}|\mathbf{Z}]\right).$$

Em particular, $\mathrm{Var}(\mathbf{X}) = \mathrm{E}[\mathrm{Var}(\mathbf{X}|\mathbf{Y})] + \mathrm{Var}(\mathrm{E}[\mathbf{X}|\mathbf{Y}])$ *e* $\mathrm{Cov}[\mathbf{X},\mathbf{Y}] = \mathrm{Cov}\left[\mathbf{X}, \mathrm{E}[\mathbf{Y}|\mathbf{X}]\right]$.

Propriedade B.2 *Sejam $\mathcal{F}_1 \subset \mathcal{F}_2$ duas σ-álgebras. Então,*

$$\mathrm{E}\left[\mathrm{E}[\mathbf{X}|\mathcal{F}_1]|\mathcal{F}_2\right] = \mathrm{E}\left[\mathrm{E}[\mathbf{X}|\mathcal{F}_2]|\mathcal{F}_1\right] = \mathrm{E}[\mathbf{X}|\mathcal{F}_1].$$

Propriedade B.3 *Seja f uma função mensurável. Então,*

$$\mathrm{E}\left[f(\mathbf{X},\mathbf{Y})|\mathbf{X}\right] = f\left(\mathbf{X}, \mathrm{E}[\mathbf{Y}|\mathbf{X}]\right).$$

B.2 Resíduos

Nesta seção, apresentamos detalhes dos resultados sobre os resíduos da Seção 3.5.

Vamos definir os seguintes resíduos.

- utilizando $\mathbf{m}^+$: $\hat{\varepsilon}_t^+ = \mathbf{Y}_t - \mathbf{F}_t\mathbf{m}_t^+$, $\hat{\boldsymbol{\xi}}_t^+ = \mathbf{m}_t^+ - \mathbf{G}_t\mathbf{m}_{t-1}^+$,
- utilizando $\mathbf{m}$: $\hat{\varepsilon}_t = \mathbf{Y}_t - \mathbf{F}_t\mathbf{m}_t$, $\hat{\boldsymbol{\xi}}_t = \mathbf{m}_t - \mathbf{G}_t\mathbf{m}_{t-1}$, (filtro)
- utilizando $\mathbf{m}^*$: $\hat{\varepsilon}_t^* = \mathbf{Y}_t - \mathbf{F}_t\mathbf{m}_t^*$, $\hat{\boldsymbol{\xi}}_t^* = \mathbf{m}_t^* - \mathbf{G}_t\mathbf{m}_{t-1}^*$, (suavizador)

em que $\mathbf{m}_t^+ = \mathbf{G}_t\mathbf{m}_{t-1}$ (ver Seção 4.4). Como qualquer preditor linear é não viciado, todos os resíduos acima têm esperança $\mathbf{0}$. Como podemos escrever $\hat{\varepsilon}_t^+ = \mathbf{Y}_t - \mathbf{f}_t$, obtemos que $\mathrm{Var}(\hat{\varepsilon}_t^+) = \mathbf{Q}_t$. Das propriedades de preditores lineares, temos que $\mathbf{Y}_{t+s} - \mathbf{f}_{t+s} \perp \mathbf{Y}^{t+s-1}$ para $s > 0$ e portanto, $\mathbf{Y}_{t+s} - \mathbf{f}_{t+s} \perp \mathbf{Y}^t$. Como $\mathbf{f}_t \in \mathcal{V}(\mathbf{Y}^{t-1})$, em que $\mathcal{V}(\mathbf{Y}^{t-1})$ denota o espaço linear gerado por $\mathbf{Y}^{t-1}$, obtemos $\mathrm{Cov}(\hat{\varepsilon}_t^+, \hat{\varepsilon}_{t+s}^+) = \mathrm{Cov}(\mathbf{Y}_t - \mathbf{f}_t, \mathbf{Y}_{t+s} - \mathbf{f}_{t+s}) = \mathbf{0}$.

Para obter as covariâncias dos resíduos para o processo latente, notamos primeiramente que os resíduos construídos a partir do filtro podem ser expressos como combinações lineares de $\hat{\varepsilon}^+$. De (2.12), podemos escrever $\hat{\boldsymbol{\xi}}_t = \mathbf{B}_t\mathbf{F}_t^\top\mathbf{Q}_t^{-1}\hat{\varepsilon}_t^+$. Logo,

$$\mathrm{Var}(\hat{\boldsymbol{\xi}}_t) = \mathbf{B}_t\mathbf{F}_t^\top\mathbf{Q}_t^{-1}\mathrm{Var}(\hat{\varepsilon}_t^+)\mathbf{Q}_t^{-1}\mathbf{F}_t\mathbf{B}_t^\top = \mathbf{B}_t\mathbf{F}_t^\top\mathbf{Q}_t^{-1}\mathbf{F}_t\mathbf{B}_t^\top.$$

Além disso, $\mathrm{Cov}(\hat{\boldsymbol{\xi}}_t, \hat{\boldsymbol{\xi}}_{t+s}) = \mathbf{B}_t\mathbf{F}_t\mathbf{Q}_t^{-1}\mathrm{Cov}(\hat{\varepsilon}_t^+, \hat{\varepsilon}_{t+s}^+)\mathbf{Q}_{t+s}^+\mathbf{F}_{t+s}^\top\mathbf{B}_{t+s}^\top = \mathbf{0}$. Agora, $\hat{\boldsymbol{\xi}}_t^+ = \mathbf{G}_t\mathbf{m}_{t-1} - \mathbf{G}_t\mathbf{G}_{t-1}\mathbf{m}_{t-2} = \mathbf{G}_t\hat{\boldsymbol{\xi}}_{t-1}$, de modo que $\mathrm{Var}(\hat{\boldsymbol{\xi}}_t^+) = \mathbf{G}_t\mathrm{Var}(\hat{\boldsymbol{\xi}}_{t-1})\mathbf{G}_t^\top = \mathbf{G}_t\mathbf{B}_{t-1}\mathbf{F}_{t-1}\mathbf{Q}_{t-1}^{-1}\mathbf{F}_{t-1}^\top\mathbf{B}_{t-1}^\top\mathbf{G}_t^\top$. Estes resíduos são não correlacionados, uma vez que $\mathrm{Cov}(\hat{\boldsymbol{\xi}}_t^+, \hat{\boldsymbol{\xi}}_{t+s}^+) = \mathbf{G}_t\mathrm{Cov}(\hat{\boldsymbol{\xi}}_{t-1}, \hat{\boldsymbol{\xi}}_{t+s-1})\mathbf{G}_{t+s}^\top = \mathbf{0}$. Escrevendo $\hat{\boldsymbol{\xi}}_t^+$ em termos de $\hat{\varepsilon}_{t-1}^+$, obtemos que $\mathrm{Cov}(\hat{\boldsymbol{\xi}}_t^+, \hat{\varepsilon}_{t+s}^+) = \mathbf{0}$ para $s \neq 1$. Para $|t-s| = 1$, temos $\mathrm{Cov}(\hat{\varepsilon}_t^+, \hat{\boldsymbol{\xi}}_s^+) = \mathbf{F}_t\mathbf{B}_t\mathbf{G}_s^\top$. Para obter as covariâncias de $\hat{\varepsilon}_t$, utilizamos as

Proposições 2.1 e (4.9), que nos dão, para $s \geq 1$, $\mathrm{Cov}(\mathbf{Y}_t, \mathbf{m}_{t+s}) = \mathbf{F}_t \mathrm{Cov}(\Theta_t, \Theta_{t+s})$. Como $\hat{\varepsilon}^+_{t+s} \perp \mathbf{m}_t \in \mathcal{V}(\mathbf{Y}^{t+s})$, obtemos $\mathrm{Cov}(\mathbf{m}_t, \mathbf{Y}_{t+s}) = \mathrm{Cov}(\mathbf{m}_t, \mathbf{f}_{t+s}) = \mathrm{Cov}(\mathbf{m}_t, \mathbf{m}_{t+s-1})\mathbf{G}^\top_{t+s}\mathbf{F}^\top_{t+s}$. Então, como $\mathrm{Cov}(\mathbf{m}_t, \mathbf{m}_{t+s}) = \mathrm{Cov}(\mathbf{m}_t, \mathbf{m}_{t+s-1})\mathbf{G}^\top_{t+s}$, obtemos, finalmente que

$$\begin{aligned}\mathrm{Cov}(\hat{\varepsilon}_t, \hat{\varepsilon}_{t+s}) &= \mathbf{F}_t \mathrm{Cov}(\Theta_t, \Theta_{t+s})\mathbf{F}^\top_{t+s} - \mathbf{F}_t \mathrm{Cov}(\mathbf{m}_t, \mathbf{Y}_{t+s}) - \\ &\quad \mathrm{Cov}(\mathbf{Y}_t, \mathbf{m}_{t+s})\mathbf{F}^\top_{t+s} + \mathbf{F}_t \mathrm{Cov}(\mathbf{m}_t, \mathbf{m}_{t+s})\mathbf{F}^\top_{t+s} = \mathbf{0}.\end{aligned}$$

Agora, é fácil ver que $\mathrm{Cov}(\hat{\boldsymbol{\xi}}_t, \hat{\varepsilon}_s) = \mathbf{0}$, se $s \neq t$ e $\mathrm{Cov}(\hat{\boldsymbol{\xi}}_t, \hat{\varepsilon}_t) = \mathbf{C}_t\mathbf{F}^\top_t$. Cálculos diretos nos dão então $\mathrm{Var}(\hat{\varepsilon}_t) = \boldsymbol{\Lambda}_t - \mathbf{F}_t\mathbf{C}_t\mathbf{F}^\top_t$.

Para obter a estrutura de covariância dos resíduos baseados no suavizador de Kalman, usaremos $\mathrm{Var}(\mathbf{m}^*) = \mathrm{Var}(\boldsymbol{\theta}) - \mathbf{C}^*$ e $\mathrm{Cov}(\mathbf{m}^*_t, \mathbf{Y}_t) = \mathrm{Cov}(\Theta_t, \mathbf{Y}_t) = \mathrm{Var}(\Theta_t)\mathbf{F}^\top_t$. Estes dois fatos, juntamente com as propriedades dos preditores lineares, nos dão $\mathrm{Var}(\hat{\varepsilon}^*_t) = \boldsymbol{\Lambda}_t - \mathbf{F}_t\mathbf{C}^*_t\mathbf{F}^\top_t$ e $\mathrm{Cov}\left(\hat{\varepsilon}^*_t, \hat{\varepsilon}^*_{t+s}\right) = -\mathbf{F}_t\mathbf{C}^*_{t,t+s}\mathbf{F}^\top_{t+s}$. Para o processo latente, temos

$$\mathrm{Var}\left(\hat{\boldsymbol{\xi}}^*_t\right) = \boldsymbol{\lambda}_t - \mathbf{G}_t\mathbf{C}^*_{t-1}\mathbf{G}^\top_t + \mathbf{C}^*_{t,t-1}\mathbf{G}^\top_t + \mathbf{G}_t\mathbf{C}^*_{t-1,t} - \mathbf{C}^*_t$$

e

$$\mathrm{Cov}\left(\hat{\boldsymbol{\xi}}^*_t, \hat{\boldsymbol{\xi}}^*_{t+s}\right) = -\mathbf{C}^*_{t,t+s} + \mathbf{G}_t\mathbf{C}^*_{t-1,t+s} + \mathbf{C}^*_{t,t+s-1}\mathbf{G}^\top_{t+s} - \mathbf{G}_t\mathbf{C}^*_{t-1,t+s-1}\mathbf{G}^\top_{t+s}.$$

Temos também que

$$\mathrm{Cov}\left(\hat{\boldsymbol{\xi}}^*_t, \hat{\varepsilon}^*_s\right) = \mathbf{C}^*_{t,s}\mathbf{F}^\top_s - \mathbf{G}_t\mathbf{C}^*_{t-1,s}\mathbf{F}^\top_s, \ s \neq t.$$

$$\mathrm{Cov}(\hat{\boldsymbol{\xi}}^*_t, \hat{\boldsymbol{\xi}}^*_s) = -\mathbf{C}^*_{t,s} + \mathbf{C}^*_{t,s-1}\mathbf{G}^\top_s - \mathbf{G}_t\mathbf{C}^*_{t-1,s-1}\mathbf{G}^\top_s + \mathbf{G}_t\mathbf{C}^*_{t-1,s}. \tag{B.1}$$

B.3 Derivadas do filtro e suavizador de Kalman

Para qualquer preditor linear $\mathbf{m}_{X|Y}$, o valor esperado de sua derivada com respeito a um dado parâmetro, digamos α, é dado por

$$-\Sigma_{XY}\Sigma_Y^{-1}\frac{\partial}{\partial\alpha}\boldsymbol{\mu}_Y,$$

uma vez que $\mathrm{E}(\mathbf{Y} - \boldsymbol{\mu}_Y) = \mathbf{0}$. Logo, o valor esperado das derivadas de qualquer preditor linear de $\boldsymbol{\theta}$ dado um subvetor de $\mathbf{Y}$, com respeito a algum componente de $(\boldsymbol{\eta}_2^\top, \boldsymbol{\eta}_3^\top)^\top$ será

nulo, pois o valor esperado marginal de $\mathbf{Y}$ não depende destes parâmetros. Desta forma, basta calcular as derivadas do suavizador de Kalman com respeito a $\boldsymbol{\eta}_1$.

Todos os valores esperados de interesse podem ser obtidas recursivamente. Os valores esperados das derivadas do suavizador podem ser obtidos por recursão, iniciada com o valor esperado da derivada do filtro de Kalman no tempo n, que calculamos abaixo. Com o objetivo de simplificar a notação, omitimos o índice j no que segue.

Seja $\boldsymbol{\mathcal{D}}_{\alpha_{ik}}(t)$ o valor esperado da derivada do filtro de Kalman em relação a α_{ik}, no tempo t. Então, como $\mathbf{m}_t = \mathbf{G}_t\mathbf{m}_{t-1} + \mathbf{B}_t\mathbf{F}_t^\top\mathbf{Q}_t^{-1}(\mathbf{Y}_t - \mathbf{f}_t)$ e $\mathrm{E}(\mathbf{Y}_t - \mathbf{f}_t) = \mathbf{0}$, obtemos

$$\boldsymbol{\mathcal{D}}_{\alpha_{ik}}(t) = (\mathbf{G}_t - \mathbf{B}_t\mathbf{F}_t^\top\mathbf{Q}_t^{-1}\mathbf{F}_t\mathbf{G}_t)\boldsymbol{\mathcal{D}}_{\alpha_{ik}}(t-1) - \mathbf{B}_t\mathbf{F}_t^\top\mathbf{Q}_t^{-1}\frac{\partial\mathbf{F}_t}{\partial\alpha_{ik}}\mathbf{G}_t\boldsymbol{\tau}_{t-1},$$

para $i = 1, \ldots, d$ e $k = 1, \ldots, l_1$. Usando uma notação análoga para os valores esperados das derivadas do filtro com respeito aos outros parâmetros de regressão, obtemos

$$\boldsymbol{\mathcal{D}}_{\beta_k}(t) = \left(\mathbf{G}_t - \mathbf{B}_t\mathbf{F}_t^\top\mathbf{Q}_t^{-1}\mathbf{F}_t\mathbf{G}_t\right)\boldsymbol{\mathcal{D}}_{\beta_k}(t-1) + \left\{\frac{\partial\mathbf{G}_t}{\partial\beta_k} - \mathbf{B}_t\mathbf{F}_t^\top\mathbf{Q}_t^{-1}\left(\frac{\partial\mathbf{F}_t}{\partial\beta_k}\mathbf{G}_t + \mathbf{F}_t\frac{\partial\mathbf{G}_t}{\partial\beta_k}\right)\right\}\boldsymbol{\tau}_{t-1},$$

para $k = 1, \ldots, l_2$ e

$$\boldsymbol{\mathcal{D}}_{\gamma_k}(t) = \left(\mathbf{G}_t - \mathbf{B}_t\mathbf{F}_t^\top\mathbf{Q}_t^{-1}\mathbf{F}_t\mathbf{G}_t\right)\boldsymbol{\mathcal{D}}_{\gamma_k}(t-1),$$

para $k = 1, \ldots, l_3$, em que

$$\frac{\partial\mathbf{F}_t}{\partial\alpha_{ik}} = \begin{pmatrix} 0 \\ \vdots \\ \mu_{it}\dfrac{x_{tk}\boldsymbol{\varphi}_i^\top}{\varphi(B)(\tau_t)} \\ \vdots \\ 0 \end{pmatrix},$$

$$\frac{\partial\mathbf{F}_t}{\partial\beta_k} = \mathrm{diag}\left\{\mu_{it}\frac{\varphi_i(B)(\tau_t)z_{tk} - \varphi_i(B)(\tau_t z_{tk})}{\varphi^2(B)(\tau_t)}\right\}_{d\times d} \times \begin{pmatrix} \boldsymbol{\varphi}_1^\top \\ \vdots \\ \boldsymbol{\varphi}_d^\top \end{pmatrix},$$

e

$$\frac{\partial\mathbf{G}_t}{\partial\beta_k} = \begin{pmatrix} 0 \\ \vdots \\ \tau_t\dfrac{\psi(B)(\tau_t)z_{tk} - \psi(B)(\tau_t z_{tk})}{\psi^2(B)(\tau_t)}\boldsymbol{\psi} \end{pmatrix}.$$

As iterações se iniciam com $\mathcal{D}_\alpha(0) = \mathbf{0}$, $\mathcal{D}_\beta(0) = \mathbf{0}$ e $\mathcal{D}_\gamma(0) = g \begin{pmatrix} \mathbf{w}^\top \\ \vdots \\ \mathbf{w}^\top \end{pmatrix}$.

Vamos denotar por $\mathcal{D}^*_{\alpha_{ik}}(t)$ a derivada do suavizador de Kalman em relação a α_{ik}, no tempo t. Como $\mathbf{m}^*_t = \mathbf{m}_t + \mathbf{D}_t\mathbf{B}^{-1}_{t+1}(\mathbf{m}^*_{t+1} - \mathbf{G}_{t+1}\mathbf{m}_t)$ e $\mathrm{E}(\mathbf{m}^*_{t+1} - \mathbf{G}_{t+1}\mathbf{m}_t) = \mathbf{0}$, obtemos que

$$\mathcal{D}^*_{\alpha_{ik}}(t) = \mathcal{D}_{\alpha_{ik}}(t) + \mathbf{D}_t\mathbf{B}^{-1}_{t+1}\left\{\mathcal{D}^*_{\alpha_{ik}}(t+1) - \mathbf{G}_{t+1}\mathcal{D}_{\alpha_{ik}}(t)\right\},$$

$$\mathcal{D}^*_{\beta_k}(t) = \mathcal{D}_{\beta_k}(t) + \mathbf{D}_t\mathbf{B}^{-1}_{t+1}\left\{\mathcal{D}^*_{\beta_k}(t+1) - \frac{\partial \mathbf{G}_{t+1}}{\partial \beta_k}\boldsymbol{\tau}_t - \mathbf{G}_{t+1}\mathcal{D}_{\beta_k}(t)\right\},$$

e

$$\mathcal{D}^*_{\gamma_k}(t) = \mathcal{D}_{\gamma_k}(t) + \mathbf{D}_t\mathbf{B}^{-1}_{t+1}\left\{\mathcal{D}^*_{\gamma_k}(t+1) - \mathbf{G}_{t+1}\mathcal{D}_{\gamma_k}(t)\right\}.$$

B.4 Matriz de variabilidade

Para obter as covariâncias dos resíduos que aparecem nas expressões que seguem, utilizamos os resultados apresentados na Tabela 4.6, notando que $\varepsilon^*_{ijt} = \boldsymbol{\varepsilon}^*_{jt}(i)$ e $\xi^*_{jt} = \boldsymbol{\xi}^*_{jt}(M)$.

Temos que $V = \mathrm{E}\{\mathcal{F}\mathcal{F}^\top\}$. Logo, como

$$\begin{aligned}
\mathcal{F}_{\alpha_{ij}}(\boldsymbol{\eta}) &= \sum_{i=1}^{n_j} \frac{\left(\boldsymbol{\varphi}^\top_{ij}\boldsymbol{\tau}_{jt}\right)^{p_i-1}}{\mu^{p_i-1}_{ijt}}\left(Y_{ijt} - \mu_{ijt}\frac{\boldsymbol{\varphi}^\top_{ij}\mathbf{m}^*_{jt}}{\boldsymbol{\varphi}^\top_{ij}\boldsymbol{\tau}_{jt}}\right)\mathbf{x}_{jt} \\
&= \sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\varphi}^\top_{ij}\boldsymbol{\tau}_{jt}\right)^{p_i-1}}{\mu^{p_i-1}_{ijt}}\varepsilon^*_{ijt}\mathbf{x}_{jt},
\end{aligned}$$

temos

$$\mathcal{F}_{\alpha_{ij}}(\boldsymbol{\eta})\mathcal{F}^\top_{\alpha_{ij}}(\boldsymbol{\eta}) = \left(\sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\varphi}^\top_{ij}\boldsymbol{\tau}_{jt}\right)^{p_i-1}}{\mu^{p_i-1}_{ijt}}\varepsilon^*_{ijt}\mathbf{x}_{jt}\right)\left(\sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\varphi}^\top_{ij}\boldsymbol{\tau}_{jt}\right)^{p_i-1}}{\mu^{p_i-1}_{ijt}}\varepsilon^*_{ijt}\mathbf{x}^\top_{jt}\right)$$

o que nos dá

$$\mathbf{V}_{\alpha_{ij}} = \sum_{s=1}^{n_j}\sum_{t=1}^{n_j} \frac{\left(\boldsymbol{\varphi}^\top_{ij}\boldsymbol{\tau}_{js}\right)^{p_i-1}}{\mu^{p_i-1}_{ijs}}\frac{\left(\boldsymbol{\varphi}^\top_{ij}\boldsymbol{\tau}_{jt}\right)^{p_i-1}}{\mu^{p_i-1}_{ijt}}\mathrm{Cov}(\varepsilon^*_{ijs}, \varepsilon^*_{ijt})\mathbf{x}_{js}\mathbf{x}^\top_{jt}.$$

Escrevendo a função de estimação para $\boldsymbol{\beta}_j$ em termos dos resíduos, obtemos

$$\mathcal{F}_{\beta_j} = \sum_{i=1}^{n_j}\sum_{i=1}^{d} \frac{\left(\boldsymbol{\varphi}^\top_{ij}\boldsymbol{\tau}_{jt}\right)^{p_i-2}}{\nu_i^2\mu^{p_i-1}_{ijt}}\varepsilon^*_{ijt}\boldsymbol{\varphi}\mathbf{z}_{jt} + \sum_{i=1} n_j \frac{\left(\boldsymbol{\psi}^\top_j\boldsymbol{\tau}_{j,t-1}\right)^{q-2}}{\sigma^2\tau^{q-1}_{jt}}\xi^*_{jt}\boldsymbol{\psi}\mathbf{z}_{jt},$$

em que $\boldsymbol{\varphi}\mathbf{z}_{jt}$ e $\boldsymbol{\psi}\mathbf{z}_{jt}$ denotam $\{\varphi_{ij}(B)(\tau_{jt})\mathbf{z}_{jt}-\varphi_{ij}(B)(\tau_{jt}\mathbf{z}_{jt})\}$ e $\{\psi_j(B)(\tau_{jt})\mathbf{z}_{jt}-\psi_j(B)(\tau_{jt}\mathbf{z}_{jt})\}$, respectivamente. Temos então

$$\begin{aligned}
\mathbf{V}_{\beta_j} &= \sum_{s=1}^{n_j}\sum_{t=1}^{n_j}\sum_{i=1}^{d}\sum_{k=1}^{d}\frac{\left(\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{js}\right)^{p_i-2}}{\nu_i^2\mu_{ijs}^{p_i-1}}\frac{\left(\boldsymbol{\varphi}_{kj}^{\top}\boldsymbol{\tau}_{jt}\right)^{p_k-2}}{\nu_k^2\mu_{kjt}^{p_k-1}}\mathrm{Cov}(\varepsilon_{ijs}^*,\varepsilon_{kjt}^*)\boldsymbol{\varphi}\mathbf{z}_{js}\boldsymbol{\varphi}\mathbf{z}_{jt}^{\top} \\
&+ \sum_{t=1}^{n_j}\sum_{s=1}^{n_j}\frac{\left(\boldsymbol{\psi}_j^{\top}\boldsymbol{\tau}_{j,s-1}\right)^{q-2}}{\sigma^4\tau_{js}^{q-1}}\frac{\left(\boldsymbol{\psi}_j^{\top}\boldsymbol{\tau}_{j,t-1}\right)^{q-2}}{\tau_{jt}^{q-1}}\mathrm{Cov}(\xi_{js}^*,\xi_{jt}^*)\boldsymbol{\psi}\mathbf{z}_{js}\boldsymbol{\psi}\mathbf{z}_{jt}^{\top} \\
&+ 2\sum_{s=1}^{n_j}\sum_{t=1}^{n_j}\sum_{i=1}^{d}\frac{\left(\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{js}\right)^{p_i-2}}{\nu_i^2\mu_{ijs}^{p_i-1}}\frac{\left(\boldsymbol{\psi}_j^{\top}\boldsymbol{\tau}_{j,t-1}\right)^{q-2}}{\sigma^2\tau_{jt}^{q-1}}\mathrm{Cov}(\varepsilon_{ijs}^*,\xi_{jt}^*)\boldsymbol{\varphi}\mathbf{z}_{js}\boldsymbol{\psi}\mathbf{z}_{jt}^{\top},
\end{aligned}$$

e

$$\mathbf{V}_{\gamma}=\sum_{i=1}^{h}\sum_{l=1}^{h}\frac{1}{g_j^{r-1}g_l^{r-1}}\mathrm{Cov}(\xi_{j0}^*,\xi_{l0}^*)\mathbf{w}_j\mathbf{w}_l^{\top},$$

em que $\xi_{j0}^*=m_{j0}^*-g_j$. Temos também

$$\begin{aligned}
\mathbf{V}_{\alpha_{ij}\beta_j} &= \sum_{s=1}^{n_j}\sum_{t=1}^{n_j}\sum_{i=1}^{d}\frac{\left(\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{js}\right)^{p_i-1}}{\mu_{ijs}^{p_i-1}}\frac{\left(\boldsymbol{\varphi}_{kj}^{\top}\boldsymbol{\tau}_{jt}\right)^{p_k-2}}{\nu_k^2\mu_{kjt}^{p_k-1}}\mathrm{Cov}(\varepsilon_{ijs}^*,\varepsilon_{kjt}^*)\mathbf{x}_{js}\boldsymbol{\varphi}\mathbf{z}_{jt}^{\top} \\
&+ \sum_{s=1}^{n_j}\sum_{t=1}^{n_j}\frac{\left(\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{js}\right)^{p_i-1}}{\mu_{ijs}^{p_i-1}}\frac{\left(\boldsymbol{\psi}_j^{\top}\boldsymbol{\tau}_{j,t-1}\right)^{q-2}}{\sigma^2\tau_{jt}^{q-1}}\mathrm{Cov}(\varepsilon_{ijs}^*,\xi_{jt}^*)\mathbf{x}_{js}\boldsymbol{\psi}\mathbf{z}_{jt}^{\top},
\end{aligned}$$

$$\mathbf{V}_{\alpha_{ij}\gamma}=\sum_{t=1}^{n_j}\frac{\left(\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{js}\right)^{p_i-1}}{\mu_{ijs}^{p_i-1}}\frac{1}{g_j^{r-1}}\mathrm{Cov}(\varepsilon_{ijs}^*\xi_{j0}^*)\mathbf{x}_{js}\mathbf{w}_j^{\top}$$

e, finalmente,

$$\begin{aligned}
\mathbf{V}_{\beta_j\gamma} &= \sum_{t=1}^{n_j}\sum_{i=1}^{d}\frac{\left(\boldsymbol{\varphi}_{ij}^{\top}\boldsymbol{\tau}_{js}\right)^{p_i-2}}{\nu_i^2\mu_{ijs}^{p_i-1}}\frac{1}{g_j^{r-1}}\mathrm{Cov}(\varepsilon_{ijs}^*,\xi_{j0}^*)\boldsymbol{\varphi}\mathbf{z}_{js}\mathbf{w}_J^{\top} \\
&+ \sum_{t=1}^{n_j}\frac{\left(\boldsymbol{\psi}_j^{\top}\boldsymbol{\tau}_{j,t-1}\right)^{q-2}}{\sigma^2\tau_{jt}^{q-1}}\mathrm{Cov}(\xi_{jt}^*,\xi_{j0}^*)\boldsymbol{\psi}\mathbf{z}_{jt}\mathbf{w}_k^{\top}.
\end{aligned}$$

Apêndice C

Gráficos

Neste apêndice, apresentamos os gráficos referentes às amostras simuladas discutidas na Seção 4.7. Para cada uma das amostras, foram construídos 4 tipos de gráficos,

1. Gráficos da função $\mathcal{L}$. Em (a), as linhas pontilhadas são gráficos das funções $\mathcal{L}\{\boldsymbol{\psi}; \hat{\boldsymbol{\eta}}(\boldsymbol{\psi}_j)\}$, para $j = 1, \ldots, 11$ ($\psi_{2,j} = 0.0, 0.1, \ldots, 1.0$); a linha cheia é o gráfico da função $\mathcal{L}\{\boldsymbol{\psi}_j; \hat{\boldsymbol{\eta}}(\boldsymbol{\psi}_j)\}$;

2. mesmo que em 1, para a função $\mathcal{L}_1$ (b);

3. mesmo que em 1, para a função $\mathcal{L}_2$ (c);

4. gráfico dos valores de $\hat{\sigma}^2$ em função de ψ_2.

Para algumas amostras, são apresentados também gráficos das funções de autocorrelação e autocorrelação parcial dos resíduos referentes a ajustes com alguns valores de ψ_2.

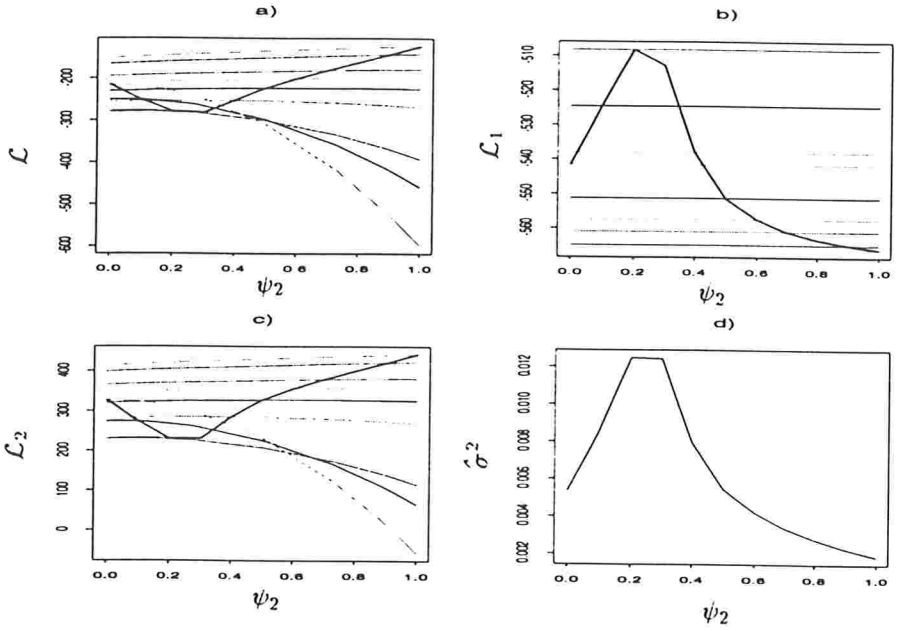

Figura C.1: Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 1

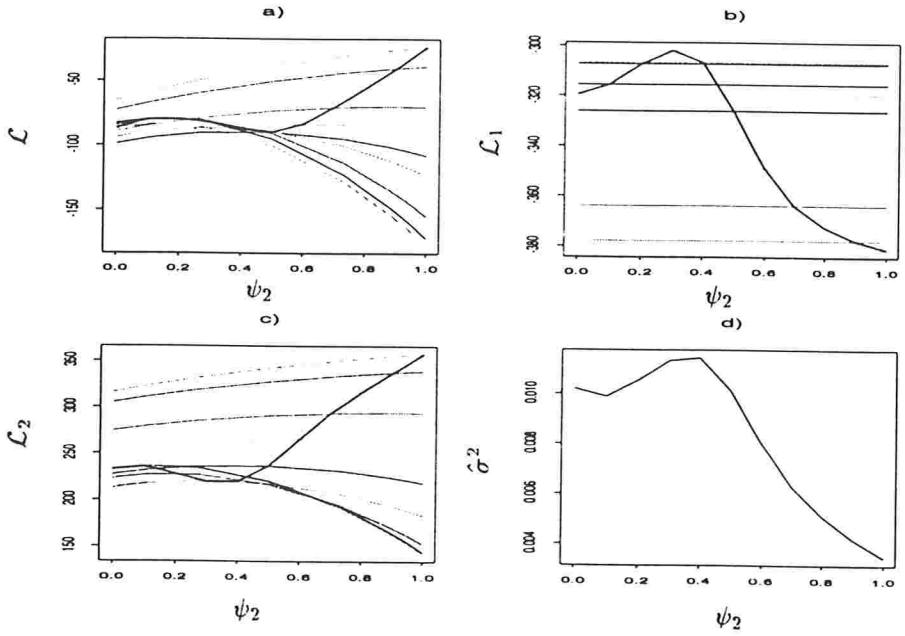

Figura C.2: Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 2

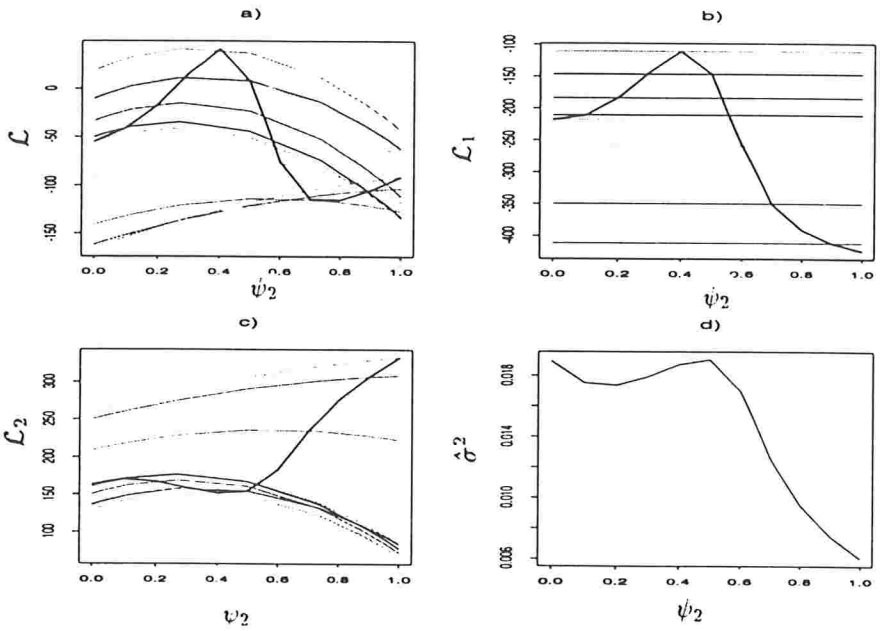

Figura C.3: Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 3

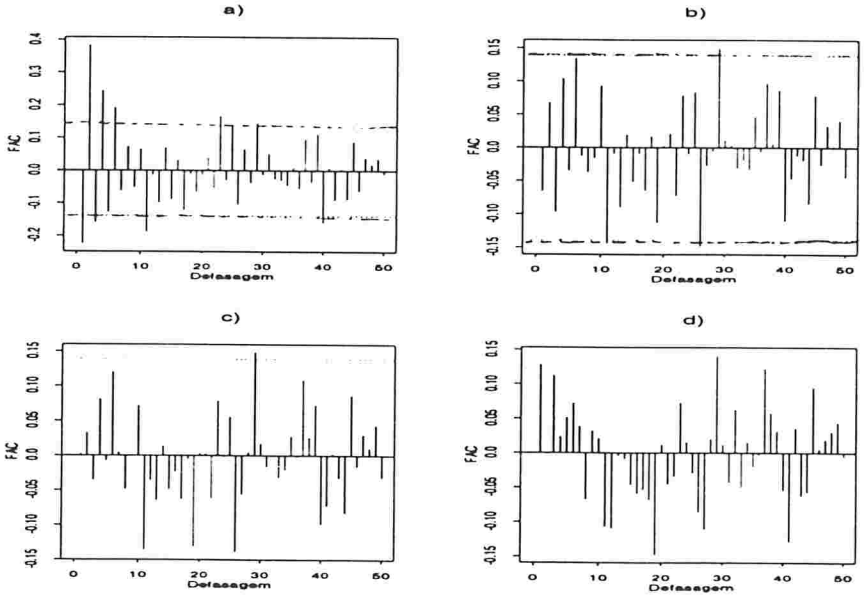

Figura C.4: Funções de autocorrelação dos resíduos dos ajustes com $\psi^{\top} = (0.0, 1.0)$, (a) $\psi^{\top} = (0.6, 0.4)$ (b), $\psi^{\top} = (0.7, 0.3)$ (c) e $\psi^{\top} = (1.0, 0.0)$ (d), para a amostra 3.

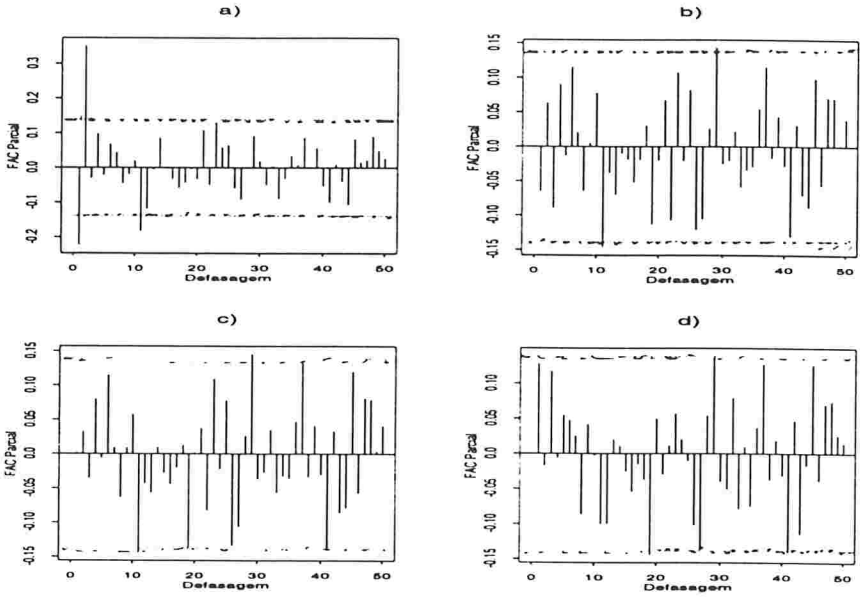

Figura C.5: Funções de autocorrelação parcial dos resíduos dos ajustes com $\psi^{\top} = (0.0, 1.0)$, (a) $\psi^{\top} = (0.6, 0.4)$ (b), $\psi^{\top} = (0.7, 0.3)$ (c) e $\psi^{\top} = (1.0, 0.0)$ (d), para a amostra 3.

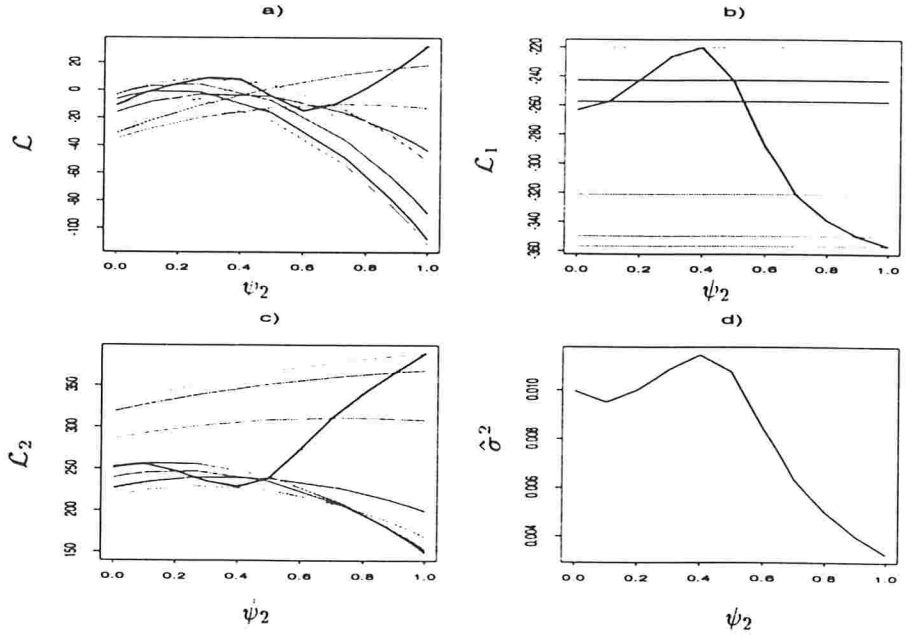

Figura C.6: Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 4

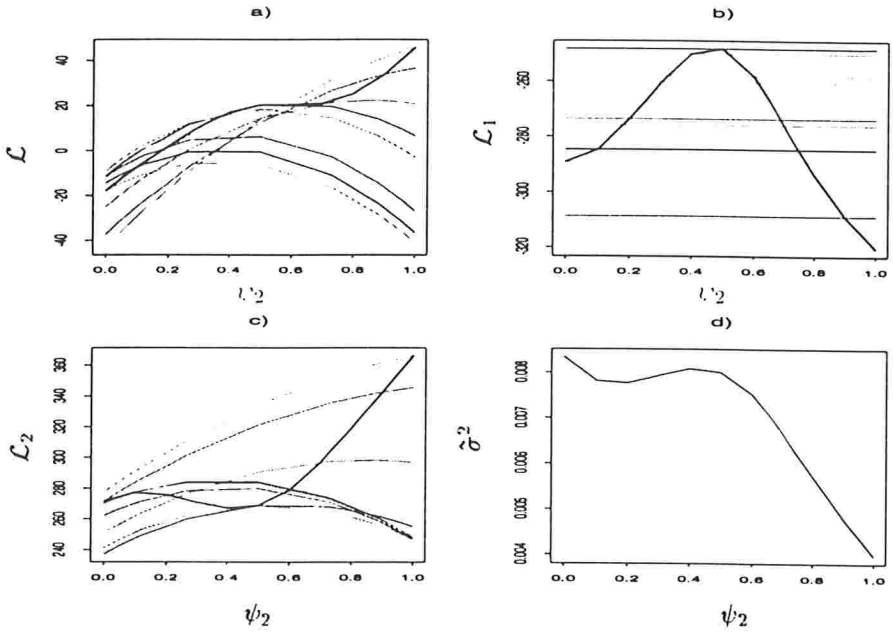

Figura C.7: Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 5

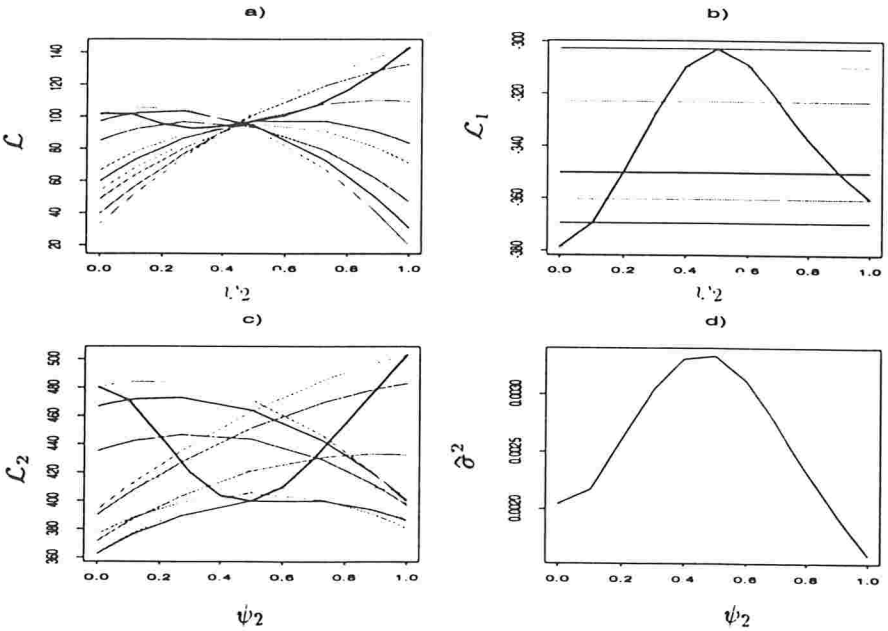

Figura C.8: Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 6

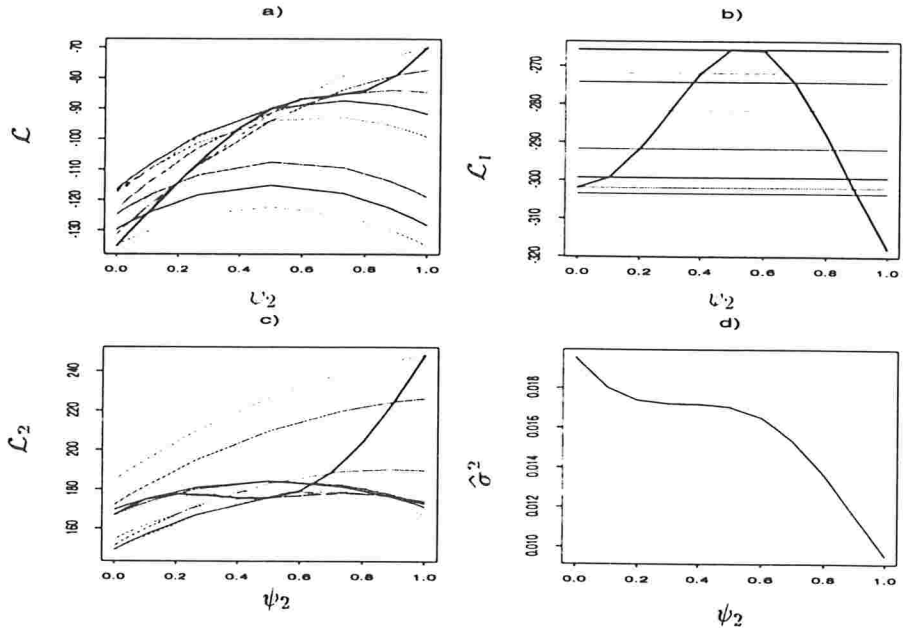

Figura C.9: Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 7

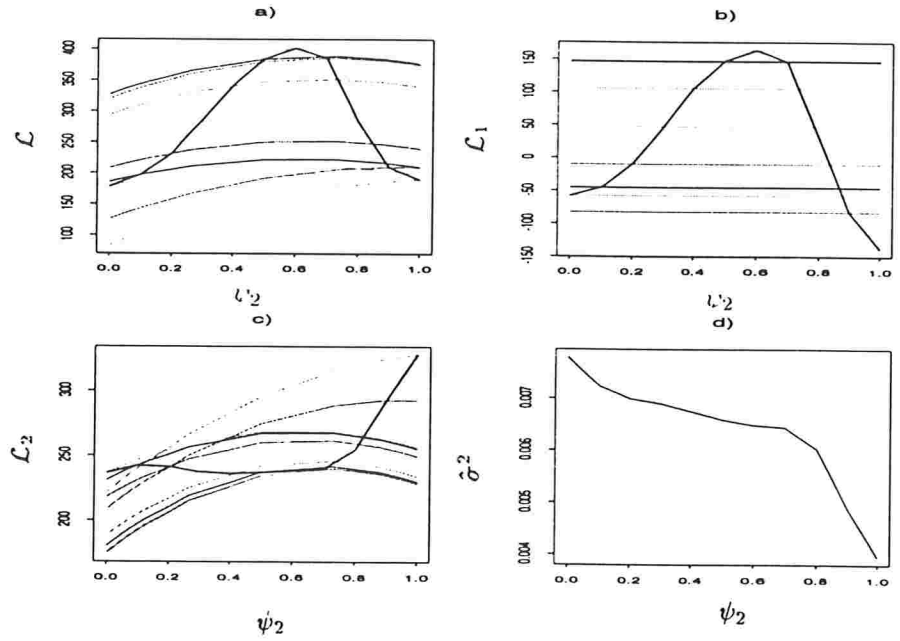

Figura C.10: Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 8

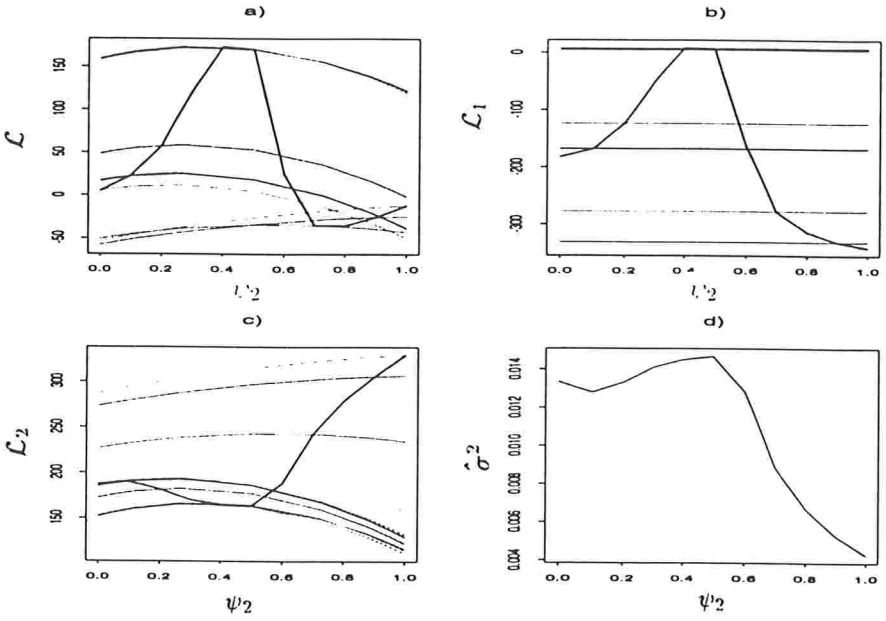

Figura C.11: Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 9

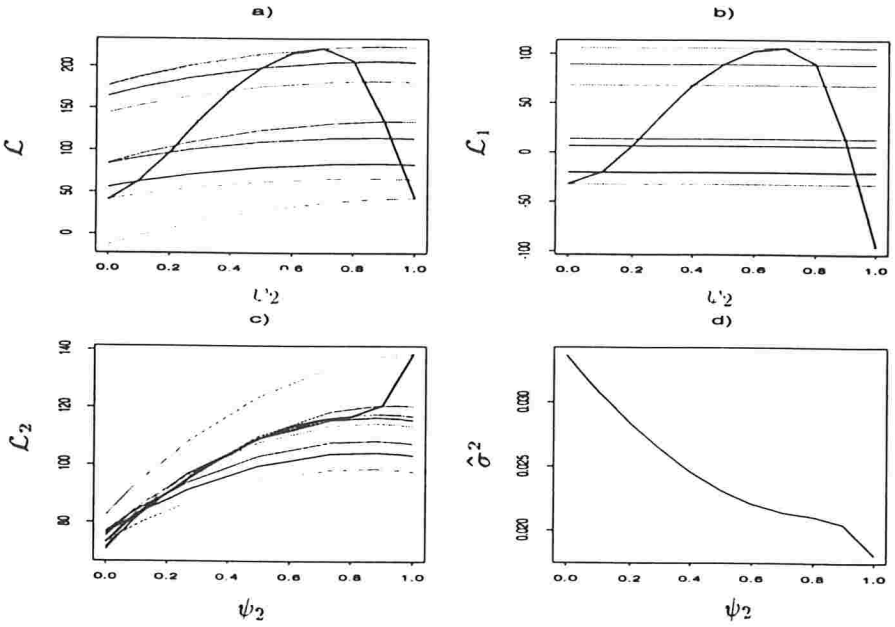

Figura C.12: Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 10

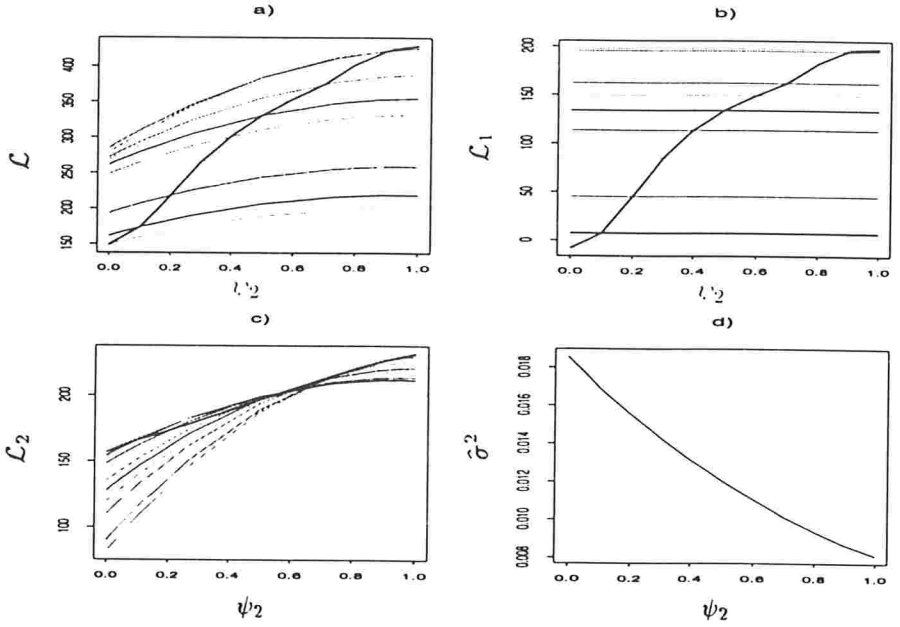

Figura C.13: Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 11

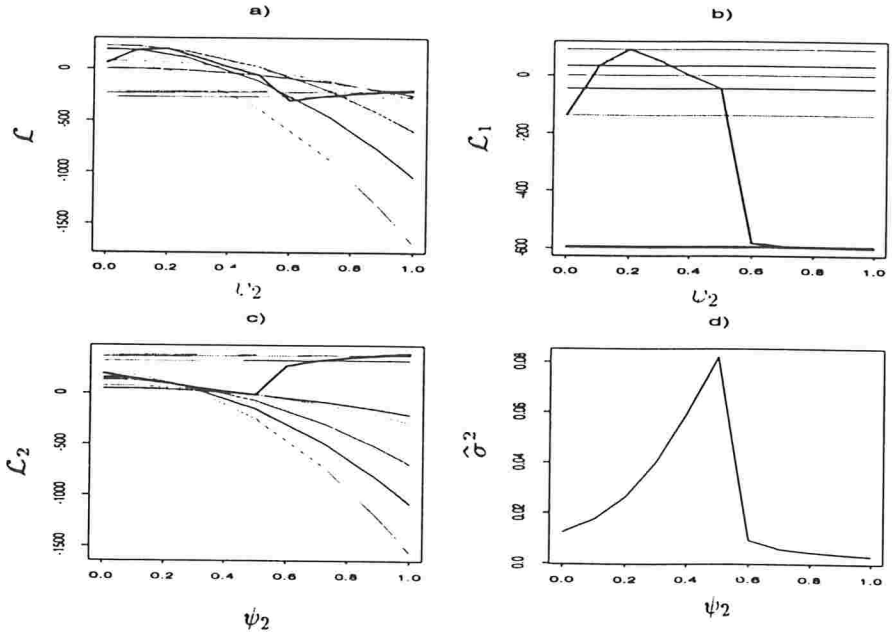

Figura C.14: Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 12

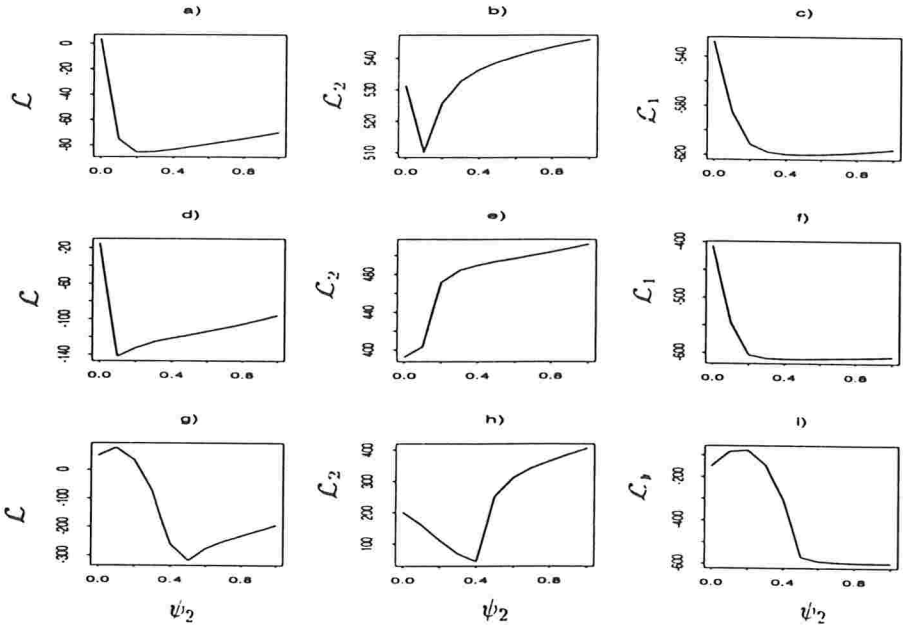

Figura C.15: Comportamento das funções $\mathcal{L}$ (a,d,g), $\mathcal{L}_1$ (c, f, i) e $\mathcal{L}_2$ (b, e, h) ao longo do processo de estimação para a amostra 12. Gráficos (a, b, c): após iteração 1; (d, e, f): após iteração 5 e (g, h , i): após iteração 50

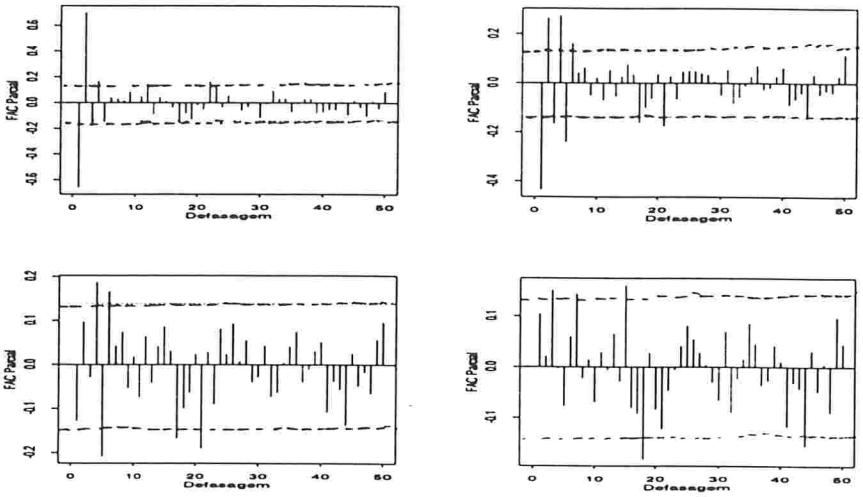

Figura C.16: Funções de autocorrelação parcial dos resíduos dos ajustes com $\psi^{\mathsf{T}} = (0.0, 1.0), (0.8, 0.2), (0.9, 0.1)$ e (1.0,0.0) para a amostra 12. O ponto (0.8,0.2) corresponde ao máximo de $\mathcal{L}_1$ (canto superior direito) e o ponto (1.0,0.0) (no canto inferior direito), foi utilizado na simulação desta amostra.

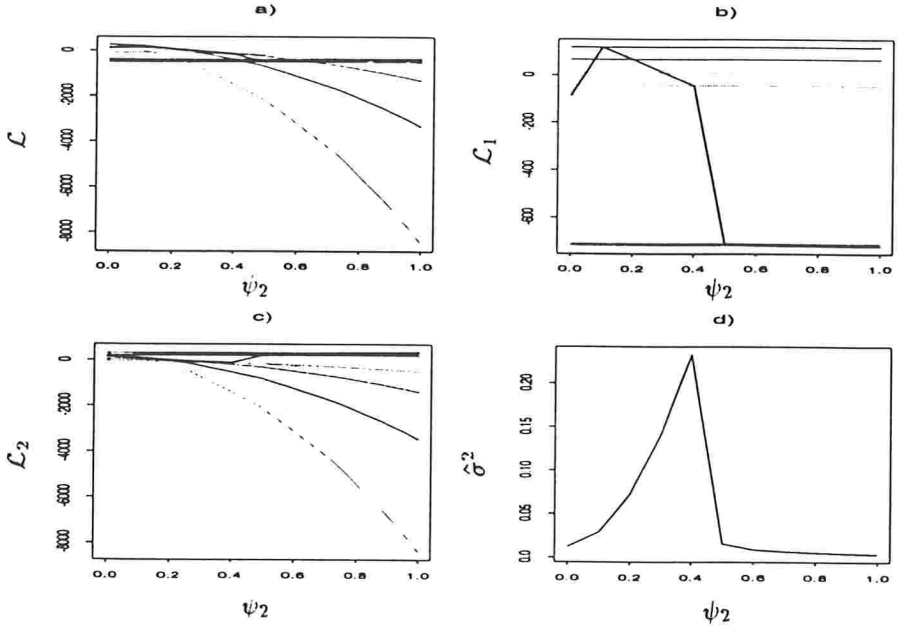

Figura C.17: Gráficos de $\mathcal{L}$ (a), $\mathcal{L}_1$ (b), $\mathcal{L}_2$ (c) e $\hat{\sigma}^2$ (d) em função de ψ_2 para a amostra 13

Referências bibliográficas

[1] Anderson, D. O. B. and Moore, J. B. (1979). *Optimal Filtering.* Englewood-Cliffs, NJ: Prentice-Hall.

[2] Artes, R. (1997). *Extensões da Teoria das Equações de Estimação Generalizadas a Dados Circulares e Modelos de Dispersão.* São Paulo: IME-USP. Tese de Doutorado.

[3] Artes, R. and Jørgensen, B. (1998). *Longitudinal Data Estimating Equations for Dispersion Models.* São Paulo: IME-USP. Relatório Técnico (RT-MAE 9802).

[4] Botter, D. A., Peres, A. A. Q and Zandonade, E. (1998). *A longitudinal study of mortality and air pollution for São Paulo, Brazil.* São Paulo: IME-USP. Relatório Técnico (RT-MAE 9834).

[5] Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

[6] Cordeiro, G. (1992). *Introdução à Teoria da Verossimilhança.* Rio de Janeiro: UFRJ. X SINAPE.

[7] Dempster, A. P., Laird, N.M. and Rubin, D. B. (1977). Maximum likelihood estimation from the EM algorithm (with discussion). *J. R. Statist. Soc.* B **39**, 1-38.

[8] Diggle, P.J., Liang, K.-Y. and Zeger, S.L. (1994). *The Analysis of Longitudinal Data.* Oxford: Oxford University Press.

[9] Gay, D. M. (1984). A trust region approach to linearly constrained optimization. In *Numerical Analysis. Proceedings*, Dundee 1983, F. A Lootsma (ed.), Berlin: Springer, p. 171-189

[10] Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, **31**, 1208-1211.

[11] Harvey, A.C. (1981). *Time Series Models.* Oxford: Allan.

[12] Huber, P. J. (1981). *Robust Statistics.* New York: John Wiley.

[13] Jørgensen, B. (1997). *The Theory of Dispersion Models.* London: Chapman and Hall.

[14] Jørgensen, B., and Labouriau, R. S. (1994). *Exponential Families and Theoretical Inference.* Vancouver: Lecture Notes, Departament of Statistics, University of British Columbia.

[15] Jørgensen, B. and Tsao, M. (1998). Dispersion Models and Longitudinal Data Analysis. *Statist. Med.* **18**, 2257–2270.

[16] Jørgensen, B., Lundbye-Christensen, S., Song, P.X.-K. and Sun, L. (1996a). A longitudinal study of emergency room visits and air pollution for Prince George, British Columbia. *Statist. Med.* **15**, 823–836.

[17] Jørgensen, B., Lundbye-Christensen, S., Song, P.X.-K. and Sun, L. (1996b). State-space models for multivariate longitudinal data of mixed types. *Can. J. Statist.* **24**, 385–402.

[18] Jørgensen, B., Lundbye-Christensen, S., Song, P.X.-K. and Sun, L. . (1999). A state space model for multivariate longitudinal count data. *Biometrika* **86**, 169–181.

[19] Knudsen, S. J. (1999). *Estimating Functions and Separate Inference.* Odense: Monographs Vol. 1, Departament of Statistics and Demography, University of Southern Denmark.

[20] Koyama, M. A. H. (1997). *Modelo de Zeger para Análise de Séries de Contagens.* São Paulo: IME-USP. Dissertação de Mestrado.

[21] Liang, K.-Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

[22] Liang, K.-Y. and Qaquish, B. (1988). Markov regression models for time series: a quasi-likelihood approach, *Biometrics* **44**, 1019-1031.

[23] Lima, A. C. P. e Sañudo, A. (1996). *Transferência entre Tarefas Sincronizatórias com Diferentes Níveis de Complexidade.* São Paulo: IME-USP (RAE-CEA-9602).

[24] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* 2th. ed. London: Chapman & Hall.

[25] Shumway, R. H. (1988). *Applied Statistical Time Series Analysis.* Englewood-Cliffs, NJ: Prentice-Hall.

[26] Singer, J.M. e Conceição, G.M.S.(1994). *Estudo da Associação entre Mortalidade e Poluição Atmosférica na Cidade de São Paulo.* São Paulo: IME-USP (RAE-CEA-9407).

[27] Singer, J.M. e Andrade, D.F. (2000). Analysis of longitudinal data. In *Handbook of Statistics,* **18***: Bio-Enviromental and Public Health Statistics*, eds. P. K. Sen and C.R. Rao. Amsterdam: North Holland. 115-160.

[28] Skjøth,F. and Lundbye-Christensen, S. (1997). An investigation of a state-space model for longitudinal count data, with an application to the modelling of cucumber yield. Foulum: Technical Report #97-9, Biometry Research Unit, DIAS, Denmark.

[29] Wedderburn (1974). Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* **61**, 439-47.

[30] West, M., Harrison, P.J. and Migon, H.S. (1985). Dynamic generalized linear models and Bayesian forecasting. *J. Amer. Statist. Assoc.* **80**, 73–97.

[31] West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic models*, 2. ed. New York: Springer Verlag.

[32] Zeger, S. L. (1988). A regression model for time series counts. *Biometrika* **75**, 621–629.