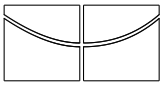

Universidade de Brasília

Instituto de Ciências Exatas

Departamento de Estatística

Dissertação de Mestrado

Modelos Dinâmicos Dirichlet

por

Guilherme Souza Rodrigues

Orientador: Prof.ª Dr.ª Cibele Queiroz da Silva

Coorientador: Prof. Dr. Helio dos Santos Migon

Julho de 2011

Guilherme Souza Rodrigues

Modelos Dinâmicos Dirichlet

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidadede de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

Universidade de Brasília

Brasília, julho de 2011

À minha esposa, por haver,

ao meu lado, superado esse desafio.

Agradecimentos

Em primeiro lugar, agradeço profundamente à minha família. Em especial, à minha esposa Thais por ter me conferido todo o amor e suporte necessário a execução dessa extensa atividade. Aos meus pais Valdir e Aida pelo apoio sempre incondicional e por terem sido, e ainda serem, responsáveis pela minha formação como pessoa e como profissional. Aos meus irmãos Leonardo e Luciano pela postura sempre bondosa e de cumplicidade. Todos estes merecem minha gratidão e admiração profunda por motivos que extrapolam infinitamente os aspectos relacionados a essa dissertação. Tenho ciência do privilegio de tê-los presentes em minha vida.

Agradeço à minha orientadora Cibele Queiroz da Silva pela enorme dedicação e envolvimento ao longo dessa parceria. Sua admirável devoção ao trabalho de pesquisa contribuiu sobremaneira ao alcance dos bons resultados obtidos nesse projeto.

Agradeço ao meu co-orientador Helio dos Santos Migon e aos professores Mariane Branco Alves e André Luis Fernandes Cançado pelas valiosas, e sempre muito oportunas, considerações e sugestões ao trabalho.

Aos gestores do TJDFT Paulo Bandeira, Liz Criciny Werlang Rauber Kopper e Laura Lima, agradeço imensamente ao apoio em todos os momentos. Sem a boa vontade dispensada, certamente não teria sido possível conciliar o trabalho e a pesquisa. Também agradeço aos meus queridos colegas de equipe Andréa, Kelly, Larissa e Rogério, pela amizade, pela compreensão, pelo incentivo e pelo sacrifício que fizeram enquanto vigorou minha licença.

Sumário

Lista de Figuras **3**

Lista de Códigos **5**

Lista de Símbolos **6**

Resumo **8**

Abstract **9**

Introdução **10**

1 Modelos Dinâmicos Lineares **13**

1.1 Motivação 13

1.2 Formulação do modelo 15

1.3 Estimação do sistema e previsão 17

1.3.1 Filtragem 17

1.3.2 Suavização 20

1.3.3 Previsão 21

1.4 Especificação do modelo 23

1.4.1 Modelos dinâmicos lineares polinomiais 23

1.4.2 Modelos dinâmicos lineares sazonais 27

1.4.3 Modelos dinâmicos lineares de regressão 32

1.4.4 Sobreposição de modelos lineares dinâmicos 34

1.5 Modelos com parâmetros desconhecidos 36

1.5.1 Procedimentos inferenciais *online* 38

1.5.1.1 Especificação das matrizes de covariância W_t 38

1.5.1.2 Modelo linear dinâmico para V desconhecido 41
1.5.2 Procedimentos inferenciais *offline* 42

2 Modelos Dinâmicos Lineares Generalizados 45
2.1 Caso geral 46
2.1.1 Procedimento inferencial 47
2.2 Modelo dinâmico Beta 51
2.2.1 Procedimento inferencial 52

3 Modelo dinâmico Dirichlet 56
3.1 Modelo Dinâmico Dirichlet - Parâmetros conhecidos 57
3.1.1 Procedimento inferencial *online* 60
3.2 Modelo dinâmico Dirichlet - Parâmetros desconhecidos 66
3.2.1 Especificação da priori e estimação *offline* via MCMC 70
3.2.1.1 Amostrando as observações faltantes 72
3.2.1.2 Amostrando os estados do sistema - Caso Dinâmico . 74
3.2.1.3 Amostrando os estados do sistema - Caso Estático . 79
3.2.1.4 Amostrando a matriz de covariância dos erros 80
3.2.1.5 Amostrando o estado inicial do sistema 84
3.2.1.6 Amostrando o parâmetro de precisão 84

4 Casos especiais e aplicações 86
4.1 Modelos univariados - Caso Beta 87
4.1.1 Ajuste de dados simulados 87
4.1.2 Regressão Beta - comparação com a abordagem clássica 93
4.1.3 MDD: Uma aplicação a dados de internações no DF 96
4.2 Modelos multivariados - Caso Dirichlet 101
4.2.1 Ajuste de dados simulados 101
4.2.2 Regressão Dirichlet - estimação num contexto simulado 108
4.2.3 MDD: Uma aplicação a dados de óbitos no Brasil 112

5 Conclusão e trabalhos futuros 118

Referências Bibliográficas 122

Lista de Figuras

1.1 Estrutura de dependência para o espaço de estados do modelo 16

1.2 Análise sequencial do modelo linear dinâmico normal 19

1.3 Serie histórica y_t (linhas cinzas), média real μ_t (linhas vermelhas) e estimativas (linhas pretas) com respectivos intervalos de credibilidade de 95% (linhas pretas tracejadas). (a) Valores filtrados para μ_t; (b) Valores previstos um passo a frente para y_t; (c) Valores suavizados para μ_t. 25

1.4 (a) Valores observados da série com dados faltantes y_t (linha cinza); (b) Valores reais (linha vermelha) e suavizados para μ_t (linha preta). Intervalos de credibilidade de 95% (linhas pretas tracejadas). 31

2.1 Análise sequencial do modelo linear generalizado dinâmico 50

4.1 (a) Serie histórica y_t (linhas cinzas); (b) Médias μ_t; (c) Níveis θ_{1t}; (d) Efeitos de sazonalidade $\theta_{3t} + \theta_{5t}$; (e) Efeito da regressão θ_{6t}. Valores reais (linhas vermelhas), valores estimados (linhas pretas) e intervalos de credibilidade de 95% (linhas pretas tracejadas). 92

4.2 (a) Serie histórica y_t (linhas cinzas); (b) Médias μ_t; (c) Níveis θ_{1t}; (d) Efeitos de sazonalidade $\theta_{3t} + \theta_{5t}$. Valores estimados (linhas pretas) e intervalos de credibilidade de 95% (linhas pretas tracejadas). 99

4.3 Proporção de internações por doenças do aparelho respiratório no DF (linhas cinzas) e valores previstos para o ano de 2010 (linha preta) com respectivos intervalos de credibilidade de 90% (linhas pretas tracejadas). 101

4.4 Proporções observadas (linhas cinzas), médias reais (linhas vermelhas) e médias estimadas (linha preta) com intervalos de credibilidade de 95% (linhas pretas tracejadas). 106
4.5 Valores reais (pontos vermelhos) e estimativas intervalares de credibilidade de 95% (linhas pretas) dos parâmetros da regressão θ_i. 112
4.6 Proporções observadas de óbitos por A - acidentes de transporte (linha cinza), B - agressões e lesões autoprovocadas intencionalmente (linha azul) e C - outras causas externas de mortalidade (linha marrom). Estimativas das médias das categorias (linhas pretas) e intervalos de credibilidade de 95% (linhas pretas tracejadas). 115
4.7 Proporções observadas de óbitos, em 2009 e 2010, por A - acidentes de transporte (linha cinza), B - agressões e lesões autoprovocadas intencionalmente (linha azul) e C - outras causas externas de mortalidade (linha marrom). Previsões das proporções para 2011 e 2012 com intervalos de credibilidade de 90% (linhas tracejadas). 116

Lista de Códigos

1.1 Modelos dinâmicos lineares polinomiais 26
1.2 Modelos dinâmicos lineares sazonais 31
1.3 Uso dos MDL em regressão linear 33
4.1 Ajuste de dados simulados univariados 89
4.2 MDD estático e modelo clássico de regressão Beta 94
4.3 Ajuste de dados de internações no Distrito Federal 97
4.4 Ajuste de dados simulados multivariados 104
4.5 Modelo de regressão Dirichlet 109
4.6 Ajuste de dados sobre óbitos por causas externas 113

Lista de Símbolos e Siglas

Símbolos

Y_t	Série temporal observável
k	Dimensão do vetor Y_t
θ_t	Vetor de parâmetros do sistema no tempo t
θ_0	Estado inicial do sistema
ϕ_t	Parâmetro de precisão da distribuição das observações no tempo t
V_t	Inverso do parâmetro de precisão ϕ no tempo t
μ_t	Esperança de $(Y_t\|\theta_t, \phi)$
η_t	Parâmetro canônico da distribuição das observações no tempo t (Cap. 2)
λ_t	Preditor linear no tempo t
$g(\cdot)$	Função de ligação
F_t	Matriz que relaciona o sistema θ_t ao preditor linear λ_t
G_t	Matriz de evolução do sistema do tempo t
w_t	Erro de evolução do sistema no tempo t
W_t	Matriz de covariância dos erros de evolução w_t.
δ	Vetor dos fatores de desconto
ψ	Vetor de parâmetros desconhecidos na especificação do modelo
Θ	Vetor $(\theta_1, \ldots, \theta_T)$
A^c	Vetor contendo os índices t das observações faltantes
D_t	Informação disponível no tempo t
T	Tempo final da série temporal Y_t
E_n	Vetor de dimensão n da forma $(1, 0, \ldots, 0)$
$J_n(\cdot)$	Matriz bloco de Jordan

m_t	Esperança a posteriori de θ_t
C_t	Variância a posteriori de θ_t
a_t	Esperança da distribuição a priori do sistema $(\theta_t \vert D_{t-1})$
R_t	Variância da distribuição a priori do sistema $(\theta_t \vert D_{t-1})$
f_t	Esperança da distribuição a priori do preditor linear $(\lambda_t \vert D_{t-1})$
Q_t	Variância da distribuição a priori do preditor linear $(\lambda_t \vert D_{t-1})$
f_t^*	Esperança da distribuição a posteriori do preditor linear $(\lambda_t \vert D_t)$
Q_t^*	Variância da distribuição a posteriori do preditor linear$(\lambda_t \vert D_t)$
η_t	Esperança da distribuição a priori da média $(\mu_t \vert D_{t-1})$ (Cap. 3)
Σ_t	Variância da distribuição a priori da média $(\mu_t \vert D_{t-1})$
$\tilde{\mu}$	Esperança a posteriori da média $(\mu_t \vert D_t)$
$\tilde{V}$	Variância a posteriori da média $(\mu_t \vert D_t)$

Siglas

MDL	Modelos Dinâmicos Lineares
MDLG	Modelos Dinâmicos Lineares Generalizados
MDB	Modelos Dinâmicos Beta
MDD	Modelos Dinâmicos Dirichlet
MCMC	Método de simulação estocástica *Markov Chain Monte Carlo*
FFBS	Esquema de amostragem *Forward Filtering Backward Sampling*
CUBS	Esquema de amostragem *Conjugate Updating Backward Sampling*
FD	Fatores de Desconto
SVD	Método de decomposição *Singular Value Decomposition*

Resumo

O interesse central desta dissertação está na modelagem estatística de dados composicionais, que são caracterizados por vetores aleatórios y_t definidos no $(k-1)$-simplex aberto padrão. Cada coordenada de y_t representa a participação (*share*), em termos percentuais, de cada uma das k categorias de resposta possíveis em um fenômeno.

Propomos um modelo dinâmico inédito, batizado como Modelo Dinâmico Dirichlet (MDD), para a descrição de dados composicionais. O MDD é útil tanto no estudo de dados relativos a uma série temporal, quanto no estudo de dados em que não há dinâmica, ou seja, dados estáticos caracterizando unidades amostrais de um estudo. Apresentamos o modelo em duas estruturas distintas, uma delineada para a estimação recursiva dos parâmetros, dita *online*, e a outra para a abordagem de estimação via simulação estocástica MCMC (*offline*), sendo este último método indicado quando há parâmetros desconhecidos na estrutura do modelo.

Discutimos a utilização prática do modelo proposto na descrição do comportamento passado da série histórica, assim como no processo de previsão. Abordamos, ainda, a aplicação do Modelo Dinâmico Dirichlet no contexto estático, um importante caso particular no qual o MDD assume a forma de um modelo de regressão Dirichlet.

Palavras Chave: *modelos dinâmicos, dados composicionais, distribuição Dirichlet, distribuição Beta, distribuição Logistica-Normal, abordagem Bayesiana, séries temporais.*

Abstract

The main purpose of this dissertation is on the study of statistical models for compositional data. Such kind of data is characterized by random vectors y_t defined on the open standard $(k-1)$-simplex. Each coordinate of y_t represents the share, in percentage, of each one of the k categories that represent a given phenomena.

We propose a new dynamic model, the Dynamic Dirichlet Model (MDD), for describing compositional data. The MDD is useful not only in the study of time series of compositional data but also for analyzing static compositional. We designed both online and offline approaches for the estimation of the parameters in the model. The online version is adequate for recursive estimation while the offline one, which is based on stochastic simulation via MCMC, can be used when there are some specific unknown parameters is the model.

We discuss the practical use of the proposed model in describing the past behavior of the series, as well as in the prediction process. We also discuss the application of the Dynamic Dirichlet Model in a static context, an important particular case in which the MDD takes the form of a Dirichlet regression model.

key words: *dynamic models, compositional data, Dirichlet distribution, Beta distribution, Logistic-Normal distribution, Bayesian approach, time series.*

Introdução

O interesse pela classe dos modelos dinâmicos vem crescendo fortemente ao longo dos anos. Tal interesse se justifica pela versatilidade e elegância que tais modelos proporcionam não apenas na descrição da estrutura de correlação temporal, mas também da relação entre variáveis de interesse. A grande flexibilidade dos modelos dinâmicos os colocam em posição de competir com modelos muito difundidos, porém mais restritos e rígidos, como os modelos de regressão e diversos modelos de séries temporais. Os modelos dinâmicos são caracterizados por permitir aos parâmetros do modelo uma evolução temporal estocástica, descrita usualmente por uma estrutura Markoviana.

Os Modelos Dinâmicos Lineares (MDL) (West e Harrison, 1997) fornecem a base para o estudo de fenômenos nos quais a normalidade das observações pode ser assumida. Grande parte dos resultados aplicáveis em tal cenário pode ser naturalmente extrapolada para casos mais gerais. Por essa razão, optamos por dedicar atenção especial a esta importante classe de modelos, antes de iniciar a discussão relativa ao tema central desta dissertação, uma vez que diversos aspectos inerentes a toda a classe de modelos dinâmicos, incluindo a especificação dos componentes do modelo, podem ser mais facilmente introduzidos nesse contexto. No Capítulo 1, além de apresentarmos aspectos teóricos e metodológicos dos modelos dinâmicos lineares, também exibimos códigos *R* que ilustram a questão computacional inerente à utilização prática desses modelos.

Para casos mais gerais, nos quais o pressuposto de normalidade não é adequado, existem os Modelos Dinâmicos Lineares Generalizados (MDLG), introduzidos por West, Harrison e Migon (1985). O processo de estimação torna-se um tanto mais complicado devido ao fato das distribuições de interesse não possuírem, em geral,

formas fechadas. Uma importante sub-classe dos MDLG, quando as observações têm distribuição Beta, está intimamente relacionada aos modelos introduzidos no presente trabalho de pesquisa.

Os Modelos Dinâmicos Beta (MDB) (da-Silva et al., 2011) constituem uma ferramenta estatística útil para lidar com séries temporais univariadas restritas ao intervalo $(0, 1)$. Usualmente, dados dessa natureza, representam proporções, taxas e probabilidades. Baseados nos MDLG e na família de distribuições Beta, os autores desenvolveram uma estratégia de estimação recursiva dos primeiros momentos dos parâmetros a posteriori do modelo. No Capítulo 2 são apresentadas as idéias básicas dos modelos MDLG e MDB.

Nesta dissertação propomos um modelo inédito, batizado como Modelo Dinâmico Dirichlet (MDD), para dados composicionais. Nesse caso, cada observação y_t da série temporal descreve um vetor que está definido no $(k-1)$-simplex aberto padrão expresso por

$$\Delta_{k-1} = \{(y_1, \ldots, y_{k-1}) \in \mathbb{R}^{k-1} | \sum_{i=1}^{k-1} y_i < 1 \, \text{e} \, y_i > 0, \forall i\}.$$

Portanto, com tal modelo é possível descrever a participação (share), em termos percentuais, de cada uma das k categorias de resposta possíveis em um fenômeno.

Um exemplo que ilustra a utilidade do MDD, em problemas ligados ao meio ambiente, está na descrição estatística da evolução do perfil da agricultura nacional. Mais precisamente, a proporção da área plantada destinada à cultura dos seguintes produtos agrícolas: soja, cana-de-açúcar, trigo e outros. Necessariamente, a soma das proporções é igual a um. Desse modo, o interesse principal da análise está na contribuição relativa de cada categoria, e não nos valores absolutos individuais. Embora o MDD seja um modelo de série temporal, um sub-modelo importante é o modelo de regressão Dirichlet, que decorre quando o processo Markoviano latente (estados) é estático.

O MDD foi formulado de acordo com a classe dos Modelos Dinâmicos Lineares Generalizados, o que confere ganhos quanto à flexibilidade na incorporação de efeitos importantes no estudo de alguns fenômenos, quando comparado ao modelo desenvolvido por Grunwald, Raftery e Guttorp (1993). Além disso, o MDD foi descrito de

modo a permitir a estimação de todos os parâmetros desconhecidos do modelo em um paradigma Bayesiano. Este trabalho representa uma extensão ao modelo desenvolvido por da-Silva et al. (2011).

O modelo de Grunwald, Raftery e Guttorp (1993), GRG, é menos flexível do que o MDD pois (1) no GRG, a priori utilizada para o vetor de proporções é baseada em argumentos muito restritivos com respeito à forma da mesma. No MDD fazemos uso de prioris bastante flexíveis, que impõem o mínimo de condições. (2) no GRG utilizam-se procedimentos inferenciais que incluem tanto métodos Clássicos quanto Bayesianos. No MDD utilizamos uma abordagem genuinamente Bayesiana; (3) No GRG a especificação dos efeitos nos parâmetros do processo latente θ_t, tais como tendência, crescimento e sazonalidade, além da inclusão de covariáveis, é bastante complicada. No MDD esse tipo de especificação é feita de forma muito simples e eficiente, mediante as matrizes de planejamento F e G do modelo.

Esta dissertação está organizada como a seguir: no Capítulo 1 revisamos os Modelos Dinâmicos Lineares (MDL). No Capítulo 2 são descritos os Modelos Dinâmicos Lineares Generalizados (MDLG) e os Modelos Dinâmicos Beta (MDB). No Capítulo 3 introduzimos o Modelo Dinâmico Dirichlet (MDD). Nesse Capítulo descrevemos, primeiramente, o modelo para a estimação recursiva, dita *online*. Em seguida, apresentamos uma formulação alternativa, desenvolvida para a estimação *offline* dos parâmetros desconhecidos via cadeias de Markov. Ainda são tratados o problema de estimação na presença de observações faltantes. No Capítulo 4 (Aplicações) discutimos seis exemplos, entre casos particulares e ajustes a dados reais. Em todos os Capítulos exibimos alguns esquemas das programações utilizadas, no intuito de permitir ao leitor uma visualização mais ampla do processo. No Capítulo 5 fazemos algumas discussões adicionais e trabalhos futuros.

Capítulo 1

Modelos Dinâmicos Lineares

Os modelos dinâmicos lineares representam um caso particular do que foi batizado na literatura estatística como modelos dinâmicos. Ainda que o tema central desta dissertação não envolva a suposição de normalidade ou linearidade, entendemos que diversas características de interesse podem ser mais bem visualizadas a partir do estudo do caso normal, no qual os resultados são sempre mais simples. Nesse sentido, optou-se por explorar, nesse capítulo, os principais aspectos dos modelos dinâmicos lineares a fim de descrever os fundamentos básicos dessa classe de modelos. Sempre que pertinente abordaremos o uso do *software* estatístico livre *R* (R Development Core Team, 2011), em especial o pacote *dlm* (Petris, 2010), nas aplicações dos modelos dinâmicos lineares.

1.1 Motivação

O processo de avaliação e monitoramento contínuo, muito presente na gestão de empresas, em estudos econômicos, em projetos de pesquisa, dentre outros, envolve a quantificação de certos atributos do sistema investigado. Em linguagem estatística, tal procedimento culmina, usualmente, na geração de dados com estrutura temporal que exigem tratamento adequado.

Não seria exagero dizer que, nesse contexto, os sistemas de interesse são, em sua maioria, dinâmicos. Ou seja, suas principais características e comportamentos sofrem modificações ao longo do tempo. Dizemos, nesses casos, que o sistema *evolui*.

Todo o esforço resultante do processo de coleta e análise dos dados justifica-se pela possibilidade de elucidar, de forma apropriada, dúvidas cruciais relacionadas ao cenário em estudo. Entre os interesses possíveis, alguns têm posição de destaque:

- Prever valores futuros;
- Diagnosticar a natureza e magnitude da relação entre variáveis;
- Estimar o efeito de cada fator (sazonalidade, por exemplo) em um resultado observado;
- Entender, de forma completa e eficiente, a trajetória do sistema.

É nesse ambiente investigativo que surgem, de modo bastante natural, as noções de modelagem. De maneira simples, pode-se dizer que um modelo é qualquer esquema descritivo, explicativo, que organiza as informações disponíveis, fornecendo, portanto, meios para a aprendizagem e previsão. Dessa forma, fica evidente o fato de que os modelos não representam a "verdade", e sim tratam-se de mecanismos poderosos capazes de simplificar a compreensão de sistemas complexos, facilitando a rotina de processamento de informações. Dessa forma, um modelo possibilita descrever, de maneira sistemática, os dados e as experiências do investigador e, sendo assim, é sempre uma representação subjetiva baseada no conhecimento passado e presente.

Em situações teóricas e aplicadas, busca-se um modelo parcimonioso. Além disso, este deve ser suficientemente robusto para que reformulações estruturais do mesmo sejam excepcionais. Há dois casos principais em que tais reformulações se fazem necessárias. O primeiro caso ocorre quando especialistas antecipam mudanças drásticas no cenário em estudo. O segundo caso acontece quando, no monitoramento do desempenho preditivo do modelo, detectam-se deficiências que apontam para uma possível falta de adequabilidade ou ajuste do modelo corrente.

Os modelos dinâmicos, estudados ao longo desta dissertação, descrevem uma classe de modelos que pode tanto incluir modelos do tipo regressão linear ou não linear, quanto os principais modelos de séries temporais, aliando sofisticação e flexibilidade na descrição de inúmeros fenômenos. A característica mais marcante desses modelos é permitir que os parâmetros associados ao sistema evoluam de acordo com alguma estrutura probabilística, fornecendo meios para tratar, adequadamente, o

caráter dinâmico do fenômeno em estudo. Este capítulo restringe-se à subclasse conhecida como modelos dinâmicos lineares (MDL).

1.2 Formulação do modelo

Considere uma série temporal Y_t, definida nos tempos $t = 1, 2, \ldots$, onde cada vetor aleatório Y_t tem dimensão $(k \times 1)$. O Modelo Dinâmico Linear (MDL), também conhecido como Modelo Linear de Espaço de Estados Gaussiano (West e Harrison, 1997), é especificado por duas equações. Uma para a série temporal observada, denominada de *equação das observações*, e outra para descrever a evolução temporal dos estados latentes θ_t, denominada *equação do sistema*. As equações são apresentadas a seguir.

Equação das observações:

$$Y_t = F_t'\theta_t + \nu_t, \quad \nu_t \sim N[0, V_t] \; ; \tag{1.1a}$$

Equação do sistema:

$$\theta_t = G_t\theta_{t-1} + \omega_t, \quad \omega_t \sim N[0, W_t] \; . \tag{1.1b}$$

A cada instante de tempo t, o modelo é caracterizado pela quadrupla $\{F_t, G_t, V_t, W_t\}$, onde:

- F_t é uma matriz conhecida $(n \times k)$;
- G_t é uma matriz conhecida $(n \times n)$;
- V_t é uma matriz de covariâncias conhecida $(k \times k)$;
- W_t é uma matriz de covariâncias conhecida $(n \times n)$.

O modelo relaciona Y_t ao vetor de parâmetros do sistema θ_t $(n \times 1)$, além de especificar a evolução temporal dos $\theta' s$. Isso se dá através das distribuições definidas sequencialmente

$$(Y_t|\theta_t) \sim N_k[F_t'\theta_t, V_t] \; ; \tag{1.2a}$$

$$(\theta_t|\theta_{t-1}) \sim N_n[G_t\theta_{t-1}, W_t] \; . \tag{1.2b}$$

A equação das observações define a distribuição amostral da série Y_t como função do estado latente θ_t. Observe que F_t está relacionada à matriz dos co-fatores (matriz de delineamento de uma regressão), cuja função é relacionar, a cada instante, o estado do sistema à média das observações.

A equação de evolução, ou equação do sistema, dada em (1.1b), implica em uma evolução *markoviana* para o sistema. Ela é definida por uma transformação linear $G_t\theta_{t-1}$ acrescida de um erro aleatório com média zero.

O conjunto de informações disponíveis no instante t, para $t = 1, 2, \ldots$, é denotado por D_t e expresso como $D_t = \{I_t, D_{t-1}\}$, onde I_t representa toda a informação adicional relevante obtida no tempo t. No caso em que a série histórica Y_t é a única fonte de informação, o sistema é dito *"fechado"* e $D_t = \{Y_t, D_{t-1}\}$. D_0 representa o conjunto contendo as informações iniciais, normalmente condensadas nos parâmetros da distribuição de θ_0, e as matrizes que fazem parte da especificação dos modelos. Dessa forma, as equações (1.2) são implicitamente condicionais a D_{t-1}. Ao longo de toda a dissertação consideraremos, apenas, sistemas fechados. Dessa forma, algumas afirmações podem não ser necessariamente extensivas a sistemas abertos.

Assume-se que $(\theta_0|D_0) \sim N_n(m_0, C_0)$ e que as sequências de erros ν_t e ω_t são internamente e mutuamente independentes. A estrutura de dependência condicional do MDL é apresentada na Figura 1.1. Nela é possível observar que, dado o presente, o passado e o futuro são independentes.

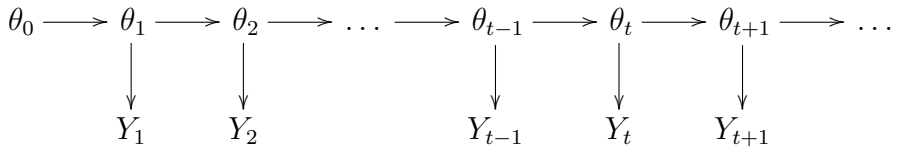

Figura 1.1: Estrutura de dependência para o espaço de estados do modelo

O fluxo apresentado na Figura 1.1 é resultado da própria definição do modelo. Assim como qualquer outro modelo de espaço de estados, os MDLs satisfazem duas propriedades fundamentais.

(A.1) (θ_t) é uma cadeia de Markov.

(A.2) Condicionalmente a (θ_t), os Y_ts são independentes e Y_t depende de θ_t apenas.

Uma consequência das propriedades (A.1) e (A.2) é podermos escrever a distribuição conjunta dos estados do sistema θ_t e das observações y_t em forma de um produto de distribuições condicionais. Isto é, para $t \geq 1$,

$$p(\theta_0, \theta_1, \ldots, \theta_t, y_1, \ldots, y_t) = p(\theta_0) \prod_{j=1}^{t} p(\theta_j|\theta_{j-1}) p(y_j|\theta_j). \tag{1.3}$$

Como poderá ser visto, tal decomposição é de grande valia no tratamento dos modelos dinâmicos.

Se as matrizes F_t e G_t são constantes para todo t, então dizemos que o modelo está na classe dos Modelos Dinâmicos Lineares de Séries Temporais (MDLST), um importante subconjunto dos MDL. De acordo com West e Harrison (1997), os MDLST incluem, essencialmente, todos os modelos lineares clássicos de séries temporais, inclusive os modelos ARIMA.

1.3 Estimação do sistema e previsão

No estudo de dados coletados ao longo do tempo, o principal interesse está em fazer inferências sobre o passado, o presente e o futuro. Em linguagem estatística, busca-se obter as distribuições condicionais do sistema $(\theta_s|D_t)$ para $s < t$ (*suavização*), $s = t$ (*filtragem*) e $s > t$ (*previsão do sistema*), respectivamente. Essa nomenclatura foi adotada em alinhamento a Petris et al. (2009). West e Harrison (1997) também denotam por *filtragem* a inferência para tempos passados, quando $s < t$. Quanto ao interesse relativo ao futuro, há também o que talvez seja a principal função dos modelos estatísticos de série temporais: prever os valores futuros para a série observável Y_t. Nesta seção abordaremos cada um dos casos citados no contexto dos MDL.

1.3.1 Filtragem

Considere agora um MDL em que Y_t é escalar, ou seja, um modelo dinâmico linear univariado. Suponha que o modelo é fechado para informações externas nos tempos $t \geq 1$. Sob essas condições, tem-se a seguinte estrutura inferencial descrita pelas equações de atualização (West e Harrison, 1997):

1. **Distribuição a posteriori no tempo $t-1$:**

   Para um vetor de médias m_{t-1} e uma matriz de covariância C_{t-1}, ambos conhecidos,

$$(\theta_{t-1}|D_{t-1}) \sim N_n[m_{t-1}, C_{t-1}] \; . \tag{1.4}$$

2. **Distribuição a priori no tempo $t-1$:**

$$\begin{gathered}(\theta_t|D_{t-1}) \sim N_n[a_t, R_t] \; , \\ \text{onde} \quad a_t = G_t m_{t-1} \quad \text{e} \quad R_t = G_t C_{t-1} G_t' + W_t \; .\end{gathered} \tag{1.5}$$

3. **Distribuição preditiva um passo à frente**:

$$\begin{gathered}(Y_t|D_{t-1}) \sim N[f_t, Q_t] \; , \\ \text{onde} \quad f_t = F_t' a_t \quad \text{e} \quad Q_t = F_t' R_t F_t + V_t \; .\end{gathered} \tag{1.6}$$

4. **Distribuição a posteriori no tempo t:**

$$\begin{gathered}(\theta_t|D_t) \sim N_n[m_t, C_t] \; , \\ \text{onde} \quad m_t = a_t + A_t e_t, \quad C_t = R_t - A_t Q_t A_t' \; ; \\ A_t = R_t F_t Q_t^{-1} \quad \text{e} \quad e_t = Y_t - f_t \; .\end{gathered} \tag{1.7}$$

As equações apresentadas fornecem o procedimento recursivo na estimação dos parâmetros do modelo. As relações (1.7) de atualização da distribuição a posteriori de θ_t são conhecidas como *Filtro de Kalman*. A prova dos quatro itens apresentados é uma imediata aplicação das propriedades da distribuição Normal multivariada. Observe que a distribuição conjunta dos elementos envolvidos é dada por:

$$\begin{pmatrix} \theta_t \\ Y_t \end{pmatrix} \Bigg| \; D_{t-1} \sim \mathrm{N}\left[\begin{pmatrix} a_t \\ f_t \end{pmatrix}, \quad \begin{pmatrix} R_t & R_t F_t \\ F_t' R_t & Q_t \end{pmatrix} \right]$$

Dessa maneira, pelas propriedades condicionais da distribuição normal, obtêm-se as equações de atualização dadas pelo *Filtro de Kalman*.

A abordagem inferencial em questão envolve três etapas: *evolução*, *previsão* e *atualização*, conforme mostra a Figura 1.2. O vetor de parâmetros θ_{t-1}, que é não observável, representa o estado corrente do sistema. Toda a informação disponível até o momento é sumarizada em seus primeiros momentos m_{t-1} e C_{t-1}. Assim, fica evidente a abordagem *Bayesiana*, em que a incerteza em relação ao parâmetro de interesse é descrita por meio de uma distribuição de probabilidade.

$$\ldots\ (\theta_{t-1}|D_{t-1}) \xRightarrow{\text{Evolução}} (\theta_t|D_{t-1}) \xRightarrow{\text{Atualização}} (\theta_t|D_t)\ \ldots$$
$$\Big\Downarrow \text{Previsão}$$
$$(Y_t|D_{t-1})$$

Figura 1.2: Análise sequencial do modelo linear dinâmico normal

A distribuição a priori dada em (1.4), em conjunto com a equação de evolução (1.1b), define a etapa de evolução, isto é:

$$p(\theta_t|D_{t-1}) = \int p(\theta_t, \theta_{t-1}|D_{t-1})d\theta_{t-1} = \int p(\theta_t|\theta_{t-1})p(\theta_{t-1}|D_{t-1})d\theta_{t-1}.$$

Consequentemente, a distribuição a priori de θ_t (1.5) é conhecida e permite prever o efeito da passagem do tempo, entre $t-1$ e t, no sistema. Com base nela e na equação das observações (1.1a), a próxima observação Y_t da série é prevista (1.6), configurando o passo de previsão:

$$p(Y_t|D_{t-1}) = \int p(Y_t|\theta_t)p(\theta_t|D_{t-1})d\theta_t.$$

Finalmente, uma vez observado o valor Y_t, passa-se à etapa de atualização da distribuição de θ_t com base no Teorema de Bayes:

$$p(\theta_t|D_t) = \frac{p(Y_t|\theta_t)p(\theta_t|D_{t-1})}{\int p(Y_t|\theta_t)p(\theta_t|D_{t-1})d\theta_t}.$$

A média a posteriori, já apresentada, é dada por $m_t = a_t + A_t e_t$ (1.7). Sendo assim, o conhecimento do analista sobre o estado do sistema é corrigido/atualizado em função do erro de previsão e das incertezas associadas ao modelo R_t e Q_t. A matriz

A_t pode ser interpretada como um peso adaptativo dado à observação Y_t. A lógica descrita torna claro o processo de aprendizagem sistemática, predição e processamento de novas informações citado anteriormente.

Caso a observação y_t seja faltante/indisponível, ela não contribuirá com qualquer informação adicional. Em notação estatística, esse caso pode ser expresso como:

$$p(\theta_t|D_t) = p(\theta_t|D_{t-1}).$$

Esse resultado significa que informações faltantes são facilmente tratadas nos MDL. Neste caso, a distribuição filtrada no tempo t é justamente a distribuição preditiva um passo a frente no tempo $t-1$.

Quanto aos aspectos computacionais do processo de estimação do MDL, a função R *dlmFilter* do pacote *dlm* (Petris, 2010) aplica o *Filtro de Kalman* a uma série observada e um modelo especificado, e retorna os valores de m_t, C_t, a_t, R_t e f_t. As matrizes de covariância são retornadas em termos de suas respectivas decomposições em valores singulares, em inglês SVD.

1.3.2 Suavização

Em diversas ocasiões, revisar inferências relativas a tempos passados, valendo-se de toda a informação disponível D_T, pode ser extremamente útil para elucidar o entendimento sobre o que, de fato, ocorreu.

Pode-se pensar, por exemplo, em uma vara judicial que, anualmente, recebe e sentencia, na figura de um Juiz de Direito, centenas de processos judiciais. Naturalmente, ao longo do ano, várias mudanças administrativas podem ocorrer, como a alteração do horário de atendimento ao público ou do tamanho do quadro de pessoal. Neste caso, o uso de um MDL poderia ser de grande valia para analisar, retrospectivamente, o impacto de cada uma dessas mudanças de cenário na produtividade da vara. Em última instância, o uso dessa ferramenta estatística pode ser crucial no que se refere à tomada de decisão, facilitando a otimização dos recursos públicos destinados a tal atividade.

Para o Modelo Dinâmico Linear apresentado na Seção 1.3.1, as distribuições retrospectivas são todas Gaussianas, com média e variância facilmente calculáveis re-

cursivamente. Pode-se mostrar que, para $t = 0, \ldots, T-1$, $(\theta_t | D_T) \sim N(s_t, S_t)$, onde

$$s_t = m_t + C_t G'_{t+1} R^{-1}_{t+1}(s_{t+1} - a_{t+1}) \; ; \tag{1.8a}$$

$$S_t = C_t - C_t G'_{t+1} R^{-1}_{t+1}(R_{t+1} - S_{t+1}) R^{-1}_{t+1} G_{t+1} C_t \; , \tag{1.8b}$$

com valores iniciais $s_T = m_T$ e $S_T = C_T$.

A função R *dlmSmooth* do pacote *dlm* (Petris, 2010) aplica as equações dadas em (1.8) a uma série observada e um modelo especificado, e retorna os valores de s_t e S_t. As matrizes S_t são retornadas também na forma decomposta em valores singulares. É também possível usar a função *dlmSmooth* para calcular os momentos suavizados a partir de um objeto da classe *dlmFiltered*, que normalmente corresponde ao resultado de uma execução anterior da função *dlmFilter*.

1.3.3 Previsão

A atividade de previsão é parte inevitável do cotidiano de cada indivíduo. Antecipar, em certo sentido, cenários e eventos futuros é indispensável para a tomada eficiente de decisões. Usualmente, as informações, numéricas ou não, são subjetivamente processadas, resultando na consolidação de uma expectativa possivelmente pertinente sobre o futuro. Ocorre que a previsão de valores futuros de uma variável qualquer, baseada no referido mecanismo de processamento, torna-se pouco confiável e eficiente em situações complexas ou quando se trata de grandes decisões. Isso se deve à dificuldade em subjetivamente se fazer uso de toda a informação disponível. Muitas vezes o analista, no intuito de avaliar o comportamento da série histórica, compara o resultado do mês corrente com apenas o do mês anterior e com o resultado do mesmo mês do ano anterior. Ou seja, pela dificuldade mencionada, tal analista contenta-se em valer-se, apenas, de fragmentos dos dados disponíveis. É fácil entender, por exemplo, que o aumento do preço da carne bovina provoca um maior consumo da carne de frango. No entanto, é difícil determinar subjetivamente em que medida isto ocorre.

A título de ilustração, suponha que uma montadora de veículos, em seu processo de planejamento, necessita decidir quantos automóveis precisará produzir no próximo

ano, em função da expectativa sobre a demanda futura do mercado. Caso a estimativa da demanda futura, feita pela montadora, seja maior do que a demanda real, a produção em excesso poderá acarretar grande prejuízo financeiro. Por outro lado, se a produção for inferior à demanda latente, também haverá prejuízo em função da perda de maior inserção no mercado.

Resumindo, a precisão e acurácia das estimativas, consequente do método de previsão utilizado, pode ditar cenários de decisão e suas implicações nos negócios. Os modelos estatísticos, em especial os modelos dinâmicos, proporcionam, quando bem empregados, um grande ganho de qualidade na tomada de decisões.

No tocante às predições obtidas com os Modelos Dinâmicos Lineares, pode ser mostrado que as distribuições condicionais de θ_{t+m} e Y_{t+m} dado D_t, para $m \geq 1$, são normais e seus momentos podem ser calculados recursivamente.

Sejam $a_t(0) = m_t$ e $R_t(0) = C_t$, então, para $m \geq 1$, valem as seguintes relações:

- $(\theta_{t+m}|D_t) \sim N(a_t(m), R_t(m))$, onde

$$a_t(m) = G_{t+m}a_t(m-1) \tag{1.9a}$$

$$R_t(m) = G_{t+m}R_t(m-1)G'_{t+m} + W_{t+m} \tag{1.9b}$$

- $(Y_{t+m}|D_t) \sim N(f_t(m), Q_t(m))$, onde

$$f_t(m) = F_{t+m}a_t(m) \tag{1.10a}$$

$$Q_t(m) = F_{t+m}R_t(m)F'_{t+m} + V_t \tag{1.10b}$$

A função $f_t(m)$ dada em (1.10a) é conhecida como *função de previsão* e desempenha um papel de destaque na especificação dos MLD, como será visto na próxima seção. É justamente ela quem define a forma da curva das previsões pontuais da série $Y_{t+1}, \ldots, Y_{t+m}$.

A função R *dlmForecast* (Petris, 2010) calcula os momentos dados em (1.9) e (1.10) para uma série observada e um modelo especificado. O usuário define quantos passos a frente se quer prever. É possível, ainda, usar essa mesma função para gerar N amostras do sistema e da série $Y_{t+1}, \ldots, Y_{t+m}$.

1.4 Especificação do modelo

Esta seção é dedicada à apresentação de alguns modelos especiais da classe dos MDL. Estamos particularmente interessados nas formas de especificação das matrizes F e G e nas consequentes funções de previsão. Serão discutidos, nessa ordem, os modelos polinomiais, sazonais e de regressão. Por fim, introduzimos a noção de sobreposição do MDL. Aspectos computacionais são abordados e alguns códigos R ilustrativos disponibilizados ao longo do texto.

1.4.1 Modelos dinâmicos lineares polinomiais

Os modelos dinâmicos lineares polinomiais representam uma subclasse dos MDL que é extensivamente usada para descrever tendências, em especial aquelas que evoluem suavemente no tempo.

Em previsões de curto ou médio prazo, modelos polinomiais de baixa ordem, digamos até 3 (correspondente a funções de crescimento quadrático), costumam resultar em boas aproximações. Este fato não surpreende, haja vista o uso frequente, na literatura estatística, de aproximações baseadas nos 3 primeiros termos da expansão em série de Taylor de diversas funções.

Um MDL polinomial de ordem n é definido como sendo qualquer MDL com matrizes F e G constantes e função de previsão, para $t \geq 0$, da forma

$$f_t(m) = E(Y_{t+m}|D_t) = a_{t0} + a_{t1}m + \ldots + a_{t,n-1}m^{n-1}, \quad m \geq 0,$$

onde os coeficientes $a_{t0}, \ldots, a_{t,n-1}$ são funções lineares da média a posteriori m_t, independentes de m.

Não há uma forma única de se especificar as matrizes F e G de modo a obter um modelo polinomial. Aqui, restringiremos nossa atenção aos modelos canônicos. A definição formal de modelos canônicos e a descrição de formas alternativas de especificação de MDL polinomiais são apresentadas, respectivamente, nos capítulos 5 e 7 de West e Harrison (1997).

Pode-se provar que um MDL, definido como na Seção (1.2), com matrizes $F = E_n$

e $G = J_n(1)$, onde

$$E_n' = (1, 0, \ldots, 0)' \quad \text{e} \quad J_n(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda & 1 & 0 & \ldots & 0 \\ 0 & 0 & \lambda & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & 0 & \ldots & \lambda \end{pmatrix},$$

é um modelo dinâmico linear polinomial. Cabe destacar que a matriz $J_n(\lambda)$ é conhecida como *bloco de Jordan*.

Com o objetivo de ilustrar os modelos em consideração, assim como os princípios de filtragem, suavização e previsão discutidos na Seção (1.3), lançamos mão do pacote *dlm* do R. Para tanto, usamos um modelo polinomial de ordem 2 definido pela quadrupla:

$$\left\{ F = E_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, G = J_2(1) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, W = \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix}, V = 1000 \right\}$$

Este modelo pode ser alternativamente escrito em função das equações das observações e do sistema. São elas:

$$\begin{aligned} &\text{Equação das observações:} && Y_t = \mu_t + \nu_t, \\ &\text{Equações do sistema:} && \mu_t = \mu_{t-1} + \beta_{t-1} + \omega_{t1}, \\ & && \beta_t = \beta_{t-1} + \omega_{t2} \\ &\text{Informação inicial:} && (\theta_0 | D_0) \sim N_n(m_0, C_0), \end{aligned}$$

onde

$$\theta_t = (\mu_t, \beta_t)', \quad \omega_t = (\omega_{t1}, \omega_{t2})' \sim N(0, W),$$
$$\nu_t \sim N(0, V), \quad m_0 = (0, 0)' \quad \text{e} \quad C_0 = 10^7 \, \mathbf{I}_2$$

Seja $E(\theta_t | D_t) = m_t = (a_t, b_t)$, a função de previsão deste modelo é dada por $f_t(m) = a_t + m b_t$. Ou seja, como deveria ser, uma reta definida em função da média a posteriori do sistema.

As funções *R dlmModPoly* e *dlmForecast* (Petris, 2010) foram usadas, respectivamente, para definir o modelo e dele simular uma possível observação. Em seguida, utilizou-se a função *dlmFilter* para estimar o nível μ_t, para $t = 1, \ldots, 60$, e prever, um passo a frente, valores futuros para a série Y_t. Ressaltamos que foi escolhida uma priori vaga $(\theta_0|D_0) \sim N_n(m_0, C_0)$ no processo de estimação. Por fim, os valores suavizados para o nível foram calculados pela função *dlmSmooth* (Petris, 2010).

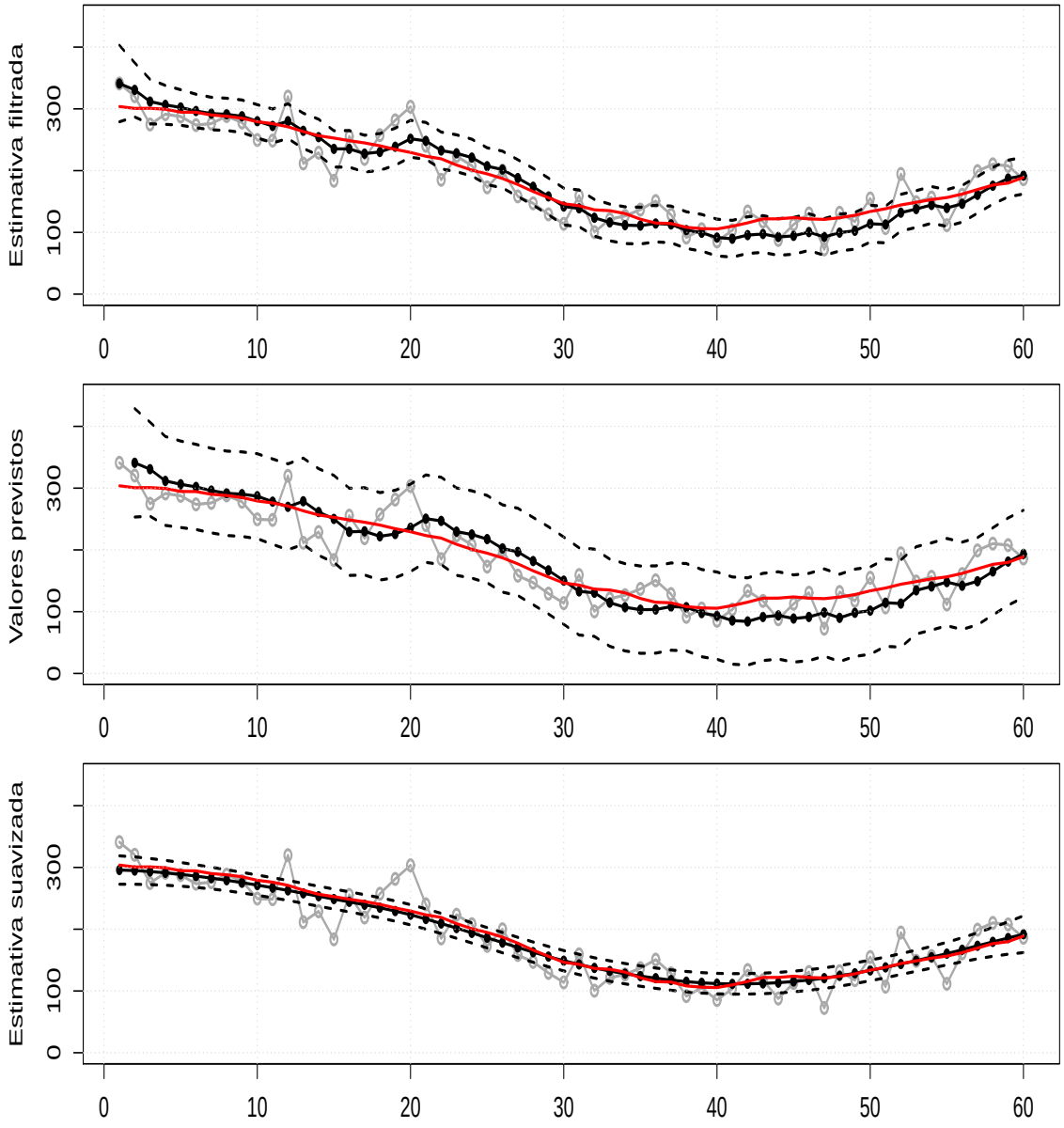

Figura 1.3: Serie histórica y_t (linhas cinzas), média real μ_t (linhas vermelhas) e estimativas (linhas pretas) com respectivos intervalos de credibilidade de 95% (linhas pretas tracejadas). (a) Valores filtrados para μ_t; (b) Valores previstos um passo a frente para y_t; (c) Valores suavizados para μ_t.

A Figura 1.3 sobrepõe a série simulada Y_t (linhas cinzas) e as estimativas filtradas e suavizadas para a média μ_t, além de comparar as previsões um passo a frente e os valores Y_t efetivamente observados. É possível notar que as estimativas, em todos os casos, são muito pertinentes. Os intervalos de credibilidade de 95% para a previsão (linhas pretas tracejadas) parecem acomodar as observações Y_t sem folga ou rigidez. Da mesma forma, as estimativas suavizadas para a média são bem mais estreitas e refletem, majoritariamente, a variabilidade de μ_t, que pouco são afetados pelos ruídos aleatórios ν_t, especialmente aqueles registrados entre os tempos 10 e 20. Em resumo, o método inferencial consegue recuperar, no sentido estatístico da palavra, os valores reais dos parâmetros e detectar as alterações (não estruturais) do sistema.

A programação utilizada (1.1) é apresentada a seguir e pode ser útil para, a partir da replicação com diferentes parâmetros, ilustrar o papel de cada elemento no comportamento da série. A programação relativa à produção do gráfico foi omitido e envolve somente os objetos R situados nas 3 últimas linhas de cada um dos 4 blocos do Código 1.1.

Código 1.1: Modelos dinâmicos lineares polinomiais

```
> # Bloco 1: Geração dos dados Y_t
> N <- 60
> mod <- dlmModPoly(order = 2, dW = c(5, 1),
+   dV = 1000, m0 = rnorm(2, c(300, 0)), C0 = diag(1, 2))
> aux <- dlmForecast(mod, nAhead = N, sampleNew = 1)
> y <- as.ts(aux$newObs[[1]])
> theta.real <- as.ts(aux$newStates[[1]])
>
> # Bloco 2: Filtragem
> mod$m0 <- rep(0, 2)
> mod$C0 <- 1e+7
> filtro <- dlmFilter(y, mod)
> C <- dlmSvd2var(filtro$U.C, filtro$D.C)
> c1 <- sapply(2:(N + 1), function(x) sqrt(C[[x]][1, 1]))
> z <- qnorm(0.975)
> mu.filtro <- dropFirst(filtro$m[, 1])
> lim.inf.C <- mu.filtro - z * c1
> lim.sup.C <- mu.filtro + z * c1
```

```
19 >
   > # Bloco 3: Previsão um passo a frente
21 > raiz.Q <- dropFirst(residuals(filtro)$sd)
   > y.previsto <- dropFirst(filtro$f)[, 1]
23 > lim.inf.Q <- y.previsto - z * raiz.Q
   > lim.sup.Q <- y.previsto + z * raiz.Q
25 >
   > # Bloco 4: Suavização
27 > suave <- dlmSmooth(filtro)
   > S <- dlmSvd2var(suave$U.S, suave$D.S)
29 > s1 <- sapply(2:(N + 1), function(x) sqrt(S[[x]][1, 1]))
   > mu.suave <- dropFirst(suave$s[, 1])
31 > lim.inf.S <- mu.suave - z * s1
   > lim.sup.S <- mu.suave + z * s1
```

1.4.2 Modelos dinâmicos lineares sazonais

Diversos fenômenos são caracterizados por apresentar comportamentos cíclicos, periódicos. O preço médio do etanol depende, dentre outros fatores, da oferta corrente de cana-de-açúcar, que, por sua vez, depende do momento na sequência cíclica de sua produção: plantio, cultivo e colheita. De maneira similar, o congestionamento nos aeroportos também apresenta períodos bem definidos. Neste caso, isso se deve ao calendário escolar, ao dia da semana, aos feriados nacionais, etc.

Em ambos os casos citados, é indispensável construir um modelo que leve em consideração esse comportamento sistemático. Há duas formas usuais para descrever sazonalidade no contexto dos MDL. A primeira abordagem, representada pelos *modelos sazonais de forma livre*, pode ser entendida como um caso particular dos modelos de regressão discutidos na Seção (1.4.3) e não será detalhada neste trabalho. Mais uma vez, sugerimos aos interessados a leitura de West e Harrison (1997). A segunda abordagem da representação da sazonalidade será descrita a seguir e está baseada na representação em *Séries de Fourier*.

Seja $g(t)$ qualquer função real definida nos inteiros $t = 0, 1, \ldots$. Dizemos que $g(t)$ é cíclica se para algum $p > 1$ e para todo t, $n \geq 0$, $g(t + np) = g(t)$. Assim, qualquer função periódica discreta assume valores somente no conjunto $\psi = \{\psi_1, \ldots, \psi_p\}$, onde

$\psi_i = g(i)$.

A idéia central na construção desses modelos é escrever o vetor ψ como combinação linear de funções trigonométricas, necessariamente periódicas. No campo da álgebra linear, diríamos que se trata de uma mudança de base do espaço vetorial.

Usando identidades trigonométricas, pode-se provar que quaisquer p números reais $\psi_1, \ldots, \psi_p$ podem ser representados por

$$\psi_j = a_0 + \sum_{r=1}^{h} [a_r \cos(\alpha r j) + b_r \text{sen}(\alpha r j)], \tag{1.11}$$

onde $\alpha = 2\pi/p$ e h é o maior inteiro não superior a $p/2$.

As quantidades a_r e b_r são funções conhecidas de ψ, referenciadas na literatura como *coeficientes de Fourier*. Usualmente, a média da série é modelada separadamente, resultando em $a_0 = 0$. Neste caso, a equação (1.11) pode ser escrita como $\psi_j = \sum_{r=1}^{h} S_r(j)$, onde

$$S_r(\cdot) = a_r \cos(\alpha r \cdot) + b_r \text{sen}(\alpha r \cdot) = A_r \cos(\alpha r \cdot + \gamma_r),$$
$$A_r = (a_r^2 + b_r^2)^{1/2} \quad \text{e} \quad \gamma_r = \arctan(-b_r/a_r).$$

O termo $S_r(\cdot)$ é chamado de r-ésimo harmônico. A_r, αr e γ_r representam, respectivamente, a *amplitude*, a *frequência* e a *fase* de $S_r(\cdot)$.

Enquanto os modelos sazonais de forma livre definem o sistema diretamente pelo vetor ψ, permutando seus elementos com repetidas aplicações da matriz de evolução G, os modelos sazonais em forma de Fourier codificam o sistema em termos dos harmônicos e seus conjugados.

Na perspectiva dos modelos dinâmicos, é preciso ser capaz de descrever cada $S_r(t+1)$ em função de $S_r(t)$. Por outro lado, o simples conhecimento de $S_r(t)$ não é suficiente para determinar o valor do harmônico no tempo seguinte. Entretanto, Petris et al. (2009) mostram que os valores do r-ésimo harmônico $S_r(t)$ e seu respectivo conjugado $S_r^*(t)$ conjuntamente determinam os valores de $S_r(t+1)$ e $S_r^*(t+1)$. Ou seja, artificialmente se cria um parâmetro adicional, $S_r^*(t)$, ao sistema com a finalidade de viabilizar a descrição da evolução de cada harmônico na estrutura dos modelos dinâmicos.

Na medida em que a frequência de cada harmônico cresce em r, a sazonalidade é expressa como a soma de funções periódicas que *vão* da mais suave para a mais brusca.

Quando o período p é par, $S_{p/2}(t+1) = -S_{p/2}(t)$, de maneira que o último harmônico simplesmente muda de sinal a cada passagem do tempo. Sendo assim, a inclusão de seu conjugado no vetor de parâmetros do sistema traz redundância ao modelo. Por esse motivo, a paridade de p é importante na especificação das matrizes F e G.

Sejam

$$J_2(1,\omega) = \lambda \begin{pmatrix} \cos(\omega) & \text{sen}(\omega) \\ -\text{sen}(\omega) & \cos(\omega) \end{pmatrix},$$

$$\mathcal{F}_{impar} = \begin{pmatrix} E_2 \\ E_2 \\ \vdots \\ E_2 \end{pmatrix}, \quad \mathcal{G}_{impar} = \begin{pmatrix} J_2(1,\omega) & 0 & \dots & 0 \\ 0 & J_2(1,2\omega) & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & J_2(1,h\omega) \end{pmatrix},$$

$$\mathcal{F}_{par} = \begin{pmatrix} E_2 \\ E_2 \\ \vdots \\ E_2 \\ 1 \end{pmatrix}, \quad \mathcal{G}_{par} = \begin{pmatrix} J_2(1,\omega) & 0 & \dots & 0 & 0 \\ 0 & J_2(1,2\omega) & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & J_2(1,(p/2-1)\omega) & 0 \\ 0 & 0 & \dots & 0 & -1 \end{pmatrix},$$

um MDL sazonal na forma de Fourier é, finalmente, definido como sendo qualquer MDL da forma

$$\{\mathcal{F}_{impar}, \mathcal{G}_{impar}, \cdot, \cdot\}, \text{ se } p \text{ é impar},$$
$$\{\mathcal{F}_{par}, \mathcal{G}_{par}, \cdot, \cdot\}, \text{ se } p \text{ é par}.$$

A estrutura em blocos da matriz G está associada à *sobreposição* de MDLs tratada na Seção (1.4.4). Os harmônicos são combinados em um mesmo modelo, mas como será visto adiante, há a possibilidade de se usar apenas alguns deles na des-

crição do efeito sazonal. Tal flexibilidade é um dos pontos fortes dessa abordagem e, quando aplicável, pode resultar em ganhos no desempenho preditivo, além da evidente simplificação do problema. Outra vantagem significativa na comparação com os modelos sazonais de forma livre está no fato de que aqui não há necessidade de se criar restrições aos parâmetros. Enquanto na abordagem de Fourier os efeitos sazonais automaticamente somam zero, nos modelos sazonais de forma livre esta restrição precisa ser imposta, de modo a tornar o modelo identificável.

Até aqui, para facilitar o entendimento, tratamos o caso em que os harmônicos evoluem deterministicamente no tempo. Por outro lado, o analista pode especificar uma matriz W diferente de zero para incorporar a natureza dinâmica da sazonalidade. Neste caso, é natural pensar um W como uma matriz também bloco-diagonal. Ressaltamos que a adição dos erros não-nulos ω_t faz com que os fatores sazonais não mais sejam, de fato, periódicos. De qualquer maneira, a função de previsão m-passos a frente, independentemente da formulação estática ou dinâmica, é periódica e depende unicamente de m e de $E(\theta_t|D_t) = m_t$.

Tendo sido discutidos os principais aspectos metodológicos dos modelos sazonais na forma de Fourier, descrevemos agora um exemplo simulado, a partir do qual ilustramos como tratar essa classe de modelos usando o *R*. Considere o MDL definido pelas matrizes

$$\left\{ F = (1,0,1,0)', G = \begin{pmatrix} J_2(1, 2\pi/5) & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & J_2(1, 4\pi/5) \end{pmatrix}, W = 10^{-4} \times \mathbf{I}_4, V = 15 \right\}.$$

Esta especificação retrata um modelo com 5 períodos sazonais. Geramos uma série de tamanho 40 do modelo citado usando o primeiro bloco da programação do Código 1.2. Observe que, na linha 9, removemos aleatoriamente 20% das observações a fim de ilustrar o funcionamento dos MDL com observações faltantes. Internamente, o pacote *dlm* lida com esses valores indisponíveis (faltantes) da maneira descrita na Seção (1.3.1).

Novamente usando uma distribuição a priori vaga para o sistema, estimamos as médias $\mu_1, \ldots, \mu_{40}$ à luz de todos os dados disponíveis. A parte superior da Figura 1.4 evidencia a dificuldade em visualizar qualquer comportamento sazonal baseado

somente no gráfico dos valores Y_t. Entretanto, pela parte inferior da figura, fica claro o comportamento cíclico da média real não observável (linha vermelha). Mais uma vez, o procedimento inferencial foi bem sucedido na estimação dos valores reais, *separando* o sinal do ruído.

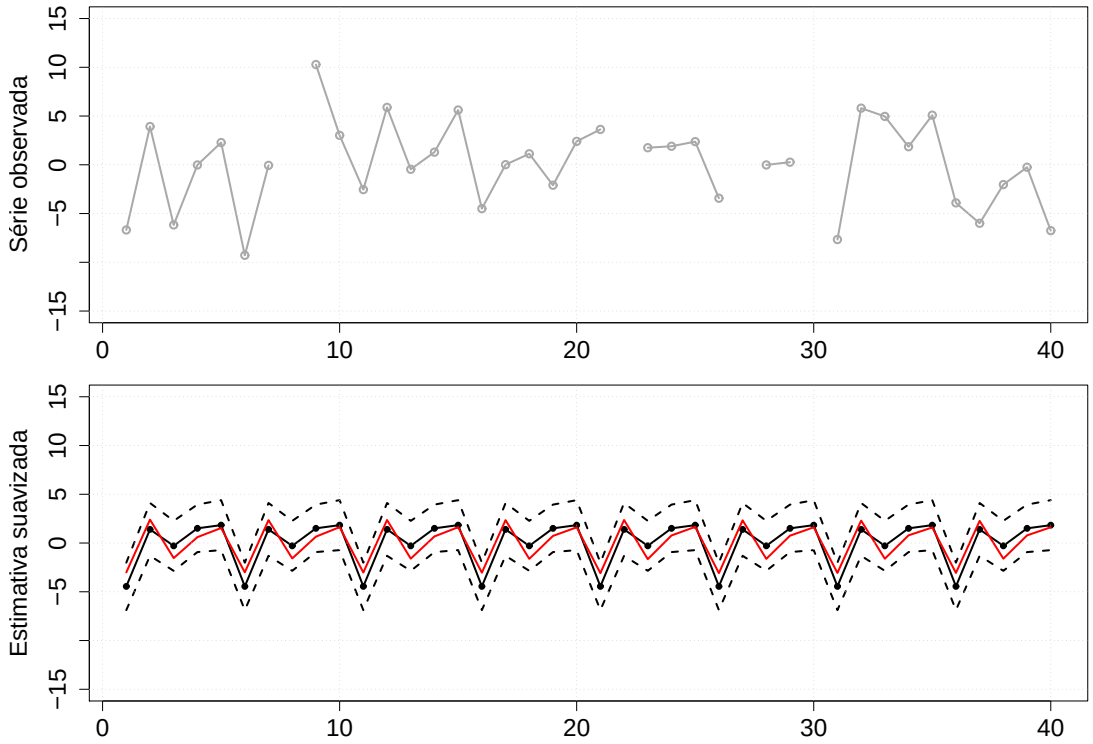

Figura 1.4: (a) Valores observados da série com dados faltantes y_t (linha cinza); (b) Valores reais (linha vermelha) e suavizados para μ_t (linha preta). Intervalos de credibilidade de 95% (linhas pretas tracejadas).

Código 1.2: Modelos dinâmicos lineares sazonais

```
> # Bloco 1: Geração dos dados Y_t
> N <- 40
> p <- 5
> mod <-  dlmModTrig(s = p, dW = rep(1e-4, p - 1),
+   dV = 15, m0 = rnorm(p - 1), C0 = diag(1, p - 1))
> aux <- dlmForecast(mod, nAhead = N, sampleNew = 1)
> y <- aux$newObs[[1]]
> theta.real <- aux$newStates[[1]]
> y[sample(2:(N - 1), N / 10), ] <- NA
>
```

```
> # Bloco 2: Suavização
> mod$m0 <- rep(0, p - 1)
> mod$C0 <- diag(1e+7, p - 1)
> aux <- dlmSmooth(y, mod)
> theta.suave <- dropFirst(aux$s)
> S <- dlmSvd2var(aux$U.S, aux$D.S)
> z <- qnorm(0.975)
> tmp <- sapply(2:(N + 1), function(x)
+   sqrt(mod$FF %*% S[[x]] %*% t(mod$FF))
> mu.suave <- theta.suave %*% t(mod$FF))
> lim.inf.S <- mu.suave - z * tmp
> lim.sup.S <- mu.suave + z * tmp
```

1.4.3 Modelos dinâmicos lineares de regressão

Não seria demais dizer que a possibilidade de incluir, com elegância, covariáveis ao modelo é uma das principais virtudes da classe dos modelos dinâmicos. Considere novamente a estrutura apresentada na Seção (1.2).

Equação das observações:

$$Y_t = F_t'\theta_t + \nu_t, \quad \nu_t \sim N[0, V_t] \; ; \tag{1.12a}$$

Equação do sistema:

$$\theta_t = G_t\theta_{t-1} + \omega_t, \quad \omega_t \sim N[0, W_t] \; . \tag{1.12b}$$

Observa-se que, se Y_t é uma série univariada, $W_t = \mathbf{0}$ e $G_t = \mathbf{I}_n$ para todo t, onde $\mathbf{I}_n$ é a matriz identidade $(n \times n)$, então o MDL coincide com o modelo clássico de regressão. Provavelmente, a notação mais utilizada em tais modelos é aquela onde F_t e $\theta_t = \theta$ são representados, respectivamente, por X_i e β. Neste caso, a ordem das observações na amostra é irrelevante.

Um MDL, como pôde ser observado, representa uma generalização do modelo de regressão que permite que ao vetor de parâmetros seja adicionada uma dinâmica evolutiva no tempo. Essa flexibilidade pode ser útil em diversas situações aplicadas.

A título de ilustração, não seria razoável assumir, por exemplo, que, para um determinado indivíduo, a relação entre a quantidade de carboidratos ingerida e a taxa de glicose no sangue seja sempre a mesma. De maneira mais geral, a relação entre essas variáveis pode evoluir ao longo do tempo, e a capacidade de descrever tal característica é a marca registrada dos modelos dinâmicos.

Visando ilustrar, computacionalmente, o uso dos MDL de regressão, bem como evidenciar a relação destes com os modelos clássico e seus respectivos mecanismos de estimação via mínimos quadrados, apresentamos o Código 1.3.

Código 1.3: Uso dos MDL em regressão linear

```
> # Bloco 1: Geração dos dados Y_t
> N <- 60
> F <- rnorm(N)
> theta.real <-  c(- 1, 1)
> y <- cbind(rep(1, N), F) % * % theta.real + rnorm(N)
>
> # Bloco 2: Estimação via mínimos quadrados
> reg.lm <- lm(y ~ F)
> print(theta.lm <- coef(reg.lm))
(Intercept)           F
  -0.970956    1.026672
> print(sd.theta <- vcov(reg.lm))
             (Intercept)            F
(Intercept)  0.015638968 -0.001457000
F           -0.001457000  0.016470073
>
> # Bloco 3: Estimação via MDL
> mod <-  dlmModReg(X = F)
> filtro <- dlmFilter(y, mod)
> print(theta.filtro <- filtro$m[N + 1, ])
[1] -0.970956  1.026672
> print(C <- dlmSvd2var(filtro$U.C, filtro$D.C)[[N + 1]])
             [,1]         [,2]
[1,]  0.016805169 -0.001565649
[2,] -0.001565649  0.017698250
```

O Código 1.3 está dividido em três Blocos. O Bloco 1 trata da geração de amostras de um modelo de regressão estático com dois parâmetros, sendo um deles referente ao intercepto. Os valores reais, atribuídos a eles na linha 4, são -1 e 1. O Bloco 2 permite a estimação clássica do vetor paramétrico (linha 11) e de sua respectiva matriz de covariância (linhas 14 e 15) usando a função *lm*.

Com o Bloco 3 estima-se as mesmas quantidades, agora com base no filtro de Kalman já apresentado. Observe que os valores estimados correspondem à média da distribuição a posteriori do vetor de parâmetros da regressão, ou equivalentemente, do sistema. A comparação dos valores retornados é bem interessante no sentido de que as retas de regressão estimadas foram rigorosamente as mesmas, embora uma pequena diferença entre as matrizes de covariância tenha sido observada.

A função *dlmModReg* (Petris, 2010) foi usada para facilitar a especificação do modelo. Como a matriz F_t varia a cada instante, os autores do pacote *dlm* adotaram uma estratégia eficiente para o armazenamento dos dados. Neste caso, ao invés de registrar todas as matrizes $F_1, \ldots, F_N$, o objeto *dlm* contém uma matriz de mesma dimensão *JFF* tal que cada uma de suas entradas é igual a zero se $F[i, j]$ não varia no tempo, e um valor inteiro positivo k caso contrário. Dessa forma, o elemento (i, j) de F_t deve ser armazenado em $X[t, k]$. O mesmo vale para os casos em que V, G ou W variam no tempo. Toda essa estratégia está bem explicada na *ajuda* da função *dlm*.

Por fim, é preciso destacar que na estimação via *dlmFilter*, os cálculos foram feitos usando a informação do valor real da variância dos ruídos $V = 1$. Isso porque o valor *default* da variância é o mesmo nas funções *rnorm* e *dlmModReg*. No entanto, em situações reais, esse valor é desconhecido e precisa ser estimado. A estimação das matrizes V e W será discutida na Seção (1.5). Visto que não alteramos os valores *default* para os momentos de $(\theta_0|D_0)$, foi adotada uma priori extremamente vaga.

1.4.4 Sobreposição de modelos lineares dinâmicos

Nas subseções anteriores, foram introduzidos os MDL de regressão, polinomiais e sazonais. Aqui, o interesse está na construção de modelos mais gerais, obtidos pela combinação dos componentes já apresentados. Como será visto a seguir, a especificação

de um modelo com tendência de crescimento linear, duas covariáveis e 4 períodos sazonais, por exemplo, é trivial. O princípio da sobreposição do MDL apresentado a seguir é a chave para tais construções (West e Harrison, 1997).

Qualquer combinação linear de MDL gaussianos independentes é um MDL gaussiano.

Formalmente, esse princípio pode ser estabelecido da seguinte forma: Considere h séries temporais Y_{it} geradas por MDL

$$M_i : \{F_{it}, G_{it}, V_{it}, W_{it}\}$$

para $i = 1, \ldots, h$. No modelo M_i, o vetor do sistema θ_{it} tem dimensão n_i, e os erros das observações e de evolução são, respectivamente, ν_{it} e ω_{it}. Os vetores do sistema são distintos, e para $i \neq j$, as séries ν_{it} e ω_{it} são mutuamente independentes das séries ν_{jt} e ω_{jt}. Nas referidas condições, a série

$$Y_t = \sum_{i=1}^{h} Y_{it}$$

segue um MDL $\{F_t, G_t, V_t, W_t\}$ n-dimensional, onde $n = n_1 + \ldots + n_h$ e o vetor do sistema θ_t e a quádrupla são dados por

$$\theta_t = \begin{pmatrix} \theta_{1t} \\ \vdots \\ \theta_{ht} \end{pmatrix}, \quad F_t = \begin{pmatrix} F_{1t} \\ \vdots \\ F_{ht} \end{pmatrix},$$

$$G_t = \text{bloco-diagonal}[G_{1t}, \ldots, G_{ht}],$$

$$W_t = \text{bloco-diagonal}[W_{1t}, \ldots, W_{ht}]$$

e

$$V_t = \sum_{i=1}^{h} V_{it}.$$

As matrizes G_t e W_t são tais que os blocos da diagonal são dados, respectivamente, pelas matrizes G_{it} e W_{it}, enquanto os elementos fora da diagonal são nulos.

Do ponto de vista prático, cada componente da observação Y_t é modelado separadamente por um bloco convenientemente especificado. Os próprios modelos sazonais na forma de Fourier podem ser entendidos como a sobreposição de submodelos, nos quais se descreve individualmente a evolução de cada harmônico. Do ponto de vista operacional, é viável testar uma grande variedade de modelos candidatos, todos eles acomodados em uma mesma estrutura. Medidas usuais de comparação, como o Critério de Informação Bayesiano *BIC* ou *Akaike*, podem ser empregados.

Embora os modelos ARIMA não tenham sido explorados nessa dissertação, pode ser mostrado que qualquer modelo ARIMA(p, d, q) pode ser representado por um MDL. Neste caso, as matrizes G_t passam a ter parâmetros desconhecidos relacionados à natureza auto-regressiva da série Y_t, enquanto W_t apresenta, entre seus componentes, os parâmetros relativos ao processo de médias móveis. O leitor interessado pode consultar West e Harrison (1997).

No pacote *dlm*, a sobreposição de modelos pode ser feita com a simples aplicação do operador lógico $'+'$. A título de ilustração, um modelo polinomial de crescimento quadrático, com 12 períodos sazonais representados na forma de Fourier, pode ser especificado a partir do seguinte comando: `mod <- dlmModPoly(3) + dlmModTrig(12)`. Como nos MDL o comportamento não-estacionário das observações é normalmente modelado diretamente por uma tendência polinomial ou por um componente sazonal, por exemplo, a correlação residual nos dados costuma ser descrita por um processo ARMA. A função *dlmModARMA* pode ser usada na especificação desses processos.

1.5 Modelos com parâmetros desconhecidos

Até o momento, assumimos que as quádruplas $\{F_t, G_t, V_t, W_t\}$ são conhecidas. Tal suposição foi útil para facilitar o entendimento de algumas das propriedades básicas dos modelos, embora, evidentemente, o conhecimento completo de todos esses elementos só se dê no campo da teoria. Em diversas situações práticas, as matrizes F_t e G_t são, de fato, conhecidas (mais precisamente, são integralmente especificadas), mas o mesmo não vale para as matrizes V_t e W_t.

De agora em diante, sem perda de generalidade, as matrizes do modelo dependem do vetor de parâmetros desconhecidos ψ. Diversas formas de lidar com essa questão

estão descritas na literatura.

Numa abordagem clássica, o pesquisador pode estimar o vetor ψ via máxima verossimilhança. Para isso, basta encontrar, algebricamente ou numericamente, os valores que maximizam a função de verossimilhança

$$L(\psi) = p(y_1, \ldots, y_n|\psi) = \prod_{t=1}^{n} p(y_t|D_{t-1}, \psi). \tag{1.13}$$

Com base em resultados assintóticos, válidos sob certas condições de regularidade, a variância do estimador de máxima verossimilhança (EMV) pode ser estimada pelo inverso da informação de Fisher, calculada no ponto que maximiza a função dada em (1.13). Nessa abordagem, uma prática comum é usar os valores estimados de ψ como se fossem constantes conhecidas e então aplicar as técnicas descritas nas seções anteriores. Esse método *plug-in*, apesar de simples, ignora a incerteza sobre o vetor ψ podendo, em última instância, comprometer a qualidade dos intervalos de credibilidade construídos.

Entre as fragilidades dos EMV no contexto dos MDL, uma tem posição de destaque. Modelos para os quais o vetor ψ tem dimensão grande ou moderada tornam inviável o processo de maximização. A superfície pode conter diversos máximos locais ou ser quase plana na região que contém seu máximo global. Em ambos os casos, mínimas variações nos dados ou nos valores iniciais usados no algoritmo de maximização podem resultar em uma drástica mudança na estimativa dos valores de ψ. Apesar das considerações feitas, o método é válido e pode ser utilizado facilmente a partir da função *dlmMLE* do pacote *dlm* (Petris, 2010).

Felizmente, existem abordagens *bayesianas* mais adequadas para lidar com a estrutura imposta nos modelos dinâmicos. Nesta direção, em harmonia com o que prescreve o paradigma Bayesiano, a incerteza em relação ao vetor ψ é modelada por uma lei de probabilidade, mais precisamente, por uma distribuição a priori $p(\psi)$. Assim, para qualquer $n \geq 1$, assume-se que

$$(\theta_0, \theta_1, \ldots, \theta_n, Y_1, \ldots, Y_n, \psi) \sim p(\theta_0|\psi)p(\psi) \prod_{t=1}^{n} p(y_t|\theta_t, \psi)p(\theta_t|\theta_{t-1}, \psi). \tag{1.14}$$

Compare a expressão em (1.14) com a Equação (1.3).

Dado as observações $(y_1, \ldots, y_t)$, inferências sobre o estado θ_s, para $s = 1, \ldots$, podem ser expressas pela equação

$$p(\theta_s|D_t) = \int p(\theta_s|\psi, D_t)p(\psi, D_t)d\psi. \tag{1.15}$$

Em certos modelos, adotando-se prioris conjugadas, as distribuições a posteriori dos estados possuem formas fechadas e podem ser obtidas recursivamente via aplicação do teorema de Bayes. Por outro lado, em situações um pouco mais complexas, tais distribuições precisam ser aproximadas por métodos de simulação estocástica *MCMC*. Há uma grande distância entre estes casos, tanto no que diz respeito à técnica de estimação, quanto às consequências da adoção de cada procedimento. A abordagem recursiva é dita *online* e será superficialmente tratada na Subseção (1.5.1). A estratégia MCMC, por sua vez, será discutida na Subseção (1.5.2).

1.5.1 Procedimentos inferenciais *online*

Considere o caso recorrente em que as matrizes F_t e G_t são conhecidas, enquanto as variâncias $V_t = V$ e W_t não o são. Apresentamos agora uma estratégia recursiva para lidar com tal situação.

1.5.1.1 Especificação das matrizes de covariância W_t

As matrizes de covariâncias do sistema - W_t - são parte da construção da priori, e portanto, podem ser especificadas pelo analista. No entanto, a calibragem inadequada da magnitude de W_t pode trazer graves consequências ao ajuste. Os valores dessas matrizes determinam quão estável é o sistema, e em que medida se dá a perda de informação, entre os tempos $t-1$ e t, sobre seus parâmetros. De acordo com (1.5), $V(\theta_t|D_{t-1}) = R_t = G_tC_{t-1}G_t' + W_t$. A quantidade $P_t = G_tC_{t-1}G_t'$ representa a incerteza associada à projeção do sistema em um cenário livre de variações estocásticas no tempo t.

Na abordagem *online*, a estimação de W_t é normalmente inviável e a solução padrão é especificá-la adequadamente. Justamente com o objetivo de definir um método para a especificação das matrizes de evolução do sistema, Harrison e Scott

(1965) propuseram o uso de fatores de desconto - FD. A idéia básica da técnica é definir uma estrutura, para cada W_t, que depende unicamente de quantidades conhecidas e do fator de desconto δ. Isso é feito diretamente assumindo-se uma taxa constante de decaimento da informação. Matematicamente, define-se

$$V(\theta_t|D_{t-1}) = R_t = \frac{1}{\delta}P_t$$

e, consequentemente,

$$W_t = \frac{1-\delta}{\delta}P_t,$$

onde $0 < \delta \leq 1$ é o fator de desconto.

Teoricamente, δ poder assumir qualquer valor no intervalo $(0, 1]$, embora usualmente adote-se valores acima de 0.8. Valores muito baixos significam que a informação é pouco duradoura e, em consequência, o modelo é pouco confiável. Como exemplo, um FD de 0.5 implica em $W_t = P_t$. Ou seja, a variância da previsão do sistema é duas vezes maior do que seria sem a adição dos erros ω_t.

A metodologia pode ser generalizada de maneira que um MDL possa ter um fator de desconto para cada bloco. Considere, novamente, o princípio da sobreposição dos MDL tratado na Seção (1.4.4). A modelagem em separado dos diversos efeitos do sistema (nível, tendência, sazonalidade, etc) sugere que a evolução dos blocos seja independente. Ou seja, da mesma forma em que as matrizes G_t são bloco-diagonal, seria razoável que as matrizes W_t também o fossem, o que não acontece quando se usa somente um FD.

Outro ponto que motiva a especificação de um fator de desconto para cada bloco é o fato de que as estabilidades de cada efeito podem ser bem diferentes umas das outras. Há na literatura um entendimento de que os componentes sazonais costumam ser mais estáveis/duradouros que os componentes de tendência, por exemplo. Neste caso, o avaliador tem autonomia para definir, por exemplo, um FD de 0.95 para nível e tendência e outro de 0.98 para sazonalidade, caso julgue necessário.

Ainda valendo-nos da notação empregada na construção de modelos sobrepostos,

defina

$$P_{it} = V[G_{it}\theta_{it}|D_{t-1}],$$

para $i = 1, \ldots, h$. Cada P_{it} mede individualmente a informação sobre o vetor do sistema do submodelo i. A idéia é, então, aplicar a técnica do FD fazendo $W_{it} = P_{it}(1 - \delta_i)/\delta_i$. Dessa maneira, a matriz de variância dos erros de evolução é definida como $W_t = \text{bloco-diagonal}(W_{1t}, \ldots, W_{ht})$. O fator de desconto é, então, denotado por $\delta = (\delta_1, \ldots, \delta_h)$. Na prática, pode-se optar por usar somente um FD para descrever todo o componente sazonal ou um conjunto de variáveis regressoras. Para maiores detalhes, veja West e Harrison (1997).

Para se ter idéia da simplicidade do método, considere um modelo de dimensão igual a 12, composto por 2 blocos. No caso de W_t descrita por meio de FD, há apenas 2 parâmetros a especificar. Já no caso de W_t com todos os parâmetros desconhecidos, é necessário estimar $12 * 13/2 = 78$ parâmetros.

De acordo com West e Harrison (1997), considerando um MDL polinomial de segunda ordem, o incremento máximo no desvio padrão dos erros de previsão a um passo à frente gerado pelo uso do FD costuma ser menor que 1%. Ou seja, há uma perda desprezível no desempenho preditivo e um forte ganho em simplicidade. Neste exemplo, usando fator de desconto, trocamos três estimações ($\omega_{1,t}$, $\omega_{2,t}$ e $\omega_{12,t}$) por uma especificação (δ). Os autores defendem, ainda, que pouco se perde em termos de capacidade de descrição do fenômeno (West e Harrison, 1997).

Nossa experiência no uso de tal técnica não é assim tão otimista. Normalmente, δ é especificado levando-se em consideração critérios relacionados ao desempenho preditivo do modelo. Se o interesse da análise for justamente esse, prever, entendemos que tal abordagem pode ser uma ótima estratégia para lidar com as matrizes W_t. Por outro lado, quando se quer avaliar o impacto de certos fatores nas observações Y_t, o modelo construído com base nos fatores de desconto pode ser consideravelmente menos eficiente no que se refere à capacidade descritiva, quando comparado ao modelo irrestrito onde W_t são matrizes quaisquer. Particularmente, sugerimos uma cautela especial quanto à utilização dos valores ajustados e retrospectivos referentes ao componente de tendência β_t apresentado na Subseção (1.4.1).

Ao longo das inúmeras simulações feitas neste trabalho de pesquisa, observamos que as medidas usuais do desempenho preditivo do modelo são, em certo sentido, muito frágeis. Considere dois modelos polinomiais de ordem 2, com 5 períodos sazonais, definidos identicamente, exceto pelo valor de δ. No primeiro $\delta = (0.8, 0.98)$ e no segundo $\delta = (0.9, 0.9)$. Uma simples investigação descritiva torna clara a enorme distância entre esses modelos. As curvas de cada parâmetro do sistema seguem trajetórias e comportamentos bem distintos. Entretanto, as medidas do desempenho preditivo podem ser facilmente muito parecidas.

Em varias situações, West e Harrison (1997) argumentam que os fatores de desconto devem refletir a visão do analista quanto à estabilidade do sistema. Ocorre que nem sempre esse mesmo analista tem uma visão a priori da mencionada estabilidade e acaba tendo enorme dificuldade em justificar sua escolha. Voltando ao exemplo dado, como escolher entre os dois modelos citados? Em resumo, acreditamos que a essência do método é muito interessante e sua aplicação pode ser de fato recomendável. Entretanto, é preciso estar ciente da fragilidade da aplicação das medidas usuais para determinar o vetor δ.

Um último comentário sobre o uso dos FD se faz necessário. Para previsões m-passos a frente, repetidas aplicações do fator de desconto δ levariam a uma variância da forma $V(\theta_{t+m}|D_t) = R_t(m) = G^m C_t G'^m/\delta^m$. Isso implicaria em um decaimento exponencial da informação, contrariando o princípio de que a perda de informação é aritmética. O mesmo ocorre quando há na série Y_t duas ou mais observações faltantes seguidas. Em ambos os casos, a solução mais conveniente é adotar, para $m = 1, \ldots,$ a variância condicionalmente constante

$$V(\omega_{t+m}|D_t) = W_t(m) = W_{t+1}.$$

1.5.1.2 Modelo linear dinâmico para V desconhecido

Há uma forma muito elegante de lidar com o caso em que as matrizes $V_t = V$ são constantes, porém desconhecidas. Como de costume na notação Bayesiana, denotamos por $\phi = V^{-1}$ o parâmetro de precisão das observações. A idéia central é atribuir uma distribuição a priori gama para ϕ escrita como $p(\phi|D_0) = G(\cdot\,, \cdot)$. Fazendo isso, todas as distribuições de interesse possuem forma fechada e a natureza recursiva do

filtro de Kalman é preservada. Quando uma nova observação se torna disponível, o analista pode atualizar tanto a informação para o estado, quanto a informação para a precisão ϕ. As distribuições marginais do sistema são agora t-Student, em substituição à distribuição normal.

As fórmulas de atualização, que não serão replicadas aqui, podem ser encontradas em West e Harrison (1997). Nelas, o valor desconhecido V é substituído pela sua estimativa no tempo t, dada por $E(\phi|D_t)$. De certo modo, pode-se dizer que as vantagens dessa estratégia são consequência de sermos capazes de resolver analiticamente a expressão

$$p(\theta_t|D_t) = \int p(\theta_t, \phi|D_t)d\phi = \int p(\theta_t|\phi, D_t)p(\phi|D_t)d\phi. \tag{1.16}$$

Para que fique claro, ressaltamos que a distribuição à esquerda da igualdade é t-Student, e as distribuições no integrando são, respectivamente, normal e gama.

Uma analogia pode ser feita com a estatística clássica, na qual se constroem intervalos de confiança, baseados na distribuição t-Student, para a média de uma variável aleatória normal com variância desconhecida.

A condição de variância constante $V_t = V$ pode não ser realista. A noção de fator de desconto descrita para especificar as matrizes W_t também pode ser empregada para impor uma evolução temporal para o parâmetro ϕ. A descrição completa dessa estratégia pode ser encontrada em West e Harrison (1997).

1.5.2 Procedimentos inferenciais *offline*

Em situações mais complexas, em geral não é possível calcular, em forma fechada, a distribuição conjunta a posteriori dos parâmetros ψ e dos estados latentes, ou seja,

$$p(\psi, \theta_0, \ldots, \theta_T|D_T). \tag{1.17}$$

A solução mais adotada é simular uma amostra dessa distribuição de interesse e calcular certas medidas resumo, como média e variância, a partir da amostra gerada. Além de resolver a tarefa de filtragem e suavização, cada ponto simulado da posteriori

pode ser usado para se obter uma observação da distribuição a priori dada por

$$p(\theta_{T+1}, \ldots, \theta_{T+m}, y_{T+1}, \ldots, y_{T+m}|D_T). \tag{1.18}$$

Nesse sentido, o problema se resume em como gerar uma amostra da distribuição (1.17). A resposta quase sempre está relacionada a métodos de simulação estocástica MCMC, em especial, amostrador de *Gibbs* e *Metropolis-Hastings* (Geweke e Tanizaki, 2001).

No problema em questão, o método de Metropolis-Hastings com passos de Gibbs consiste em gerar valores de (1.17), simulando alternadamente das distribuições condicionais completas $p(\psi|\theta_0, \ldots, \theta_T, D_T)$ e $p(\theta_0, \ldots, \theta_T|\psi, D_T)$. Enquanto a primeira densidade depende do problema específico em consideração e pode não ter forma fechada, a segunda pode ser sempre amostrada pelo método *FFBS* (do inglês *Foward Filtering Backward Sampling*), proposto por Frühwirth-Schnatter (1994) e Carter e Kohn (1994). Não descreveremos esse amostrador aqui mas, na Seção 3.2.1.2, apresentaremos, em detalhe, o amostrador *CUBS*, que representa uma adaptação do *FFBS* para os modelos dinâmicos lineares generalizados descritos no Capítulo 2. Em resumo, costuma-se empregar o seguinte algoritmo.

1. Inicialização: Defina um valor inicial $\psi = \psi^{(0)}$.

2. Para $i = 1, \ldots, N$:

   (a) Amostre $\theta_0^{(i)}, \ldots, \theta_T^{(i)}$ de $p(\theta_0, \ldots, \theta_T|D_t, \psi = \psi^{(i-1)})$ usando o *FFBS*.

   (b) Amostre $\psi^{(i)}$ de $p(\psi|D_t, \theta_0 = \theta_0^{(i)}, \ldots, \theta_T = \theta_T^{(i)})$.

Há diversas maneiras de executar o passo (b). O vetor ψ pode ser amostrado em um passo único, elemento por elemento, ou em blocos. Tudo depende dos parâmetros que o compõem e de suas respectivas densidades condicionais. Veremos nos próximos capítulos que nem sempre é fácil amostrar diretamente da distribuição condicional completa de ψ (ou de um subvetor ψ_i) e, nesses casos, usar o amostrador de Metropolis-Hastings pode ser a melhor saída. No *R*, a função *dlmBSample*, em conjunto com a função *dlmFilter*, executa o passo (a).

São muitas as vantagens da abordagem *offline*. Entretanto, para atualizar a distribuição a posteriori no momento em que uma nova observação se faz disponível, é preciso gerar uma amostra Gibbs totalmente nova, o que pode ser extremamente ineficiente. A depender do caso, o tempo computacional para realizar essa tarefa pode ser proibitivo.

Métodos de simulação sequencial Monte Carlo, em especial o filtro de partícula, surgem como alternativa às abordagens *online* e *offline*. Basicamente, tais métodos permitem o aproveitamento das amostras geradas anteriormente para atualizar a posteriori dada em (1.17). Embora o tema não tenha sido tratado nesse trabalho de pesquisa, nos parece que sua aplicação pode trazer bons resultados. Como não poderia deixar de ser, essa abordagem também tem limitações que devem ser consideradas na comparação dos métodos.

Capítulo 2

Modelos Dinâmicos Lineares Generalizados

Os modelos apresentados no Capítulo 1 são aplicáveis a diversas situações práticas. Ainda assim, há duas suposições um tanto restritivas em sua construção. A primeira refere-se ao fato de assumir-se normalidade para a distribuição das observações. Evidentemente, tal suposição nem sempre é adequada para a análise de, por exemplo, dados de proporções e contagens, ou dados com comportamento assimétrico. A segunda suposição diz respeito à forma com que os parâmetros do sistema se relacionam, em cada instante, com a média μ_t da série Y_t. Nos MDL assume-se que essa relação é caracterizada por uma transformação linear. Quando são válidas, as duas suposições acima referidas implicam em grandes facilidades computacionais e metodológicas. Porém, quando tais suposições não são válidas, há necessidade de se buscar modelos mais adequados.

Justamente visando atenuar a mencionada rigidez resultante das suposições citadas, West et al. (1985) combinaram a abrangência dos Modelos Lineares Generalizados (MLG), introduzidos por Nelder e Wedderburn (1972), com a natureza dinâmica dos parâmetros dos Modelos Dinâmicos Lineares (MDL) para propor a classe dos Modelos Dinâmicos Lineares Generalizados (MDLG). Essa classe de modelos assume que a distribuição de probabilidade da série Y_t pertence à família exponencial, e que o vetor de parâmetros do sistema relaciona-se à média de Y_t por meio de uma função de ligação. Nessas condições, a série pode ser modelada, a critério do analista, por uma distri-

buição Poisson, Gama, Beta, Binomial, Normal, entre outras. Além disso, é possível relacionar o preditor linear $\lambda_t = F_t'\theta_t$ à média $\mu_t = E(Y_t|\lambda_t)$ via transformação logito, probito, complemento log-log, identidade, etc.

Este capítulo está dividido em duas seções. A primeira delas apresenta formalmente os MDLG e descreve seu procedimento inferencial. A segunda seção aborda, especificamente, o caso em que a distribuição condicional da série Y_t é da família beta.

2.1 Caso geral

Considere uma série temporal Y_t univariada. O modelo dinâmico linear generalizado é caracterizado pelos seguintes componentes.

- **Equação da observação:**

$$(Y_t|\eta_t) \sim \exp\{\phi_t[t(y_t)\eta_t - b(\eta_t)] + c(y_t, \phi_t)\}, \quad t = 1, 2, \ldots \tag{2.1}$$

- **Função de ligação:**

$$g(\mu_t) = g(E(Y_t|\eta_t)) = F_t'\theta_t = \lambda_t \tag{2.2}$$

- **Equação do sistema:**

$$\theta_t = G_t\theta_{t-1} + w_t \; ; \quad w_t \sim (0, W_t) \tag{2.3}$$

- **Informação inicial:**

$$(\theta_0|D_0) \sim (m_0, C_0), \tag{2.4}$$

onde $b(\cdot)$, $c(\cdot\,,\,\cdot)$ e $t(\cdot)$ são funções conhecidas, η_t é o parâmetro natural e $\phi_t = V_t^{-1}$, o parâmetro de precisão da distribuição.

Observe que, nessa formulação, tanto a distribuição de ω_t quanto a de θ_0 são especificadas somente por seus momentos, não se impondo distribuição alguma para ω_t e θ_0. Assume-se que, dado η_t, as observações Y_t são independentes entre si e dos

erros de evolução ω_t, para $t = 1, 2, \ldots$. É assumido, ainda, que os erros de evolução também são a priori independentes entre si.

A estrutura apresentada é parecida com a dos MDL, embora haja, aqui, alguns elementos adicionais. Da equação (2.2) vemos que μ_t, η_t e λ_t se relacionam deterministicamente. Ou seja, o conhecimento de qualquer um deles é suficiente para se calcular o valor dos outros dois. A função de ligação $g(\cdot)$ é especificada pelo analista e define como variações do preditor linear interferem na média da série. Por outro lado, a relação entre o parâmetro natural η_t e a média da distribuição de Y_t, μ_t, é determinada diretamente pela distribuição das observações. Usando propriedades da função *escore*, definida como a primeira derivada do logaritmo da função de verossimilhança, é possível provar que

$$E(Y_t|\eta_t, \phi_t) = b'(\eta_t) = \mu_t \ ;$$
$$V(Y_t|\eta_t, \phi_t) = \frac{b''(\eta_t)}{\phi_t}.$$

Convém destacar que a função $b(\cdot)$ é duas vezes diferenciável, convexa, e pode ser encontrada escrevendo-se a distribuição das observações na forma da família exponencial.

Os MDLG são extremamente flexíveis. Observe que, embora apresentem uma estrutura aparentemente simples, dependem de mais de vinte componentes, entre parâmetros, momentos e funções. O tratamento ou entendimento inadequado relativo a qualquer um deles pode determinar o fracasso do ajuste.

2.1.1 Procedimento inferencial

Nesta seção, assumimos conhecidas as matrizes F_t, G_t, W_t e o parâmetro de precisão ϕ_t. Quando ao menos um desses elementos não é conhecido, técnicas (do tipo *online* e *offline*) filosoficamente iguais às apresentadas no caso normal são empregadas (vide Seções 1.5.1 e 1.5.2).

O processo inferencial tem o mesmo caráter sequencial dos MDL. Porém, alguns passos adicionais são necessários na estimação dos parâmetros do modelo. A seguir, será descrito, de forma resumida, como funciona esse processo.

1. *Reconhecimento dos momentos a priori de θ_t e λ_t*

   A distribuição a posteriori de θ_{t-1} tem vetor de médias m_{t-1} e matriz de covariâncias C_{t-1}. Portanto, pela equação de evolução (2.3), $(\theta_t|D_{t-1}) \sim (a_t, R_t)$, onde $a_t = G_t m_{t-1}$ e $R_t = G_t C_{t-1} G_t' + W_t$. Como $\lambda_t = F_t'\theta_t$, a distribuição a priori do preditor linear é parcialmente especificada por

   $$(\lambda_t|D_{t-1}) \sim (f_t, q_t), \tag{2.5}$$

   com $f_t = F_t' a_t$ e $q_t = F_t' R_t F_t$.

2. *Especificação da priori*

   Para que se possa calcular a distribuição preditiva, bem como processar a informação contida em novos dados, é preciso especificar a forma funcional da priori de λ_t, η_t ou μ_t. Convém destacar que essa escolha é arbitrária, tendo como única restrição que os momentos de $(\lambda_t|D_{t-1})$ sejam especificados de acordo com a Equação (2.5). Como mencionado anteriormente, a relação determinística entre a média, o preditor linear e o parâmetro natural, permite ao analista decidir a qual deles será dada uma distribuição a priori.

   West e Harrison (1997) ressaltam que a distribuição conjugada, na escala de η_t, tem forma fechada conhecida e normalmente é a melhor opção. Por outro lado, na situação em que a distribuição das observações é da família Beta, da-Silva, Migon e Correia (2011) mostram que há vantagens em se definir uma priori não conjugada. Ou seja, é útil avaliar caso a caso.

   Suponha, sem perda de generalidade, que o analista optou por estipular a distribuição a priori, definida pelos hiper-parâmetros r_t e s_t, diretamente para a média μ_t. O próximo passo do ciclo inferencial é, então, elicitar os valores de r_t e s_t em conformidade com a relação $g(\mu_t) = \lambda_t$ e com a Equação (2.5). Isso é feito resolvendo-se, em r_t e s_t, o sistema de equações simultâneas

   $$\begin{cases} f_t = E(g(\mu_t)|D_{t-1}) = h_1(r_t, s_t) \\ q_t = V(g(\mu_t)|D_{t-1}) = h_2(r_t, s_t) \end{cases}.$$

3. *Atualização da distribuição de μ_t*

   Uma vez que a observação Y_t torna-se disponível, é possível, ainda que nem sempre muito facilmente, estimar os momentos a posteriori de μ_t, cuja distribuição é obtida pelo *Teorema de Bayes*:

$$p(\mu_t|D_t, \phi_t) \propto p(y_t|\mu_t, D_{t-1}, \phi_t)p(\mu_t|D_{t-1}) \ . \tag{2.6}$$

4. *Atualização dos momentos de λ_t*

   Uma vez estimados os dois primeiros momentos de $(\mu_t|D_t)$, faz-se o caminho inverso para obter os momentos correspondentes, f_t^* e q_t^*, da posteriori de λ_t. Ou seja, resolve-se o sistema de equações

$$\begin{cases} E(\lambda_t|D_t) = f_t^* = E(g(\mu_t)|D_t) \\ V(\lambda_t|D_t) = q_t^* = V(g(\mu_t)|D_t) \ . \end{cases}$$

5. *Atualização dos momentos de θ_t*

   A solução, muito elegante, proposta pelos autores dos MDLG para calcular os momentos (m_t, C_t) baseia-se no uso do estimador conhecido como *Linear Bayes*. Sendo a distribuição a priori conjunta de (λ_t, θ_t) parcialmente especificada por

$$\begin{pmatrix} \lambda_t \\ \theta_t \end{pmatrix} \Bigg| \ D_{t-1} \sim \left[\begin{pmatrix} f_t \\ a_t \end{pmatrix}, \ \begin{pmatrix} q_t & F_t'R_t \\ R_tF_t & R_t \end{pmatrix} \right],$$

   a estimação dos momentos de $(\theta_t|\lambda_t, D_{t-1})$ via linear Bayes, resulta em

$$\begin{aligned} \hat{E}(\theta_t|\lambda_t, D_{t-1}) &= a_t + R_tF_t(\lambda_t - f_t)/q_t \ ; \\ \hat{V}(\theta_t|\lambda_t, D_{t-1}) &= R_t - R_tF_tF_t'R_t/q_t \ . \end{aligned}$$

   Vale observar que as equações obtidas funcionam como uma regressão linear de θ_t em λ_t. Em outras palavras, a esperança de $(\theta_t|\lambda_t)$ é igual a média a priori de θ_t, a_t, corrigida linearmente em função da distância de Mahalanobis entre λ_t e sua média f_t. Além disso, o peso de λ_t na esperança condicional de θ_t é determinado unicamente pela covariância entre estes dois vetores, como deveria

ser. O estimador apresenta algumas propriedades ótimas. O leitor interessado pode consultar West e Harrison (1997) para maiores detalhes.

Finalmente, pelo teorema de Bayes, a distribuição a posteriori de θ_t é obtida pela relação

$$p(\theta_t|D_t) = \int P(\theta_t|\lambda_t, D_{t-1})P(\lambda_t|D_t)d\lambda_t \ .$$

Via de regra, $p(\theta_t|D_t)$ não tem forma fechada conhecida, mas seus momentos são estimados por

$$\begin{aligned} m_t &= E(\theta_t|D_t) \\ &= E[\hat{E}(\theta_t|\lambda_t, D_{t-1})|D_t] \\ &= a_t + R_t F_t(f_t^* - f_t)/q_t \ ; \qquad (2.7) \\ C_t &= V(\theta_t|D_t) \\ &= V[\hat{E}(\theta_t|\lambda_t, D_{t-1})|D_t] + E[\hat{V}(\theta_t|\lambda_t, D_{t-1})|D_t] \\ &= R_t - R_t F_t F_t' R_t(1 - q_t^*/q_t)/q_t \ . \qquad (2.8) \end{aligned}$$

É importante citar que λ_t faz a ponte entre as observações Y_t e o sistema, descrito pelo vetor θ_t. O fluxo da análise sequencial do MLDG está representado na Figura 2.1.

$$\begin{array}{ccccc} \ldots\ (\theta_{t-1}|D_{t-1}) & \xRightarrow{\text{Evolução}} & (\theta_t|D_{t-1}) & & (\theta_t|D_t)\ \ldots \\ & & \Downarrow & & \Uparrow \\ & & (\lambda_t|D_{t-1}) & \xRightarrow{\text{Atualização}} & (\lambda_t|D_t) \end{array}$$

Figura 2.1: Análise sequencial do modelo linear generalizado dinâmico

2.2 Modelo dinâmico Beta

Nesta seção, apresenta-se a metodologia proposta por da-Silva et al. (2011) para a modelagem de séries temporais definidas no intervalo $(0, 1)$. Dados com essa característica surgem em inúmeras aplicações e áreas do conhecimento e, em geral, representam proporções, probabilidades e taxas. Pense, por exemplo, na modelagem longitudinal da porcentagem 1) de vôos atrasados em um dado aeroporto, 2) da área de plantio destinada à produção de soja ou 3) de carros importados, entre aqueles comercializados no período.

Os autores abordam as estratégias associadas ao processo de estimação num contexto em que a distribuição das observações, considerando o modelo apresentado na Seção 2.1, pertence, especificamente, à família Beta. Essa família de distribuições, com suporte no segmento $(0, 1)$, é bastante flexível e assume diversas formas para diferentes valores dos parâmetros.

Considerando a parametrização proposta por Ferrari e Cribari-Neto (2004), na qual a média é o próprio parâmetro canônico da distribuição, o modelo é definido pelos seguintes componentes:

- **Equação da observação:**

$$p(y_t|\mu_t, \phi) = \frac{\Gamma(\phi)}{\Gamma(\phi\mu_t)\Gamma(\phi(1-\mu_t))} y_t^{\phi\mu_t - 1}(1-y_t)^{\phi(1-\mu_t)-1}$$

- **Distribuições a priori:**

$$(\mu_t|D_{t-1}) \sim \text{Beta}(r_t, s_t) \quad \text{e} \quad \phi \sim \text{Gama}(\alpha, \beta)$$

- **Função de ligação: logito**

$$\lambda_t = F_t'\theta_t = \log\left(\frac{\mu_t}{1-\mu_t}\right)$$

- **Equação do sistema:**

$$\theta_t = G_t\theta_{t-1} + w_t \; ; \quad w_t \sim (0, W_t)$$

- **Informação inicial:**

$$(\theta_0|D_0) \sim (m_0, C_0)$$

Observe que não foi adotada uma priori conjugada, ou seja, a distribuição a posteriori de μ_t não é da família Beta. Outro ponto importante a se notar é a escolha da função de ligação. A função logito é, provavelmente, a mais utilizada nesse tipo de situação. Tal transformação garante que, para qualquer que seja o valor do preditor linear, a estimativa de μ_t esteja dentro do espaço paramétrico. Outras funções podem ser adotadas sem que sejam adicionadas ao problema dificuldades significativas de natureza metodológica ou computacional. Por fim, ressaltamos que embora o modelo proposto, e seu respectivo procedimento de estimação, atribua uma priori Gama para ϕ, optamos por abordar o caso em que este parâmetro é conhecido. A decisão se justifica pela facilidade de comunicação do modelo dinâmico Beta com o modelo dinâmico Dirichlet apresentado no Capítulo 3. Enquanto da-Silva et al. (2011) estimam recursivamente o parâmetro ϕ, nós adotamos a abordagem de estimação *offline* via MCMC.

2.2.1 Procedimento inferencial

Baseado no modelo apresentado e na sequência de procedimentos na estimação dos parâmetros do modelo descritos na Seção 2.1, e, novamente, assumindo conhecidas as matrizes F_t, G_t, W_t e o parâmetro de precisão ϕ_t, obtém-se a seguinte sequência de passos:

1. *Reconhecimento dos momentos a priori de θ_t e λ_t*

$$(\theta_t|D_{t-1}, \phi) \sim (a_t, R_t) \quad \text{e} \quad (\lambda_t|D_{t-1}, \phi) \sim (f_t, q_t), \quad \text{onde:}$$
$$a_t = G_t m_{t-1}, \quad R_t = G_t C_{t-1} G_t' + W_t, \quad f_t = F_t' a_t \quad \text{e} \quad q_t = F_t' R_t F_t.$$

2. *Obtenção dos parâmetros da priori de* μ_t

da-Silva, Migon e Correia (2011) provam que

$$E(\lambda_t|D_{t-1},\phi) = E\left[\log\left(\frac{\mu_t}{1-\mu_t}\right)|D_{t-1},\phi\right] = \psi(r_t) - \psi(s_t) = f_t \ ;$$
$$V(\lambda_t|D_{t-1},\phi) = V\left[\log\left(\frac{\mu_t}{1-\mu_t}\right)|D_{t-1},\phi\right] = \psi'(r_t) + \psi'(s_t) = q_t \ ,$$

onde $\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$ e $\psi'(z)$ são, respectivamente, as conhecidas funções *digamma* e *trigamma*. É sabido que tais funções podem ser aproximadas, via expansões de Taylor de ordem 1 por, respectivamente, $\psi(z) \approx \log(z)$ e $\psi'(z) \approx 1/z$. As aproximações usualmente são satisfatórias. Adotando tais aproximações obtêm-se

$$f_t \approx \log\left(\frac{r_t}{s_t}\right) \ ;$$
$$q_t \approx \frac{1}{s_t} + \frac{1}{r_t} \ .$$

Resolvendo o sistema para r_t e s_t, chega-se às seguintes relações:

$$s_t \approx \frac{1+\exp(-f_t)}{q_t}$$
$$r_t \approx \frac{1+\exp(f_t)}{q_t}.$$

Vale destacar que se for escolhida outra função de ligação, é possível encontrar r_t e s_t via expansão de Taylor para a função $f(\lambda_t) = g^{-1}(\lambda_t)$.

3. *Atualização dos parâmetros de* μ_t

A distribuição a posteriori de μ_t é obtida pelo *Teorema de Bayes*,

$$\begin{aligned} p(\mu_t|D_t,\phi) &\propto p(y_t|\mu_t, D_{t-1},\phi)p(\mu_t|D_{t-1}) \\ &\propto \frac{\Gamma(\phi)}{\Gamma(\phi\mu_t)\Gamma(\phi(1-\mu_t))} y_t^{\phi\mu_t-1}(1-y_t)^{\phi(1-\mu_t)-1} \\ &\quad \times \frac{\Gamma(s_t+r_t)}{\Gamma(r_t)\Gamma(s_t)} \mu_t^{r_t-1}(1-\mu_t)^{s_t-1}. \end{aligned}$$

Como $p(\mu_t|D_t,\phi)$ não tem forma fechada conhecida, é preciso calcular os mo-

mentos de primeira e segunda ordem, $\tilde{\mu}_t$ e $\tilde{V}_t$, de forma aproximada. da-Silva, Migon e Correia (2011) recomendam o uso do método de quadratura de Gauss-Legendre (Lange, 1999).

4. *Atualização dos parâmetros de* λ_t

   Queremos, neste momento, obter os parâmetros f_t^* e q_t^* satisfazendo

$$E(\lambda_t|D_t,\phi) = E(g(\mu_t)|D_t,\phi) = f_t^* \ ;$$
$$V(\lambda_t|D_t,\phi) = V(g(\mu_t)|D_t,\phi) = q_t^* \ .$$

   Pela função de ligação,

$$\begin{aligned} \mu_t = g^{-1}(\lambda_t) &= \frac{\exp(\lambda_t)}{1+\exp(\lambda_t)} \\ &\approx \frac{\exp(f_t^*)}{1+\exp(f_t^*)} + (\lambda_t - f_t^*)\frac{\exp(f_t^*)}{[1+\exp(f_t^*)]^2} \ . \end{aligned} \tag{2.9}$$

   A Equação (3.3) representa a expansão de $g^{-1}(\lambda_t)$ em série de Taylor de primeira ordem ao redor de f_t^*. Daí segue que

$$E(\mu_t|D_t,\phi) = E\left(\frac{\exp(\lambda_t)}{1+\exp(\lambda_t)}\ |D_t,\phi\right) \approx \frac{\exp(f_t^*)}{1+\exp(f_t^*)} = \tilde{\mu}_t \ ; \tag{2.10a}$$
$$V(\mu_t|D_t,\phi) = V\left(\frac{\exp(\lambda_t)}{1+\exp(\lambda_t)}\ |D_t,\phi\right) \approx \left(\frac{\exp(f_t^*)}{[1+\exp(f_t^*)]^2}\right)^2 q_t^* = \tilde{V}_t \ . \tag{2.10b}$$

   Resolvendo em f_t^* e q_t^* as Equações (2.10), obtém-se:

$$f_t^* = \log\left(\frac{\tilde{\mu}_t}{1-\tilde{\mu}_t}\right)$$
$$q_t^* = \frac{\tilde{V}_t}{(\tilde{\mu}_t)^2}\left[\frac{1}{1-\tilde{\mu}_t}\right]^2$$

5. *Atualização dos parâmetros de* θ_t

   Neste momento já se tem todas as informações necessárias para completar o ciclo de atualização. Basta computar, usando-se as Equações (2.7) e (2.8), os

momentos m_t e C_t dados por

$$m_t = a_t + R_t F_t (f_t^* - f_t)/q_t \ ;$$
$$C_t = R_t - R_t F_t F_t' R_t (1 - q_t^*/q_t)/q_t \ .$$

A atualização dos parâmetros do sistema, descrita na sequência de passos apresentada acima, viabiliza o processo de previsão. A distribuição preditiva um passo a frente é dada por

$$p(y_t|D_{t-1}, \phi) = \int_0^1 p(y_t|\mu_t, D_{t-1}, \phi) p(\mu_t|D_{t-1}) d\mu_t \ ,$$

que não pode, em geral, ser calculada analiticamente. Todavia, seus primeiros momentos são conhecidos,

$$\begin{aligned}
E(Y_t|D_{t-1}, \phi) &= E[E(Y_t|\mu_t, D_{t-1}, \phi)] = \frac{r_t}{r_t + s_t} \ ; \\
V(Y_t|D_{t-1}, \phi) &= E[V(Y_t|\mu_t, D_{t-1}, \phi)] + V[E(Y_t|\mu_t, D_{t-1}, \phi)] \\
&= E[(\mu_t - \mu_t^2)/(1 + \phi)|D_{t-1}, \phi] + V[\mu_t|D_{t-1}, \phi] \\
&= \frac{1}{1+\phi} \left[\frac{r_t}{r_t + s_t} \left(1 - \frac{r_t}{r_t + s_t} \right) + \frac{\phi r_t s_t}{(r_t + s_t)^2 (r_t + s_t + 1)} \right] .
\end{aligned}$$

Tais momentos são extremamente úteis para a comparação do desempenho preditivo de modelos candidatos. Além disso, eles permitem o monitoramento da adequabilidade do modelo ao longo do tempo, identificando mudanças significativas no cenário que não tenham sido antecipadas por especialistas.

Naturalmente, em aplicações práticas, os elementos F_t, G_t, W_t e ϕ_t, em especial os dois últimos, não são conhecidos. Para lidar com eles, da-Silva et al. (2011) propõem uma abordagem *online*, recursiva. Neste caso, as matrizes W_t são especificadas via fator de desconto em blocos, discutidos na Subseção (1.5.1). ϕ_t, por sua vez, é tratado como um parâmetro desconhecido com distribuição a priori Gama(α, β). Em consequência, a atualização da distribuição de μ_t envolve a marginalização da distribuição conjunta $(\mu_t, \phi_t|D_t)$. Em outras palavras, é preciso integrar, numericamente, esta distribuição em relação a ϕ_t. O mesmo ocorre quanto à distribuição preditiva.

Capítulo 3

Modelo dinâmico Dirichlet

No Capítulo 2, discutimos modelos para análise de séries temporais univariadas restritas ao intervalo $(0, 1)$, em que os dados observados referem-se a taxas, proporções ou porcentagens. Em estatística, tais situações surgem naturalmente em fenômenos em que se deseja monitorar a proporção da ocorrência de uma dada categoria de resposta. Por exemplo, a proporção de adultos que estão empregados em uma população ou a proporção do orçamento estadual destinado à área da saúde. Nesse caso, a série temporal observada é univariada.

Voltamos agora nossa atenção a situações em que k categorias de resposta são possíveis. No caso das taxas de desemprego, as seguintes categorias: empregado na iniciativa privada, empregado no setor público, desempregado e outros. No caso da partição do orçamento estadual, as seguintes categorias: saúde, educação, segurança, transporte e outros. Esses exemplos tratam de um problema multivariado onde o resultado observado, a cada tempo, é um vetor de proporções (dados composicionais) e não mais um escalar.

Apresentamos, neste capítulo, o Modelo Dinâmico Dirichlet (MDD), representando uma contribuição inédita para descrever séries temporais multivariadas expressas em termos de dados composicionais. O modelo em questão foi formulado de acordo com a classe dos Modelos Dinâmicos Lineares Generalizados, o que confere ganhos quanto à flexibilidade na incorporação de efeitos importantes no estudo de alguns fenômenos quando comparado ao modelo desenvolvido por Grunwald et al. (1993). Além disso, o MDD foi implementado de modo a permitir a estimação de

todos os parâmetros desconhecidos do modelo em um paradigma genuinamente Bayesiano. Este trabalho representa uma extensão ao modelo desenvolvido por da-Silva et al. (2011).

3.1 Modelo Dinâmico Dirichlet - Parâmetros conhecidos

Seja Y_t, $t = 1, 2, \ldots$, uma série temporal k-variada de dados composicionais, o Modelo Dinâmico Dirichlet é definido pela seguinte estrutura probabilística:

- **Equação das observações:**

$$(Y_t|\mu_t, \phi) \sim \text{Dirichlet}(\mu_{1t}\phi, \mu_{2t}\phi, \ldots, \mu_{kt}\phi) \text{ , isto é:}$$

$$p(y_t|\mu_t, \phi) = \frac{\Gamma(\phi)}{\prod_{i=1}^{k}\Gamma(\phi\mu_{it})}\prod_{i=1}^{k} {y_{it}}^{\phi\mu_{it}-1} \text{ , onde:}$$

$$\mu_{kt} = 1 - \sum_{i=1}^{k-1}\mu_{it}, \quad y_{kt} = 1 - \sum_{i=1}^{k-1} y_{it}; \quad 0 < y_{it}, \mu_{it} < 1 \quad \forall i;$$

$$\phi > 0 \quad \text{e} \quad \mu_t = (\mu_{1t}, \ldots, \mu_{k-1,t}).$$

- **Distribuição a priori:**

$$(\mu_t|D_{t-1}) \sim \text{LN}_{k-1}(\eta_t, \Sigma_t)$$

- **Função de ligação: logito aditiva**

$$\lambda_t = F_t'\theta_t = \left[\log\left(\frac{\mu_{1t}}{\mu_{kt}}\right), \ldots, \log\left(\frac{\mu_{k-1,t}}{\mu_{kt}}\right)\right]'$$

- **Equação de evolução do sistema:**

$$\theta_t = G_t\theta_{t-1} + w_t \text{ ; } \quad w_t \sim (0, W_t)$$

- **Informação inicial:**

$$(\theta_0|D_0) \sim (m_0, C_0)$$

Denota-se por $\text{LN}_{k-1}(\eta_t, \Sigma_t)$ a distribuição Logística-normal parametrizada por η_t e Σ_t.

Neste trabalho, ao parâmetro de precisão ϕ não permitimos dinâmica, sendo esse um tópico para trabalhos futuros. Observe que, para viabilizar o uso do estimador *linear bayes* na atualização de θ_t, as distribuições de $(\theta_0|D_0)$ e w_t, para $t = 0, \ldots$, foram especificadas apenas parcialmente por seus momentos.

Como já mencionado, o modelo acima é uma generalização do Modelo Dinâmico Beta apresentado na Seção 2.2. Uma diferença importante está na distribuição a priori de μ_t. Enquanto lá se adotou a distribuição Beta, aqui a priori de μ_t é definida por uma distribuição Logística-normal. Equivalentemente, pode-se dizer que aqui foi adotada uma priori Normal, na escala do preditor linear λ_t, como será visto adiante. Nas diversas simulações feitas ao longo deste trabalho de pesquisa, os ajustes sempre se mostraram muito parecidos, qualquer que fosse a priori utilizada, quer seja Logística-normal ou Beta. A razão pela qual isto acontece está, justamente, na grande flexibilidade destas duas distribuições. A título de exemplo, no caso específico da distribuição Beta, um par (η_t, Σ_t) adequadamente escolhido como parâmetro da distribuição Logística-normal, tornará as duas densidades parecidas.

Antes de seguirmos aos procedimentos de estimação dos parâmetros, vale elencar algumas propriedades das distribuições envolvidas.

A distribuição das observações, Dirichlet, é a generalização multivariada da distribuição Beta. A primeira distribuição é conjugada à distribuição Multinomial e tem suporte definido no $(k-1)$-simplex aberto padrão. Formalmente, o conjunto de pontos em que a função de densidade é diferente de zero é expresso por

$$\Delta_{k-1} = \{(y_1, \ldots, y_{k-1}) \in \mathbb{R}^{k-1} | \sum_{i=1}^{k-1} y_i < 1 \,\text{e}\, y_i > 0, \forall i\}.$$

Utilizando essa parametrização, os momentos da distribuição Dirichlet são dados por:

$$\begin{aligned} E(y_t|\mu_t,\phi) &= \mu_t = (\mu_{1t},\ldots,\mu_{k-1,t})' \\ V(y_t|\mu_t,\phi) &= \frac{1}{1+\phi}[\text{diag}(\mu_t) - \mu_t\mu_t'] \end{aligned} \tag{3.1}$$

De todas as propriedades dessa família de distribuições, a estrutura da matriz de covariâncias é que melhor justifica a nossa proposta de modelo dinâmico Dirichlet. Se a média é conhecida, ou especificada, só resta ϕ como parâmetro livre. Tal simplicidade é extremamente útil, uma vez que ϕ pode ser estimado facilmente via simulação estocástica MCMC. Mais adiante serão dados maiores esclarecimentos sobre benefícios dessa estrutura simples da Equação (3.1).

Adotamos como priori de μ_t, a distribuição Logística-normal (Aitchison e Shen, 1980). Também definida no simplex, essa distribuição atende a todas as condições necessárias para completar o ciclo inferencial de atualizações do modelo apresentado. Sendo $(\mu_t|D_{t-1}) \sim \text{LN}_{k-1}(\eta_t, \Sigma_t)$, sua densidade é dada por:

$$p(\mu_t|D_{t-1}) = \left(\frac{1}{2\pi}\right)^{(k-1)/2} |\Sigma_t|^{-1/2} \left(\frac{1}{\prod_{i=1}^{k}\mu_{it}}\right) \exp\left[-\frac{1}{2}(z_t-\eta_t)'\Sigma_t^{-1}(z_t-\eta_t)\right],$$

onde μ_t pertence ao conjunto dado por

$$\begin{aligned} &\{(\mu_{1t},\ldots,\mu_{k-1,t}) \in \mathbb{R}^{k-1} \,|\, \sum_{i=1}^{k-1}\mu_{it} < 1 \;\text{ e }\; \mu_{it} > 0, \forall i\} \quad \text{e} \\ z_t = \text{alz}(\mu_t) &= \left[\log\left(\frac{\mu_{1t}}{\mu_{kt}}\right),\ldots,\log\left(\frac{\mu_{k-1,t}}{\mu_{kt}}\right)\right] = (z_{1t},\ldots,z_{k-1,t})', \end{aligned}$$

onde $\mu_t \in \Delta_{k-1}$ e alz$(\cdot)$ descreve a transformação de Δ_{k-1} em $\mathbb{R}^{k-1}$, representando, também, a função de ligação multivariada usada em nosso modelo. É importante salientar que μ_{kt} não é um parâmetro livre, e pode ser escrito deterministicamente como função de $(\mu_{1t},\ldots,\mu_{k-1,t})$.

Nesses termos, é fácil mostrar que o vetor aleatório Z_t tem distribuição $N_{k-1}(\eta_t, \Sigma_t)$.

3.1.1 Procedimento inferencial *online*

As etapas do processo de estimação *online* dos parâmetros do sistema são basicamente as mesmas do MDLG e estão detalhadas a seguir.

1. *Obtenção dos momentos a priori de θ_t e λ_t*

   Como de costume, temos as seguintes distribuições a priori parcialmente especificadas por $(\theta_t|D_{t-1},\phi) \sim (a_t, R_t)$ e $(\lambda_t|D_{t-1},\phi) \sim (f_t, Q_t)$, onde:

$$a_t = G_t m_{t-1}, \quad R_t = G_t C_{t-1} G_t' + W_t, \quad f_t = F_t' a_t \quad \text{e} \quad Q_t = F_t' R_t F_t.$$

2. *Elicitação dos parâmetros da priori de μ_t*

   Combinando os resultados já apresentados, é possível estabelecer facilmente que

$$\begin{aligned} f_t &= E(\lambda_t|D_{t-1},\phi) = E(\text{alz}(\mu_t)|D_{t-1},\phi) = E(z_t|D_{t-1},\phi) = \eta_t \ ; \\ Q_t &= V(\lambda_t|D_{t-1},\phi) = V(\text{alz}(\mu_t)|D_{t-1},\phi) = V(z_t|D_{t-1},\phi) = \Sigma_t \ . \end{aligned}$$

   As igualdades acima evidenciam que os momentos da priori de λ_t são rigorosamente iguais aos parâmetros, mas não aos momentos, da priori de μ_t. Esse resultado, inicialmente surpreendente, mas agora óbvio, torna extremamente simples a execução desta etapa da elicitação da priori de μ_t. Recorde que, mesmo no caso univariado, modelo Beta da Seção 2.2, esta etapa não era simples, sendo necessário utilizar propriedades da família exponencial, ou expansão em série de Taylor, para resolver o sistema e determinar os parâmetros r_t e s_t da priori Beta. Adotando-se uma priori Logística-normal, tem-se um resultado exato e imediato, não sendo necessário o uso de aproximações do tipo $\psi(z) \approx \log(z)$ e $\psi'(z) \approx 1/z$ que, a depender dos valores de r_t e s_t, podem não ser inadequadas.

3. *Atualização dos parâmetros de μ_t*

   A distribuição a posteriori de μ_t é obtida pelo *Teorema de Bayes*,

$$p(\mu_t|D_t,\phi) \propto p(y_t|\mu_t, D_{t-1},\phi) p(\mu_t|D_{t-1}). \tag{3.2}$$

   Precisamos obter os momentos de $(\mu_t|D_t,\phi)$ de modo a atualizar, posterior-

mente, os momentos de $(\lambda_t|D_t, \phi)$. Como tal posteriori não tem forma fechada, seus momentos de primeira e segunda ordem, $\tilde{\mu}_t$ e $\tilde{V}_t$, devem ser calculados numericamente. Para tanto, sugerimos um procedimento composto por duas etapas.

Etapa 1. A região de integração é definida pelo $(k-1)$-simplex aberto, ou seja, pelo conjunto $\Delta_{k-1} = \{(\mu_1, \ldots, \mu_{k-1}) \in \mathbb{R}^{k-1} | \sum_{i=1}^{k-1} \mu_i < 1 \, \text{e} \, \mu_i > 0, \forall i\}$. É possível, e útil, fazer uma mudança de coordenadas a fim de transformar a região de integração em um hipercubo unitário $(0,1)^{k-1}$, formalmente escrito como $\Omega_{k-1} = \{(x_1, \ldots, x_{k-1}) \in \mathbb{R}^{k-1} \mid 0 < x_1, \ldots, x_{k-1} < 1\}$. Adotamos, então, a seguinte transformação:

$$x_i = \begin{cases} \sum_{i=1}^{k-1} \mu_i, & \text{para} \quad i = 1 \\ \mu_{i-1} / \sum_{j=i-1}^{k-1} \mu_j, & \text{para} \quad 2 \leq i \leq k-1, \end{cases}$$

cujo jacobiano tem módulo dado por

$$J = \prod_{i=2}^{k-2} (1 - x_i)^{k-1-i}.$$

No *R*, a referida transformação está implementada na função *p_to_e*, enquanto o jacobiano pode facilmente ser calculado usando a função *Jacobian*. Vale destacar que essas duas funções são do pacote *hyperdirichlet* (Hankin, 2010).

Etapa 2. Agora, com as variáveis já transformadas, podemos proceder à integração no hipercubo usando a técnica de integração multidimensional adaptativa, também conhecida como *cubatura*. Tal método é, por vezes, referenciado como a generalização, para o contexto multivariado, do método da *quadratura*. A função *adaptIntegrate*, do pacote *cubature* (Johnson e Narasimhan, 2009), é muito atrativa por permitir o cálculo da integral de vetores.

Para completar a tarefa de estimação de $\tilde{\mu}_t$ e $\tilde{V}_t$, precisamos calcular um conjunto de integrais: uma delas para encontrar a constante de normalização da posteriori de μ_t, $k-1$ integrais para calcular as médias das coordenadas de μ_t, e $k(k-1)/2$ para calcular os momentos $E(\mu_i \mu_j | D_t)$, $i, j = 1, \ldots, k-1$. En-

tretanto, utilizando-se a função *adaptIntegrate*, todas essas integrais são computadas calculando-se uma única vez os valores da função dada em (3.2). Essa ferramenta torna bastante eficiente o processo de integração. De qualquer forma, a magnitude do problema, mensurado pela quantidade de pontos a serem calculados para que se tenha uma dada precisão, cresce exponencialmente em função do número k de categorias possíveis para Y_t.

É sabido que o *R* não é um *software* especialmente rápido. Uma possibilidade interessante para diminuir o tempo computacional seria realizar este cálculo em linguagem *C*, de dentro do programa *R*. Para que se tenha idéia, em um computador doméstico regular, a integração leva aproximadamente 0, 1, 5 e 30 segundos para os casos $k = 2$, 3, 4 e 5, respectivamente. Para o caso univariado, correspondente ao modelo Beta, adotamos o mesmo método de integração, porém utilizando a função *integrate*.

4. *Atualização dos parâmetros de* λ_t

   Nesta etapa o nosso foco está na obtenção dos parâmetros f_t^* e Q_t^* satisfazendo a

$$E(\lambda_t|D_t,\phi) = E(\text{alz}(\mu_t)|D_t,\phi) = f_t^* \ ;$$
$$V(\lambda_t|D_t,\phi) = V(\text{alz}(\mu_t)|D_t,\phi) = Q_t^* \ .$$

   Observando que, pela função de ligação,

$$\mu_t = \text{alz}^{-1}(\lambda_t) = \frac{1}{1+\sum_{i=1}^{k-1} e^{\lambda_{it}}}(e^{\lambda_{1t}},\ldots,e^{\lambda_{k-1,t}})', \tag{3.3}$$

   é possível encontrar, de maneira aproximada, os parâmetros a posteriori de λ_t expandindo em série de Taylor de primeira ordem, em torno de f_t^*, a função $\text{alz}^{-1}(\lambda_t)$. Assim, temos:

$$\text{alz}^{-1}(\lambda_t) \approx \text{alz}^{-1}(f_t^*) + \bigtriangledown[\text{alz}^{-1}(f_t^*)](\lambda_t - f_t^*), \tag{3.4}$$

onde o símbolo $\bigtriangledown$ denota o gradiente da função, que, nesse caso, é dado por:

$$\bigtriangledown = \bigtriangledown[\text{alz}^{-1}(f_t^*)] = \frac{1}{(1+\beta'\mathbf{1}_{k-1})^2}[\text{diag}(\beta) + \text{diag}(\beta\mathbf{1}'_{k-1}\beta) - \beta\beta']. \quad (3.5)$$

O vetor coluna $\beta = (\beta_1, \ldots, \beta_{k-1})' = (e^{f_{1t}^*}, \ldots, e^{f_{k-1,t}^*})'$, é usado aqui somente para simplificar a notação. O termo $\mathbf{1}_{k-1}$ representa o vetor coluna de dimensão $k-1$, tal que todas as suas entradas são iguais a 1. Portanto, temos as equações que relacionam os momentos a posteriori de μ_t e λ_t:

$$\begin{aligned}
\tilde{\mu}_t = E(\mu_t|D_t) &= E(\text{alz}^{-1}(\lambda_t)|D_t) \\
&\approx E(\text{alz}^{-1}(f_t^*) + \bigtriangledown(\lambda_t - f_t^*)|D_t) \\
&= E(\text{alz}^{-1}(f_t^*)|D_t) + E(\bigtriangledown(\lambda_t - f_t^*)|D_t) \\
&= \text{alz}^{-1}(f_t^*)\ ; \quad (3.6) \\
\tilde{V}_t = V(\mu_t|D_t) &= V(\text{alz}^{-1}(\lambda_t)|D_t) \\
&\approx V(\text{alz}^{-1}(f_t^*) + \bigtriangledown(\lambda_t - f_t^*)|D_t) \\
&= \bigtriangledown V(\lambda_t|D_t)\bigtriangledown' \\
&= \bigtriangledown Q_t^* \bigtriangledown\ . \quad (3.7)
\end{aligned}$$

Finalmente, resolvendo as Equações 3.6 e 3.7 em função de f_t^* e Q_t^*, tem-se que $f_t^* = \text{alz}(\tilde{\mu}_t)$ e $Q_t^* = \bigtriangledown^{-1}\tilde{V}_t\bigtriangledown^{-1}$. A distribuição a posteriori de λ_t é aproximada por:

$$(\lambda_t|D_t) \sim (\text{alz}(\tilde{\mu}_t), \bigtriangledown^{-1}\tilde{V}_t\bigtriangledown^{-1}).$$

Para o modelo dinâmico Beta, da-Silva, Migon e Correia (2011) avaliam a qualidade da aproximação baseada na expansão em série de Taylor. A partir da comparação entre as expansões de primeira e segunda ordem, eles concluem que o ganho da inclusão do termo quadrático não é significativo para um grande leque de aplicações. No contexto multivariado, a inclusão do segundo termo complicaria demasiadamente a atualização de λ_t.

5. *Atualização dos parâmetros de θ_t*

A exemplo do caso univariado, para a atualização dos momentos de θ_t, recorremos ao método de estimação via *Linear Bayes*. A distribuição a priori conjunta de (λ_t, θ_t) é parcialmente especificada por

$$\begin{pmatrix} \lambda_t \\ \theta_t \end{pmatrix} \Bigg| \; D_{t-1} \sim \left[\begin{pmatrix} f_t \\ a_t \end{pmatrix}, \quad \begin{pmatrix} Q_t & F_t' R_t \\ R_t F_t & R_t \end{pmatrix} \right]$$

Por outro lado, a distribuição conjunta a posteriori pode ser escrita convenientemente como mostra a Equação (3.8).

$$\begin{aligned} p(\lambda_t, \theta_t | D_t) &= p(\lambda_t, \theta_t | D_{t-1}) p(Y_t | \lambda_t) / p(Y_t | D_{t-1}) \\ &= [p(\theta_t | \lambda_t, D_{t-1}) p(\lambda_t | D_{t-1})] p(Y_t | \lambda_t) / p(Y_t | D_{t-1}) \\ &= p(\theta_t | \lambda_t, D_{t-1}) [p(\lambda_t | D_{t-1}) p(Y_t | \lambda_t) / p(Y_t | D_{t-1})] \\ &= p(\theta_t | \lambda_t, D_{t-1}) p(\lambda_t | D_t). \end{aligned} \tag{3.8}$$

A combinação de tais estruturas, juntamente com as propriedades de esperança e variância condicionais, nos permite passar adiante a informação contida em y_t aos parâmetros do sistema. Os momentos a posteriori de $(\theta_t | D_t)$, m_t e C_t, podem ser estimados via linear bayes, por:

$$\begin{aligned} m_t &= E(\theta_t | D_t) = E[E(\theta_t | \lambda_t, D_{t-1}) | D_t] \\ &\approx E[\hat{E}(\theta_t | \lambda_t, D_{t-1}) | D_t] \\ &= E[a_t + R_t F_t Q_t^{-1} (\lambda_t - f_t) | D_t] \\ &= a_t + R_t F_t Q_t^{-1} (f_t^* - f_t), \end{aligned}$$

e

$$\begin{aligned} C_t &= V(\theta_t | D_t) = V[E(\theta_t | \lambda_t, D_{t-1}) | D_t] + E[V(\theta_t | \lambda_t, D_{t-1}) | D_t] \\ &\approx V[a_t + R_t F_t Q_t^{-1} (\lambda_t - f_t) | D_t] + E[R_t - R_t F_t Q_t^{-1} F_t' R_t | D_t] \\ &= V[R_t F_t Q_t^{-1} \lambda_t | D_t] + R_t - R_t F_t Q_t^{-1} F_t' R_t \\ &= R_t - R_t F_t Q_t^{-1} [I - Q_t^* Q_t^{-1}] F_t' R_t. \end{aligned}$$

A sequência de passos envolvida no procedimento inferencial *online*, acima descrita, viabiliza não somente o processamento de novos dados, como também o processo de previsão. A distribuição preditiva um passo a frente é dada por

$$p(y_t|D_{t-1},\phi) = \int_{\Delta} p(y_t|\mu_t, D_{t-1}, \phi)p(\mu_t|D_{t-1})d\mu_t \ ,$$

e não pode ser calculada analiticamente. Vale lembrar que a região de integração, denotada por Δ_{k-1}, consiste no $(k-1)$-simplex aberto padrão. O procedimento de integração em duas etapas, usado na atualização dos parâmetros de μ_t, pode ser empregado na integração numérica da distribuição preditiva.

Como subproduto deste trabalho de pesquisa, implementamos, embora ainda não seja de domínio público, uma rotina em *R* capaz de executar toda a sequência de cálculos citados no processo de filtragem do MDD. Batizada de *mddFilter*, a função desempenha, para os modelos dinâmicos Dirichlet, o mesmo papel da função *dlmFilter* em relação aos MLD. Ambas foram construídas a partir de uma mesma estratégia, embora algumas diferenças precisam ser citadas:

• A rotina *mddFilter* pode ser usada tanto quando W é conhecida, quanto na situação em que esta é especificada via fatores de desconto em blocos. Neste último caso, é preciso incluir entre os *inputs* da função um vetor contendo os fatores de desconto δ e um vetor, de mesmo tamanho, com a dimensão de cada bloco do sistema. Como exemplo, para um modelo polinomial de ordem 3 e 5 períodos sazonais, com fatores de desconto de 0.9 e 0.95, respectivamente, entre os inputs da função estariam os objetos `disc = c(0.9, 0.95)` e `dim.disc = c(3, 4)`. Quando tais objetos são omitidos, a função *busca* a matriz W no objeto `mod$W`.

• Outra importante diferença está na forma de lidar com o parâmetro ϕ. Este deve ser armazenado fora do objeto `mod`. Por esse motivo, o objeto `mod$V`, caso esteja definido, é ignorado no processo de filtragem.

• O usuário pode, se desejar, mudar o erro máximo do processo de integração. Esse valor é internamente passado adiante para a função *adaptIntegrate*. O *default* é 0.01. A função é bem robusta e trata, adequadamente, as observações faltantes (mas não as incompletas) com o procedimento descrito na subseção (1.3.1). Os objetos retornados são os mesmos da função *dlmFilter*, além dos inputs utilizados e das

matrizes W_t (também na forma decomposta em valores singulares), caso tenha sido usada a abordagem via fatores de desconto.

• As fórmulas de suavização da Subseção 1.3.2 são aplicáveis ao MDD. A função *dlmSmooth*, apesar de funcionar perfeitamente quando W é conhecida, não foi construída para o caso em que o processo de filtragem envolve a especificação via FD. Para contornar o problema, a partir de uma pequena adaptação, construímos a função *mddSmooth* capaz de lidar adequadamente com ambas as situações.

• As especificações das matrizes F_t e G_t seguem os mesmos princípios do caso normal, embora estejam sujeitas a certas particularidades. Os exemplos apresentados no Capítulo 4 exploram, com mais detalhes, o assunto. No caso multivariado, o produto de *Kronecker* costuma ser bastante útil na representação compacta das matrizes em questão.

O modelo de Grunwald et al. (1993) é menos flexível do que o proposto neste trabalho pois:

(1) É baseado em uma priori do tipo *"power steady"* (Smith, 1979), cuja característica principal é que a moda de $(\lambda_t|D_{t-1})$ é a mesma de $(\lambda_{t-1}|D_{t-1})$, porém, com maior dispersão;

(2) No processo inferencial, utiliza-se tanto os métodos Bayesianos quanto os métodos clássicos, o que é um tanto desconcertante;

(3) A descrição dos efeitos nos parâmetros do processo latente θ_t, tais como tendência, crescimento, sazonalidade e inclusão de covariáveis é bastante complicado, contrariamente às inúmeras possibilidades disponíveis mediante as especificações das matrizes F e G.

3.2 Modelo dinâmico Dirichlet - Parâmetros desconhecidos

Na Seção 3.1 apresentamos uma nova classe de modelos para dados composicionais, detalhando a metodologia de estimação recursiva, no caso dos componentes F_t, G_t, W e ϕ serem todos conhecidos. Frequentemente, as matrizes F_t e G_t são determinadas

integralmente, só restando W e ϕ na especificação do modelo. Neste caso, uma possibilidade prática, ainda que por vezes um tanto imprecisa, é especificar W via fator de descontos e *escolher* ϕ de modo a otimizar, sob algum critério, o desempenho preditivo observado do modelo. Isto é, calcula-se o vetor $(\widetilde{\delta}, \widetilde{\phi})$ que, de certo modo, melhor descreve o comportamento passado da série histórica. Esses valores, então, determinam W e ϕ, que passam a ser tratados como quantidades conhecidas, o que viabiliza a aplicação do procedimento inferencial descrito. É preciso ter em mente que a incerteza relativa a esses elementos não é devidamente considerada. Além disso, como tal determinação de δ e ϕ é feita *"ad-hoc"*, a adoção de um processo paralelo de controle, baseado nos erros de previsão, é útil na detecção de uma possível deterioração do desempenho preditivo do modelo.

Por mais simplista que possa parecer, a estratégia descrita no parágrafo anterior se mostrou muitas vezes razoável para a previsão, filtragem e suavização. Percebemos, em diversos casos simulados, que ao se fixar um valor $\widetilde{\phi}$ não muito próximo ao valor real de ϕ, a perda de qualidade das predições não foi apreciável.

Nesta seção, introduzimos um modelo dinâmico Dirichlet diferente, em certos aspectos, do apresentado anteriormente. A nova formulação tem por objetivo viabilizar a abordagem de estimação *offline* via simulação estocástica MCMC, aplicável a situações mais complexas e abrangentes. Sugerimos a releitura da Subseção 1.5.2 que, embora seja aplicável aos modelos normais, guarda bastante similaridade com o caso geral. Uma vez exposta a nova formulação do modelo, discutiremos como gerar amostras da posteriori e, a partir delas, simular valores da distribuição preditiva. A simulação relativa a cada um dos elementos integrantes do modelo dinâmico também será apresentada.

Seja Y_t, $t = 1, \ldots, T$, uma série temporal k-variada, o modelo dinâmico Dirichlet usado na abordagem *offline* é definido, para $t = 1, \ldots, T$, pelas seguintes equações:

- **Equação das observações:**

$$(Y_t|\mu_t, \phi) \sim \text{Dirichlet}(\mu_{1t}\phi, \mu_{2t}\phi, \ldots, \mu_{kt}\phi), \tag{3.9a}$$

$$\text{onde:} \quad \mu_{kt} = 1 - \sum_{i=1}^{k-1} \mu_{it}, \quad y_{kt} = 1 - \sum_{i=1}^{k-1} y_{it}; \tag{3.9b}$$

$$0 < y_{it}, \mu_{it} < 1 \quad \forall i, \quad \phi > 0 \quad \text{e} \quad \mu_t = (\mu_{1t}, \ldots, \mu_{k-1,t}).$$

- **Função de ligação: logito aditiva**

$$\lambda_t = F_t'\theta_t = \left[\log\left(\frac{\mu_{1t}}{\mu_{kt}}\right), \ldots, \log\left(\frac{\mu_{k-1,t}}{\mu_{kt}}\right)\right]' \tag{3.9c}$$

- **Equação de evolução do sistema:**

$$\theta_t = G_t\theta_{t-1} + w_t \; ; \quad w_t \sim N(0, W_t) \tag{3.9d}$$

- **Distribuição a priori:**

$$(\Theta, \psi) \sim p(\Theta, \psi), \tag{3.9e}$$

onde: $\Theta = (\theta_1, \ldots, \theta_T)$ e ψ é o vetor contendo as demais quantidades desconhecidas do modelo.

Do ponto de vista teórico e operacional, a estrutura do modelo dinâmico Dirichelt (*offline*), representado pelas Equações (3.9), apresenta diferenças substanciais em relação ao MDD (*online*), dado na Seção 3.1. Enquanto no segundo a elicitação da priori é feita a cada tempo, à medida que as observações y_t são sequencialmente processadas, no primeiro é preciso atribuir uma distribuição a priori conjunta para todos os parâmetros desconhecidos, (Θ, ψ), independente dos valores y_t.

O vetor ψ pode assumir diferentes formatos, a depender do modelo em consideração. O caso mais típico possivelmente sendo aquele em que $\psi = (W, \phi, \theta_0)$. Dessa forma, a distribuição a priori dada na Equação (3.9e) precisa ser especificada respeitando-se as particularidades do fenômeno estudado e a estrutura imposta ao modelo. A Subseção 3.2.1 discute, em detalhes, como definir esta priori e como amostrar

da distribuição a posteriori expressa por

$$p(\psi, \Theta|D_T) \propto p(D_T|\Theta, \psi)p(\Theta, \psi).$$

Outra diferença significativa entre os modelos diz respeito à distribuição dos erros do sistema (w_t). Na abordagem recursiva, essas distribuições foram especificadas apenas parcialmente. Por outro lado, no MDD *offline*, assumimos normalidade para w_t. Note que, equivalentemente, assumimos, para $t = 1, \ldots, T$,

$$p(\theta_t|\theta_{t-1}, W) \sim N(G_t\theta_{t-1}, W).$$

Dessa forma, se a distribuição a priori dada na Equação (3.9e) for definida como

$$p(\Theta, \psi) = [p(\Theta|\theta_0, W)\ p(\theta_0)]\, p(W, \phi), \tag{3.10}$$

a suposição de normalidade para a priori de θ_0, isto é $\theta_0 \sim N(m_0, C_0)$, é suficiente para que $\mu_t = \text{alz}^{-1}(F_t\theta_t) \sim LN_{k-1}(\eta_t^*, \Sigma_t^*)$.

Assumir normalidade da distribuição a priori $(\theta_0|D_0)$ não traz qualquer prejuízo prático relevante. Muitas vezes as prioris são vagas, e quando não o são, é possível escolher parâmetros m_0 e C_0 que tornem compatível a informação inicial do analista com a distribuição normal escolhida.

A suposição sobre os erros de evolução do sistema w_t é mais delicada. Nenhuma distribuição de probabilidade é tão usada quanto a Normal na modelagem estatística de erros. Esse fato resulta de uma série de propriedades importantes dessa distribuição. Pelo teorema do limite central, podemos entender uma variável Normal como a soma de infinitos fatores independentes que *per se* não tem significância, mas conjuntamente provocam um efeito relevante. No contexto dos modelos dinâmicos, os erros são contínuos, frequentemente simétricos.

Entendemos que a suposição de normalidade dos erros w_t é historicamente amparada e justificável em diversas aplicações práticas. Isso não significa, contudo, que será sempre válida. Por vezes ficará clara a inadequação desta premissa. Do mesmo modo, assumir que os dados composicionais Y_t seguem uma distribuição Dirichlet, ou que os parâmetros do sistema se relacionam à média via ligação logito, pode ser

igualmente equivocado. Ou seja, qualquer modelo paramétrico tem suposições, fracas ou fortes, que podem não ser compatíveis com o fenômeno.

Argumentamos que a tríplice (observações Dirichlet, função de ligação logito e erros Normais) constitui uma base natural, sólida e ampla para a modelagem de dados composicionais. O modelo assim construído é ainda muito flexível e pode assumir diversos formatos. Será visto, adiante, que tanto se pode usá-lo para a regressão Dirichlet estática como para a análise de séries temporais de dados composicionais, por exemplo.

Em resumo, no modelo da Seção 3.1, definimos, a cada instante t, a distribuição a priori $p(\mu_t|D_{t-1}) \sim LN_{k-1}(\eta_t, \Sigma_t)$. Essa atribuição tem como propósito viabilizar o processamento da nova informação y_t, ou mais precisamente a atualização de θ_t. Na abordagem *offline*, por sua vez, a suposição de normalidade de θ_0 e w_t, em conjunto com a função de ligação logito aditiva, implica que, adotando-se a estratégia natural de definição da priori como em 3.10, o preditor linear λ_t é Normal e a média μ_t é Logística-normal, para $t = 1, 2, \ldots, T$.

Observe que, na abordagem *online*, a distribuição a priori para μ_t é **condicional** aos dados já processados $y_1, \ldots, y_{t-1}$. No modelo apresentado nesta seção, a priori de μ_t é construída sem que seja usada qualquer informação proveniente dos dados $y_1, \ldots, y_T$. Essa diferença, que está relacionada ao processo de estimação de cada modelo, precisa ficar clara. Não há recursividade na estimação do modelo *offline*. Isto é, quando novas informações se tornam disponíveis, o analista precisa reiniciar o processo de estimação. Ou seja, definir o modelo, incluindo a priori para os parâmetros desconhecidos, e usar o algoritmo MCMC para gerar amostras da posteriori. Assim, o fato de já haver sido feita uma análise anterior com parte dos dados não interfere no processo inferencial baseado na série completa $(y_1, \ldots, y_t)$.

3.2.1 Especificação da priori e estimação *offline* via MCMC

Nesta subseção será discutido em detalhes a estratégia de definição da priori dada na Equação (3.9e). Como mencionado anteriormente, ψ pode ou não incluir certos elementos. Como exemplos, se o estudo se refere a um cenário estático, W é uma matriz nula e, necessariamente, não estará em ψ. Por outro lado, se os dados incluem

observações faltantes, estas deverão estar contidas em ψ. Ou seja, na abordagem via simulação estocástica, cada caso requer um tratamento particular.

A fim de apresentar uma metodologia de simulação abrangente, será descrito ao longo dessa seção, entre os casos típicos, aquele mais completo. Consideraremos, especificamente, que $\psi = (\theta_0, \phi, W_t = W, Y_{\mathcal{A}^c})$. A terminologia $\mathcal{A}^{\mathcal{C}}$ representa o conjunto de índices t tais que Y_t é uma observação faltante. Denotemos, ainda, o conjunto dos tempos para os quais Y_t é conhecido por $\mathcal{A}$, e o vetor $(\theta_1, \ldots, \theta_T)$ por Θ. O vetor ψ pode facilmente acomodar outros cenários, simplesmente removendo-se, ou incluindo-se, alguns de seus elementos. Em particular, podemos citar modelos auto-regressivos, nos quais a matriz G_t contém parâmetros a serem estimados.

A estimação do modelo apresentado na Seção 3.2 envolve a simulação da distribuição a posteriori

$$p(\psi, \Theta | D_T) \propto p(D_T | \Theta, \psi) p(\Theta, \psi) \tag{3.11a}$$

$$= \prod_{t=\mathcal{A}} p(y_t | \theta_t, \phi) \prod_{t=\mathcal{A}^{\mathcal{C}}} p(Y_t | \theta_t, \phi) \tag{3.11b}$$

$$\times \prod_{t=1}^{T} p(\theta_t | \theta_{t-1}, W) p(W) p(\phi) p(\theta_0). \tag{3.11c}$$

É importante que fique claro o significado de cada um dos produtos de (3.11b). O primeiro retrata, exatamente, a verossimilhança. O segundo, por sua vez, corresponde à distribuição a priori das distribuições faltantes. Cabe destacar, ainda, que a equação (3.11c) não é unicamente determinada por (3.11a) e representa a situação em que W, ϕ e θ_0 são independentes a priori. Em (3.11c), as densidades do produtório são definidas pela equação de evolução (3.9d) e os termos restantes representam prioris que devem ser definidas pelo analista. Por fim, ressaltamos que será abordado somente o caso em que $W_t = W$, isto é, quando a matriz de covariâncias dos erros do sistema é a mesma para todo t.

A amostragem de (3.11a) é feita por meio do amostrador de Gibbs, resumido no algoritmo que se segue:

1. Inicialização: Defina valores iniciais $\psi = \psi^{(0)}$ e $\Theta = \Theta^{(0)}$.

2. Para $i = 1, \ldots, N$:

   (a) Amostre as observações faltantes $Y_{\mathcal{A}^c}^{(i)}$ de $p(Y_{\mathcal{A}^c}|\Theta = \Theta^{(i-1)}, \phi = \phi^{(i-1)})$

   (b) Amostre os estados do sistema $\Theta^{(i)}$ de
   $p(\Theta|D_T, Y_{\mathcal{A}^c} = Y_{\mathcal{A}^c}^{(i)}, \theta_0 = \theta_0^{(i-1)}, \phi = \phi^{(i-1)}, W = W^{(i-1)})$

   (c) Amostre a matriz de covariância dos erros $W^{(i)}$ de
   $p(W|\theta_0 = \theta_0^{(i-1)}, \Theta = \Theta^{(i)})$

   (d) Amostre o estado inicial $\theta_0^{(i)}$ de $p(\theta_0|\theta_1 = \theta_1^{(i)}, W = W^{(i)})$

   (e) Amostre o parâmetro de precisão $\phi^{(i)}$ de
   $p(\phi|D_T, Y_{\mathcal{A}^c} = Y_{\mathcal{A}^c}^{(i)}, \theta_0 = \theta_0^{(i)}, \Theta = \Theta^{(i)})$.

É possível mostrar que a distribuição estacionária das amostras (dependentes) geradas por esse mecanismo é justamente a distribuição conjunta a posteriori. A seguir será descrito, para cada item do algoritmo, como definir uma priori e executar a tarefa de amostrar da distribuição condicional completa.

3.2.1.1 Amostrando as observações faltantes

Como foi visto na Subseção 1.3.1, as observações faltantes são facilmente tratadas no processo de filtragem dos modelos dinâmicos. Na abordagem via MCMC, essas observações são entendidas como parâmetros desconhecidos e passam a compor o conjunto ψ, sendo a elas atribuída uma distribuição a priori.

De acordo com as Equações (3.11), a distribuição a priori conjunta $p(\Theta, \psi)$ é convenientemente definida como

$$p(\Theta, \psi) = p(Y_{\mathcal{A}^c}|\Theta, \phi)p(\Theta|\theta_0, W)p(W)p(\phi)p(\theta_0).$$

O primeiro termo do lado direito da igualdade pode ser naturalmente especificado pela equação das distribuições (3.9a). Ou seja, a distribuição a priori de $Y_{\mathcal{A}^c}$, induzida pela estrutura do modelo, é decomposta em $p(Y_{\mathcal{A}^c}|\Theta, \phi) = \prod_{t=\mathcal{A}^c} p(Y_t|\theta_t, \phi)$, onde cada termo do produtório segue uma distribuição Dirichlet com parâmetros definidos em função de θ_t e ϕ.

Novamente, considerando (3.11), tem-se que a distribuição condicional completa de $Y_{\mathcal{A}^C}$ é da forma:

$$p(Y_{\mathcal{A}^C}|\ldots) = \prod_{t=\mathcal{A}^C} p(Y_t|\theta_t, \phi). \tag{3.12}$$

Visando a simplificação da notação, de agora em diante usaremos nas distribuições condicionais completas o símbolo "..." para denotar a informação final D_T e todas as quantidades desconhecidas do modelo, exceto o próprio termo à esquerda da barra vertical usada para representar o condicionamento da distribuição.

É instrutivo notar que a Equação (3.12) não depende diretamente dos dados observados, ou seja, da função de verossimilhança. Ainda que esse resultado possa, em primeiro momento, parecer contra-intuitivo, ele é uma consequência imediata da propriedade (A.2) da página 16. Isto é, nos modelos de espaço de estado, que incluem os modelos dinâmicos Dirichlet, condicionalmente a θ, os Y_t's são independentes e Y_t depende apenas de θ_t.

A fim de facilitar o processo de amostragem das observações faltantes, criamos a função *R mddYSample*, ainda não disponibilizada, capaz de executar o passo (a) do algorítmo de Gibbs. Sua lista de *entradas* inclui os objetos:

(1) `mod`, especificado como no pacote *dlm*;

(2) Theta, a matriz $\Theta^{\cdot}$ tal que cada uma de suas linhas $i = 1, \ldots, T$ contém o respectivo vetor θ_i;

(3) `phi`, o parâmetro ϕ;

(4) `tempos`, o conjunto de índices $\mathcal{A}^C$.

A função *mddYSample*, que internamente faz uso da rotina *rdirichlet* do pacote *MCMCpack* (Martin et al., 2011), retorna uma matriz contendo as amostras geradas de $Y_{\mathcal{A}^C}$. Nela, a i-ésima linha corresponde ao valor gerado de Y_{k_i}, onde k_i representa o i-ésimo elemento de $\mathcal{A}^C$.

Como é possível gerar valores da distribuição condicional completa de forma fácil e direta, todas as amostras produzidas são aceitas. O analista pode salvar cada uma das amostras geradas caso queira fazer inferências sobre os dados faltantes. Se não

há interesse direto nessas quantidades, é possível apenas reciclar, a cada iteração, os valores de $Y_{\mathcal{A}^c}$, evitando, portanto, o armazenamento desnecessário dos dados simulados. De uma forma ou de outra, a amostragem dos demais elementos discutidos nas subseções que se seguem exige que as observações geradas para $Y_{\mathcal{A}^c}$ sejam gravadas na mesma matriz das observações, substituindo os valores faltantes, denotados por `NA`. Isto é, as observações $Y_{\mathcal{A}}$ e os valores correntes $Y_{\mathcal{A}^c}^{(i)}$ são passados adiante em um mesmo objeto.

3.2.1.2 Amostrando os estados do sistema - Caso Dinâmico

No que se refere à amostragem dos parâmetros do sistema, é preciso considerar, em separado, dois casos possíveis: dinâmico e estático. O primeiro deles será tratado nesta subseção, enquanto o segundo será discutido posteriormente na Subseção 3.2.1.3.

Da Equação (3.11), a distribuição condicional completa de Θ é dada por

$$p(\Theta|\ldots) \propto \prod_{t=1}^{T} p(y_t|\theta_t, \phi)p(\theta_t|\theta_{t-1}, W). \tag{3.13}$$

Os termos $p(y_t|\theta_t, \phi)$ seguem uma distribuição Dirichlet, em conformidade com a equação das observações. Note que, aqui, pouco importa se os valores y_t representam observações de fato ou valores correntes dos parâmetros desconhecidos $Y_{\mathcal{A}^c}^{(i)}$. Os termos $p(\theta_t|\theta_{t-1}, W)$, por sua vez, são normalmente distribuídos em decorrência da normalidade assumida para θ_0 na Subseção 3.2.1.5 e para w_t na Equação (3.9d). Desse modo, a distribuição dada em (3.13) não tem forma fechada e a simulação de amostras de Θ não é uma tarefa trivial. Além dos vários benefícios já mencionados da abordagem *offline*, temos a vantagem de poder, não somente conhecer os momentos, mas toda a distribuição a posteriori dos parâmetros do sistema.

Há, na literatura estatística, um conjunto de soluções possíveis para o problema de se gerar amostras para o sistema, no contexto dos modelos dinâmicos não-normais e não-lineares. Em especial, podemos citar Shephard e Pitt (1997), Gamerman (1998), Geweke e Tanizaki (2001) e Ravines et al. (2007). Apenas o último, denominado método *CUBS*, será descrito nesta dissertação.

A técnica do ***CUBS*** (do inglês *Conjugate Updating Backward Sampling*), pro-

posta por Ravines et al. (2007), equivale, basicamente, ao método *FFBS* (vide Seção 1.5.2) com a utilização do *Conjugate Updating* em substituição ao *Filtro de Kalman*. No nosso caso, o *Conjugate Updating* equivale, essencialmente, ao método de filtragem descrito na Seção 3.1.

O algoritmo *CUBS* propõe uma *distribuição geradora de candidatos* para Θ, numa amostragem em um movimento múltiplo de Metropolis-Hastings. Considerando a natureza sequencial dos modelos dinâmicos, a referida proposta baseia-se na decomposição da distribuição condicional completa conjunta dos estados dada por

$$p(\Theta|D_T, \psi) = p(\theta_T|D_T, \psi) \prod_{t=1}^{T-1} p(\theta_t|\theta_{t+1}, D_t, \psi) \ .$$

Em nosso caso específico, as distribuições à direita do sinal de igualdade não são conhecidas. Entretanto, seus momentos podem ser estimados usando-se, conjuntamente, o método de filtragem da Seção 3.1 e os resultados relativos às distribuições retrospectivas. Pode-se mostrar que $(\theta_t|\theta_{t+1}, D_t, \psi)$ tem primeiro e segundo momentos dados por (m_t^s, C_t^s), que são descritos por

$$\begin{aligned} m_t^s &= m_t + C_t G_{t+1}' R_{t+1}^{-1}(\theta_{t+1} - a_{t+1}) \ ; \\ C_t^s &= C_t - C_t G_{t+1}' R_{t+1}^{-1} G_{t+1} C_t \ , \end{aligned} \tag{3.14}$$

onde m_t e C_t são, respectivamente, a média e variância *online* aproximadas pela metodologia de filtragem.

A função geradora de candidatos proposta é construída simplesmente amostrando sequencialmente de distribuições normais, respeitando os momentos dados em (3.14). A grande virtude do *CUBS* está justamente em valer-se da estrutura de correlação temporal dos parâmetros do sistema na geração de candidatos.

Antes de prosseguir aos aspectos computacionais, faz-se necessário tornar claro alguns pontos técnicos da teoria, quando aplicados ao nosso modelo.

Observe que a função geradora de candidatos é construída a partir da metodologia de estimação proposta para o modelo dinâmico Dirichlet apresentado na Seção 3.1. Recorde, porém, que as suposições sobre a distribuição dos parâmetros θ_0 e w são diferentes para os modelos com ψ conhecido (Seção 3.1) e desconhecido (Seção 3.2).

Ora, não estaríamos com isso criando uma inconsistência probabilística ao modelo? Definitivamente, não.

Do ponto de vista teórico, a função geradora de candidatos é de livre escolha do analista. Isto é, o algoritmo de Metropolis-Hastings estaria válido quaisquer que fossem os momentos utilizados na distribuição geradora de candidatos normal. Do ponto de vista prático, entendemos que a diferença nas suposições dos modelos é uma questão de menor importância e, na pior das hipóteses, haveria apenas uma redução da taxa de aceitação.

Vale destacar, ainda, a impossibilidade de assumir normalidade para θ_0 e w_t no modelo estimado recursivamente via Linear Bayes. Para visualizar este ponto, suponha, por exemplo, que $F_t = \mathbf{I}$ e $G_t = \mathbf{I}$, e que os erros w_t sejam normais. Neste caso, a distribuição a posteriori $(\theta_1|D_1)$, e consequentemente o lado direito da equação de evolução $(G_2\theta_1 + w_2|D_1)$, não seriam normais. Por outro lado, ao se atribuir uma priori Logística-normal para $(\mu_2|D_1)$, teríamos, pela função de ligação, normalidade no lado esquerdo da equação de evolução $(\theta_2|D_1)$. Resumindo, assumir normalidade para θ_0 e w_t produziria uma contradição no modelo. A rigor, estaríamos atribuindo prioris diferentes para a mesma quantidade.

Voltando ao *CUBS*, a idéia do método é amostrar θ_t^* sequencialmente, para $t = T, \ldots, 1$, e aceitar o bloco Θ^* com probabilidade definida pelo método Metropolis-Hastings. O esquema de amostragem *CUBS* é sumarizado no seguinte algoritmo:

1. Calcule os momentos $m^{(i)}$ e $C^{(i)}$ de $p(\theta_t|D_t, \psi^{(i-1)})$ com o procedimento de filtragem descrito na Seção 3.1;

2. Amostre Θ^* via *Backward Sampling*:

   (a) Amostre o candidato θ_T^* da distribuição $N(m_T^{(i)}, C_T^{(i)})$

   (b) Amostre θ_t^*, $t = T-1, \ldots, 1$, de $p(\theta_t|\theta_{t+1}^*, \psi^{(i-1)})$

3. Faça $\Theta^{(i)} = \Theta^*$ com probabilidade π_t e $\Theta^{(i)} = \Theta^{(i-1)}$ com probabilidade $1 - \pi_t$, onde $\pi_t = \min\{1, A\}$ e A é a razão de aceitação do Metropolis-Hastings:

$$A = \min\left[1, \frac{\omega(\theta^*)}{\omega(\theta)}\right], \quad \omega(\theta^*) = \frac{p(\theta^*)}{q(\theta^*)},$$

onde $q(.)$ é a densidade proposta obtida pela combinação dos passos anteriores.

O passo (1) do algoritmo é aquele com maior custo computacional. Para executá-lo é preciso seguir todos os passos do processo de filtragem, em especial calcular as T integrais definidas num espaço de dimensão $k-1$, envolvidos na aproximação dos momentos a posteriori de μ_t. Como mencionado anteriormente, o tempo computacional cresce exponencialmente em função de k. Dessa forma, repetir as etapas descritas na Seção 3.1 milhares de vezes, uma para cada amostra gerada, seria impraticável em problemas com mais de duas categorias possíveis para cada tempo da série observada.

Felizmente, o método de amostragem Metropolis-Hastings continua válido, independente da execução repetida do passo (1). Esse resultado nos permite atualizar a função geradora de candidatos eventualmente, digamos a cada 100 iterações, por exemplo. Temos assim um grande ganho em eficiência computacional. Todavia, é preciso estar ciente que há, em contrapartida, certa perda de "qualidade"da função geradora de candidatos. Os momentos filtrados vão ficando desatualizados, no sentido de que foram construídos com base no conjunto ψ relativo a iterações "antigas".

Um caso particular dessa estratégia seria executar o passo (1) uma única vez. Suponha primeiro que conjunto $\psi^{(0)}$ seja igual ao conjunto real desconhecido ψ. Ou seja, os valores iniciais foram especificados perfeitamente. A função geradora de candidatos seria a melhor possível e, portanto, poderia ser mantida ao longo de todo o processo MCMC. Por outro lado, na situação em que os valores iniciais $\psi^{(0)}$ forem mal especificados, a cadeia evoluirá, enquanto os momentos m_t e C_t, usados para gerar os candidatos, se manterão inalterados e inadequados. Dessa forma, a função $q(\cdot)$ se mostrará proibitivamente ineficiente. Na prática, o que se deseja é atualizar mais frequentemente a função geradora de candidatos (executar o passo (1)) enquanto a convergência não foi atingida.

O *CUBS* amostra, em um único movimento, centenas de parâmetros (dimensão de Θ). Talvez, por isso, as taxas de aceitação nem sempre são razoáveis. Não é incomum observar-se taxas próximas ou inferiores a 1%. Em uma tentativa de aumentar as taxas de aceitação, implementamos em nossa rotina computacional a amostragem baseada numa *partição* de Θ. Suponha que o vetor Θ seja dividido em m partes. Uma para cada 12 tempos, por exemplo. Então, usando rigorosamente o mesmo

princípio do *CUBS*, é possível amostrar cada uma dessas partes sequencialmente, aceitando ou rejeitando-as em função da respectiva razão de Metropolis-Hastings. Probabilisticamente, se estaria amostrando da distribuição condicional completa de cada uma das partes de Θ. Ainda que a técnica da partição de Θ funcione bem em vários contextos, não observamos ganhos significativos no caso do modelo dinâmico Dirichlet.

O método *CUBS*, embora muito interessante, não é o único disponível para amostrar os vetores do sistema. Imaginamos que outras propostas podem trazer ganhos de eficiência em alguns cenários.

O primeiro passo do *CUBS*, calcular os momentos filtrados m_t e C_t, pode ser executado usando a função *mddFilter* apresentado na Subseção 3.1.1. Para gerar uma amostra candidata Θ^* e aceitá-la com probabilidade dada pela razão Metropolis-Hastings, criamos a função *R mddThetaSample*, cujos *inputs* são os seguintes objetos:

(1) `filtro`, retornado pela função *mddFilter*. É nesse objeto que a função busca as quantidades envolvidas no processo de filtragem, incluindo m_t e C_t;

(2) `y`, matriz dos dados e dos valores $Y_{\mathcal{A}^C}^{(i)}$;

(3) `mod`, especificação do modelo;

(4) `Theta`, matriz representado $\Theta^{(i-1)}$. Cada linha contém o respectivo valor de $\theta_t^{(i-1)}$. Observe que esse objeto será usado no cálculo da probabilidade de aceitação da amostra candidata.

(5) `theta0`, o valor corrente do vetor θ_0;

(6) `phi`, o valor atual do parâmetro de precisão ϕ;

A função *mddThetaSample*, cujo código não é tão simples, retorna (em forma de matriz) a amostra $\Theta^{(i)}$. Como está implementada, é possível utilizar a função para fazer a amostragem em partes, conforme a estratégia de partição de Θ citada anteriormente. Por não termos observado benefícios reais no modelo dinâmico Dirichlet, não abordaremos a questão de como usar a função nesse contexto.

3.2.1.3 Amostrando os estados do sistema - Caso Estático

O caso estático do modelo dinâmico Dirichlet é especial e requer um tratamento particular. Nele, as matrizes W_t são nulas e os vetores θ_t, para $t = 0, \ldots, T$, se relacionam deterministicamente. A fim de evitar confusão com o caso dinâmico, denotaremos o vetor de parâmetros do sistema, de dimensão p, por θ. Dessa maneira, tem-se que $\theta_t = G_t\theta_{t-1} = G_tG_{t-1}\ldots G_1\theta$, para $t = 0, \ldots, T$, e $\theta_0 = \theta$. Ou seja, para qualquer t, θ_t é determinado pelo vetor θ. Assim, ao invés de $p * (T + 1)$ parâmetros livres relativos ao sistema, no caso estático há apenas p a estimar. Normalmente tem-se, no caso estático, $G_t = \mathbf{I}$, para $t = 1, \ldots, T$.

A distribuição condicional completa de θ é dada por

$$p(\theta|\ldots) \propto \prod_{t=1}^{T} p(y_t|\theta_t, \phi)p(\theta). \tag{3.15}$$

Utilizamoscomo priori para Θ a distribuição Gaussiana. Dessa forma, (3.15) não tem forma fechada, sendo o produto da distribuição a priori normal de θ e das distribuições Dirichlet das observações.

Sugerimos novamente o uso do amostrador de Metropolis-Hastings. Aqui, no entanto, substituímos o *CUBS* por uma função geradora de candidatos Normal, centrada no valor corrente de θ, escrita como $q(\theta) \sim N(\theta^{(i-1)}, \Sigma)$. Essa escolha torna a cadeia um passeio aleatório gaussiano, e pela simetria da função $q(\theta^{(i-1)}, \theta^*) = q(\theta^*, \theta^{(i-1)})$, a probabilidade de aceitação da amostra gerada só depende das verossimilhanças, mais precisamente: $\pi(\theta^{(i-1)}, \theta^*) = \min\{1, p(\theta^*)/p(\theta^{(i-1)})\}$.

A boa escolha da matriz Σ é crucial para a eficiência do algoritmo e será discutida nas Subseções 4.1.2 e 4.2.2. Há uma forma muito interessante de especificá-la por meio da abordagem de estimação *online* e do uso de fatores de desconto. Neste caso, só é preciso fazer a filtragem uma única vez e as taxas de aceitação são calibráveis, preferencialmente assumindo valores entre 50 e 25%, a depender da dimensão p (Roberts et al., 1994).

Para amostrar o vetor θ seguindo a estratégia proposta, criamos a função *mddStaticThetaSample*. As entradas da função são as seguintes:

(1) `y`, matriz das observações;

(2) `mod`, especificação do modelo. Note que os parâmetros da priori $p(\theta)$ são m_0 e C_0, e devem ser armazenados nesse objeto;

(3) `phi`, valor corrente do parâmetro de precisão das observações ϕ;

(4) `theta`, valor corrente do vetor θ, ou seja, $\theta^{(i-1)}$;

(5) Σ, matriz de covariância da função geradora de candidatos.

A função *mddStaticThetaSample* devolve a amostra $\theta^{(i)}$ e a matriz Θ associada, contendo, nas respectivas linhas, os vetores $\theta_t^{(i)}$. O objetivo é facilitar o chamado das demais funções, calculadas a partir desta matriz.

3.2.1.4 Amostrando a matriz de covariância dos erros

A matriz de covariância, W, dos erros de evolução do sistema desempenha um papel crucial nos modelos dinâmicos. Considerando os princípios de sobreposição de MLDs discutidos na Seção 1.4.4, é benéfico escrever W como uma matriz bloco-diagonal com elementos $(W_1, \ldots, W_h)$, com W_i de dimensão $p_i \times p_i$. Note que não há qualquer perda de generalidade, visto que é possível, simplesmente, fazer $h = 1$. Normalmente, esta partição de W decorre da composição do vetor do sistema e da matriz de evolução, expressos, respectivamente, por

$$\theta_t = \begin{bmatrix} \theta_{1,t} \\ \vdots \\ \theta_{h,t} \end{bmatrix} \quad \text{e} \quad G_t = \begin{bmatrix} G_{1,t} & \ldots & 0 \\ 0 & \ddots & 0 \\ 0 & \ldots & G_{h,t} \end{bmatrix}.$$

Como de costume na abordagem Bayesiana, denotamos por $\Phi = W^{-1}$ a matriz de precisão bloco-diagonal com elementos $\Phi_i = W_i^{-1}$, $i = 1, \ldots, h$. Antes de prosseguirmos para o processo de amostragem propriamente dito, apresentamos brevemente a distribuição Wishart.

Se W é uma matriz simétrica positiva-definida, a distribuição de W é a distribuição conjunta de suas entradas $(w_{i,j})$, $i, j = 1, \ldots, k$. Dizemos, então, que W tem

distribuição *Wishart* com parâmetros α e B se sua densidade é da forma

$$\mathcal{W}_k(W;\alpha,B) = c|W|^{\alpha-(k+1)/2}\exp(-\mathrm{tr}(BW))\,,$$

onde $\alpha \geq (k-1)/2$, B é uma matriz simétrica positiva-definida, $c = |B|^{\alpha}/\Gamma_k(\alpha)$, $\Gamma_k(\alpha) = \pi^{k(k-1)/4}\prod_{i=1}^{k}\Gamma((2\alpha+1-i)/2)$ é a função gama generalizada e $\mathrm{tr}(\cdot)$ denota o traço da matriz. Com esta parametrização, $E(W) = \alpha B^{-1}$.

Há uma forma alternativa de parametrização muito adotada na literatura. Nela, a representação $W \sim \mathcal{W}_k(W;\alpha,B)$ é escrita como $W \sim \mathcal{W}_k(W;n/2,\Sigma^{-1}/2)$. Neste caso, n é chamado de *graus de liberdade*.

A distribuição Wishart, que pode ser entendida como uma generalização multivariada das distribuições Gama e Chi-quadrado, surge como a distribuição da matriz de covariância amostral de observações retiradas de uma Normal multivariada. No contexto Bayesiano, a relevância da Wishart é consequência desta ser a distribuição conjugada da matriz de precisão de uma distribuição normal multivariada. Vale destacar que se $W \sim \mathcal{W}_k(W;\alpha = n/2, B = \Sigma^{-1}/2)$, então $\Phi = W^{-1}$ tem distribuição Wishart-Invertida, e para $n \geq k+2$, $E(\Phi) = \Sigma^{-1}/(n-k-1)$.

Ao adotarmos prioris Wishart(ν_i, S_i) independentes para cada uma das submatrizes de precisão $\Phi_i = W_i^{-1}$, para $i = 1,\ldots,h$, obtém-se, pela Equação (3.11), a seguinte distribuição condicional completa de Φ_i:

$$\begin{aligned}
p(\Phi_i|\ldots) &\propto \prod_{t=1}^{T} p(\theta_t|\theta_{t-1},\Phi)p(\Phi) \\
&\propto \prod_{t=1}^{T} N(\theta_t; G_t\theta_{t-1}, \Phi^{-1})\mathcal{W}(\Phi_i;\nu_i,S_i) \\
&\propto \prod_{t=1}^{T}\prod_{j=1}^{h} |\Phi_j|^{1/2}\exp\{-\frac{1}{2}(\theta_t - G_t\theta_{t-1})'\Phi(\theta_t - G_t\theta_{t-1})\} \\
&\quad \times |\Phi_i|^{\nu_i-(p_i+1)/2}\exp\{-\mathrm{tr}(S_i\Phi_i)\} \\
&\propto |\Phi_i|^{T/2+\nu_i-(p_i+1)/2} \\
&\quad \times \exp\{-\mathrm{tr}(\frac{1}{2}\sum_{t=1}^{T}(\theta_t - G_t\theta_{t-1})(\theta_t - G_t\theta_{t-1})'\Phi) - \mathrm{tr}(S_i\Phi_i)\}. \qquad (3.16)
\end{aligned}$$

O último passo da Equação (3.16) justifica-se pelas seguintes propriedades algébricas:

Se A e B são matrizes tais que AB é um escalar, então $AB = \text{tr}(AB)$ e $\text{tr}(AB) = \text{tr}(BA)$.

Para simplificar ainda mais a Equação (3.16), faça

$$SS_t = (\theta_t - G_t\theta_{t-1})(\theta_t - G_t\theta_{t-1})' \quad \text{e}$$
$$SS_{ii,t} = (\theta_{i,t} - G_{i,t}\theta_{i,t-1})(\theta_{i,t} - G_{i,t}\theta_{i,t-1})'.$$

É possível mostrar que $\text{tr}(SS_t\Phi) = \sum_{j=1}^{h} \text{tr}(SS_{jj,t}\Phi_j)$, e consequentemente, a distribuição condicional completa de Φ_i pode ser escrita como

$$p(\Phi_i|\ldots) \propto |\Phi_i|^{T/2+\nu_i-(p_i+1)/2} \exp\left\{-\text{tr}\left(\left(\frac{1}{2}SS_{i\cdot} + S_i\right)\Phi_i\right)\right\}, \tag{3.17}$$

onde $SS_{i\cdot} = \sum_{t=1}^{T} SS_{ii,t}$. Essa densidade é justamente o núcleo de uma distribuição Wishart$(T/2 + \nu_i,\ S_i + SS_{i\cdot}/2)$, o que garante ser esta a distribuição $p(\Phi_i|\ldots)$.

A grande vantagem da adoção da priori Wishart é permitir que se amostre de (3.17) de maneira simples e direta. Por outro lado, não é sempre fácil definir uma distribuição a priori suficientemente vaga. Neste caso, por se inserir no modelo mais informação do que de fato se tem previamente, é possível que o efeito indevido da priori seja demasiado, contaminando, portanto, o ajuste.

Os algoritmos de geração de amostras da distribuição Wishart, tanto do pacote *dlm* quando do pacote *MCMCpack*, são baseados na decomposição de *Bartlett* e só valem quando o grau de liberdade n é um número inteiro tal que $n \geq k$. Assim, para representar uma priori vaga se utiliza o menor valor de n possível, qual seja, k. Essa estratégia costuma funcionar bem em inúmeras situações, em especial quando T é grande e/ou existe algum conhecimento prévio acerca da matriz W. No modelo dinâmico Dirichlet, essas matrizes W em geral são compostas por entradas próximas a zero. Suponha, por exemplo, que W real seja uma matriz diagonal com elementos iguais a 10^{-5}. É fácil perceber que uma priori baseada em uma estimativa pontual $\hat{W}$ proporcionalmente longe do valor real de W pode deslocar a posteriori, ainda que os valores absolutos de $\hat{W}$ e W sejam parecidos, em certo sentido.

Pela dificuldade na definição da priori, as taxas de aceitação do CUBS, por vezes, se mostraram instáveis e bastante dependentes dos parâmetros ν_i e S_i adotados. Além

disso, alguns parâmetros da priori provocaram complicações numéricas do código desenvolvido. Observe que no processo de estimação é preciso lidar com milhares de cálculos de integrais delicadas e inversões de matrizes com determinante quase nulo. A estabilidade numérica e eficiência das rotinas computacionais envolvidas requerem um projeto a parte. Assumidamente, não temos uma avaliação conclusiva sobre como especificar as priori Wishart. Sugerimos, portanto, cautela ao lidar com esses elementos e, se possível, fazer uma análise de sensibilidade para diagnosticar o impacto da priori no ajuste do modelo. Em todas as aplicações do Capítulo 4 adotamos o par $(\nu_i, S_i) = (k/2, 10^{-3}\mathbf{I})$, que mesmo sujeito aos comentários tecidos acima, produziram bons ajustes. Note, por fim, que não há nenhum ponto teórico que impeça a utilização de outras prioris, que não a Wishart. A busca por novas possibilidades constitui um tópico para trabalhos futuros.

Com o intuito de simplificar a execução da amostragem das matrizes Φ_i, desenvolvemos a função *mddWSample*, que retorna a matriz W_i simulada. Internamente usamos a função *rWishart* do pacote *dlm* para amostrar da distribuição condicional completa. Suas entradas são as seguintes:

(1) `mod`, especificação do modelo;

(2) `Theta`, matriz representado $\Theta^{(i-1)}$;

(3) `theta0`, o valor corrente do vetor θ_0;

(4) `nu.i`, primeiro parâmetro da priori de Φ_i. Um valor inteiro, correspondente a duas vezes o número de graus de liberdade, maior ou igual a $k/2$;

(5) `Si`, segundo parâmetro da priori. Uma matriz positiva-definida tal que $E(\Phi_i) = \nu S_i$;

(6) `pos.inicial`, valor inteiro correspondente à posição do primeiro elemento de $SS_{ii,t}$ em SS_t. Como exemplo, em um modelo formado por dois blocos, um polinomial e um sazonal, se a matriz amostrada corresponde ao segundo, então esse argumento deve ser igual à dimensão do primeiro mais um;

(7) `pos.final`, valor inteiro correspondente à posição do último elemento de $SS_{ii,t}$ em SS_t. No exemplo anterior, esse argumento seria igual à dimensão do sistema.

3.2.1.5 Amostrando o estado inicial do sistema

Assumindo-se normalidade para a distribuição a priori do estado inicial do sistema, a Equação (3.11) garante que a distribuição condicional completa de θ_0 é tal que

$$\begin{aligned} p(\theta_0|\ldots) &\propto p(\theta_1|\theta_0, W)p(\theta_0) \\ &\propto \exp\{-(\theta_1 - G_1\theta_0)'W^{-1}(\theta_1 - G_1\theta_0)/2\} \\ &\quad \times \exp\{-(\theta_0 - m_0)'C_0^{-1}(\theta_0 - m_0)/2\}. \end{aligned}$$

A expansão e rearranjo da função quadrática em θ_0, versão multivariada da técnica de "completar quadrados", nos permite mostrar que a distribuição condicional completa de θ_0 também é Normal, com momentos conhecidos, isto é:

$$\begin{gathered} (\theta_0|\ldots) \sim N(\mu, \Sigma), \quad \text{onde} \\ \mu = \Sigma(G_1'W^{-1}\theta_1 + C_0^{-1}m_0) \quad \text{e} \quad \Sigma^{-1} = G_1'W^{-1}G_1 + C_0^{-1}. \end{gathered}$$

Para executar a amostragem de θ_0, construímos a função *R* *mddTheta0Sample*. As entradas da função são os usuais objetos `mod`, onde a rotina busca pelas matrizes G_1, m_0, C_0 e W, e `Theta`, a matriz Θ, cuja primeira linha corresponde exatamente ao vetor θ_1 atual. A rotina *mddTheta0Sample*, que internamente faz uso da função *rmvnorm*, do pacote *mvtnorm* (Genz et al., 2011), retorna o valor gerado de θ_0.

3.2.1.6 Amostrando o parâmetro de precisão

O último parâmetro a ser amostrado no algoritmo de Gibbs é exatamente ϕ. Sua distribuição condicional completa é dada por

$$p(\phi|\ldots) \propto \prod_{t=1}^{T} p(y_t|\theta_t, \phi)p(\phi). \tag{3.18}$$

Novamente, as distribuições que envolvem os dados y_t são Dirichlet. A única restrição quanto à distribuição a priori de ϕ é que tenha suporte na reta real positiva. Recomendamos a adoção de uma priori não conjugada Gama(α, β). A equação (3.18) evidencia o fato da distribuição alvo não ter forma fechada, indicando mais uma

vez o uso do método de Metropolis-Hastings. Como função geradora de candidatos, sugerimos a função Log-normal com média igual ao valor atual de ϕ. Isto é, um passeio aleatório Log-normal.

Podemos dizer que esse esquema de amostragem de ϕ se mostrou muito robusto. As taxas de aceitação são sempre boas, a especificação do valor inicial não é uma questão crucial e a convergência da cadeia é normalmente rápida. Em resumo, esse certamente não é um parâmetro que nos traga preocupação.

Computacionalmente, desenvolvemos a rotina *mddPhiSample* capaz de amostrar o parâmetro ϕ nas referidas condições. Para se utiliza-la, é preciso especificar os seguintes objetos

(1) `y`, matriz das observações;

(2) mod, especificação do modelo;

(3) `Theta`, a representação matricial de Θ;

(4) `phi`, valor corrente de ϕ;

(5) `alpha`, parâmetro α da priori. O valor padrão é 0.001;

(6) `beta`, parâmetro β da priori. O valor padrão é 0.001;

(7) `sdlog`, desvio padrão da função geradora de candidatos Log-normal na escala do logaritmo. O valor padrão adotado foi 0.5. Em todas as aplicações do Capítulo 4 foi utilizado este desvio padrão na geração de candidatos.

Os valores da priori gama correspondem à situação na qual pouca informação está previamente disponível. Não surpreendentemente, este é justamente o caso de todas as aplicações consideradas nesta pesquisa. A função *mddPhiSample* utiliza as funções *rlnorm* e *dlnorm* para, respectivamente, gerar uma amostra ϕ^* e avaliar a densidade de $q(\cdot)$ nos pontos $\phi^{(i-1)}$ e ϕ^*. O objeto retornado por nossa função é o escalar $\phi^{(i)}$.

Capítulo 4

Casos especiais e aplicações

Neste capítulo serão ajustados alguns modelos particulares, da classe dos modelos dinâmicos Dirichlet, de grande relevância. Trata-se de exemplos simples que visam, essencialmente, ilustrar o funcionamento das técnicas descritas nos capítulos anteriores. Destacamos que os casos relativos a dados reais são meramente motivacionais, uma vez que a elaboração de um modelo de interesse prático requer uma profundidade de discussão que extrapola o escopo dessa dissertação.

Ressaltamos que se optou por uma abordagem superficial no que se refere à análise de diagnóstico, em especial acerca das cadeias de Markov. Buscamos, dessa maneira, evitar possíveis desvios quanto ao foco central desta pesquisa, qual seja propor o Modelo Dinâmico Dirichlet (MDD) e discutir suas principais características. De qualquer maneira, os principais métodos propostos na literatura para identificar a convergência das cadeias de Markov foram usados e, no limite de nosso conhecimento, todos os exemplos respeitaram os requisitos básicos que viabilizam a utilização do método MCMC nas estimativas dos parâmetros. No *R*, o pacote *coda* (Plummer et al., 2010) permite fácil aplicação dos testes mais populares. O pacote *MCMCpack* também é útil no tratamento de amostras MCMC. Aos interessados no aprofundamento teórico do tema, sugerimos a leitura de Gamerman e Lopes (2006).

Incluímos os códigos utilizados nos ajustes, por entender que o aspecto computacional desempenha um papel fundamental no processo de análise estatística. Acreditamos ser esta a melhor forma de permitir ao leitor a visualização plena da aplicação do modelo. Um benefício adicional é mostrar como todas as diferentes situações abor-

dadas podem ser modeladas simplesmente redefinindo-se os elementos que definem o MDD.

A Seção 4.1 restringe-se ao estudo de modelos para dados univariados, a importante sub-classe do MDD na qual as observações têm distribuição Beta. Na Seção 4.2, discutem-se situações relativas a cenários em que, a cada instante de tempo, as observações são multivariadas.

4.1 Modelos univariados - Caso Beta

Serão apresentados a seguir 3 ajustes do MDD univariado. Na Subseção 4.1.1 tratamos um conjunto de dados simulados. O conhecimento dos parâmetros que geraram a série nos permite avaliar a qualidade das estimativas.

A Subseção 4.1.2 compara o ajuste dos modelos dinâmicos Dirichlet das Seções 3.1 e 3.2 com o modelo clássico de regressão Beta (Ferrari e Cribari-Neto, 2004). O conjunto de dados está entre aqueles usados nos exemplos contidos na *ajuda* da função *R betareg* (Cribari-Neto e Zeileis, 2010).

Encerramos essa seção ajustando um conjunto de dados reais referentes a internações decorrentes de doenças do aparelho respiratório no sistema de saúde do Distrito Federal. O objetivo é descrever, superficialmente, quais são os benefícios potenciais da utilização dos MDD univariados em situações práticas.

4.1.1 Ajuste de dados simulados

No intuito de ilustrar o funcionamento do modelo dinâmico Dirichlet univariado, construímos um cenário em que as observações exibem uma tendência localmente linear e um comportamento cíclico com 4 períodos, além de depender de uma variável regressora conhecida X_t. Com base nos procedimentos de especificação de modelos apresentados na Seção 1.4, adotamos o MDD, apresentado na Seção 3.2, definido pelas seguintes quantidades:

$$\left\{ F_t = (E_2, E_2, 1, x_t)',\ G = \begin{pmatrix} J_2(1) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{G}_{par} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix},\ W = \begin{pmatrix} 10^{-2} & \mathbf{0} \\ \mathbf{0} & 10^{-4}\mathbf{I}_{5\times 5} \end{pmatrix},\ \phi = 300 \right\},$$

onde

$$E_2 = (1, 0),\ J_2(1) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{e} \quad \mathcal{G}_{par} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Para gerar dados desse modelo é preciso especificar os momentos à priori m_0 e C_0 do estado inicial θ_0. Enquanto o primeiro, média m_0, foi definido aleatoriamente, o segundo, matriz de covariância C_0, foi construído de maneira a forçar a relação $\theta_0 \approx m_0$ e evitar preditores lineares demasiadamente distantes de 0. Observe, por exemplo, que pela função logito, valores de λ_t iguais a -7 e 7 estão associados, respectivamente, a médias μ_t iguais a 0.001 e 0.999. Ou seja, valores muito altos (ou muito baixos) de λ_t produzem séries extremas de pouco interesse prático, com $\mu_t \approx 1$ (ou $\mu_t \approx 0$). É importante destacar que não estamos adotando uma priori precisa no processo de estimação, mas sim na geração dos dados. A definição da priori de θ_0 usada na etapa inferencial será feita adiante.

Computacionalmente, para simular uma observação desse modelo, utilizamos a função *rMdd*, desenvolvida nessa pesquisa. Ela retorna ao usuário não somente a série simulada y_t mas também as médias μ_t e os estados θ_t, ambas na forma matricial. É preciso explicitar, no chamado da função, o modelo por meio dos objetos `mod` e `phi`, e o tamanho `N` desejado para a série simulada.

Observe que removemos, aleatoriamente, 20% das 50 observações geradas a fim de também descrever a amostragem das observações faltantes. O primeiro bloco da programação está destinado à especificação e geração de uma observação do modelo proposto.

Agora, já tendo os dados *reais* que nos servirão de gabarito, seguimos à fase de estimação. Para tanto, nos limitamos a nos valer apenas dos dados y_t e do *design* do modelo, representado pelas matrizes F_t e G_t, desconsiderando, portanto, os valores conhecidos de ϕ, W e θ_t no processo de estimação.

Estimação via MCMC

Embora estejamos adotando a abordagem via MCMC, a estimação recursiva, em conjunto com a técnica dos fatores de desconto, é muito útil na especificação dos valores iniciais $\Theta^{(0)}$ e $W^{(0)}$ do algoritmo de Gibbs. Mais precisamente, executamos o procedimento de filtragem com fatores de desconto e utilizamos a matriz W_N retornada para fixar o valor inicial $W^{(0)}$ do algoritmo de Gibbs. Em seguida, repetimos a filtragem considerando $W = W^{(0)}$ (agora essa matriz é tida como conhecida) e adotamos as médias suavizadas de θ_t para definir $\Theta^{(0)}$. Podemos afirmar que essa estratégia culmina em bons valores iniciais de Θ. Novamente, não há qualquer contradição quanto ao uso simbiótico dos modelos dados nas Seções 3.1 e 3.2. A justificativa sendo a liberdade do analista em determinar valores iniciais no método de Gibbs.

De acordo com o bloco 2 do Código 4.1, adotamos uma priori vaga $\theta_0 \sim N(\mathbf{0}, 10\ \mathbf{I_{6\times 6}})$, um vetor de fatores de desconto $\delta = (0.8, 0.95, 0.95)$ e um valor inicial do parâmetro de precisão $\phi^{(0)} = 100$ (recorde que o valor real é 300). Além disso, criamos nesse bloco os objetos que receberão as amostras geradas no algoritmo MCMC e definimos os valores iniciais para W e Θ usando a função *mddFilter* e o procedimento descrito no parágrafo anterior.

Código 4.1: Ajuste de dados simulados univariados

```
> # Bloco 1: Geração dos dados e definição do modelo
> set.seed(111)
> N <- 50
> p <- 4
> X <- rnorm(N)
> mod <- dlmModPoly(2) + dlmModTrig(p) +
+   dlmModReg(X, addInt = F)
> mod$GG[3, 3] <- mod$GG[4, 4] <- 0
> W.real <- diag(mod$W) <- c(1e-2, rep(1e-4, 5))
> mod$m0 <- rnorm(6, sd = c(0.2, 0.01, rep(0.2, 4)))
> mod$C0 <- diag(1e-5, 6)
> phi.real <- 300
> modObs <- rMdd(mod, phi.real, N)
> y <- modObs$y
> mu.real <- modObs$mu
> theta.real <- modObs$theta
```

```
> y[sample(2:(N - 1), N / 10), ] <- NA
>
> # Bloco 2: Especificação dos valores iniciais do Gibbs
> disc <- c(0.8, 0.95, 0.95)
> dim.disc <- c(2, p - 1, 1)
> dim.sys <- length(mod$m0)
> mod$m0 <- rep(0, dim.sys)
> mod$C0 <- diag(10, dim.sys)
> phi0 <- 100
> mcmc <- 300000
> Theta <- array(dim = c(N, dim.sys, mcmc + 1))
> W <- array(0, dim = c(dim.sys, dim.sys, mcmc + 1))
> theta0 <- rbind(mod$m0, matrix(nr = mcmc, nc = dim.sys))
> phi <- c(phi0, rep(0, mcmc))
> aux <- mddFilter(y, mod, phi0 , disc, dim.disc)
> W[, , 1] <- mod$W <- dlmSvd2var(aux$U.W, aux$D.W)[[N]]
> filtro <- mddFilter(y, mod, phi0)
> Theta[, , 1] <- dropFirst(mddSmooth(filtro)$s)
> linhasNA <- order(y[, 1], na.last = F)[1:sum(is.na(y[, 1]))]
> Si <- diag(1e-3, dim.sys)
>
> # Bloco 3: Geração, via Gibbs, de amostras da posteriori
> for (i in 2:(mcmc + 1)) {
+   filtro <- mddFilter(y, mod, phi[i - 1])
+   y[linhasNA, ] <- mddYSample(mod,
+     Theta[, , i - 1], phi[i - 1], linhasNA)
+   Theta[, , i] <- mddThetaSample(filtro, y, mod,
+     Theta[, , i - 1], theta0[i - 1, ], phi[i - 1])
+   W[1:2, 1:2, i] <- mddWSample(mod, Theta[, , i],
+     theta0[i - 1, ], 2 / 2, Si[1:2, 1:2], 1, 2)
+   W[3:5, 3:5, i] <- mddWSample(mod, Theta[, , i],
+     theta0[i - 1, ], 3 / 2, Si[3:5, 3:5], 3, 5)
+   W[6, 6, i] <- mddWSample(mod, Theta[, , i],
+     theta0[i - 1, ], 1 / 2, Si[6, 6, drop = F], 6, 6)
+   mod$W <- W[, , i]
+   theta0[i, ] <- mddTheta0Sample(mod, Theta[, , i])
+   phi[i] <- mddPhiSample(y, mod, Theta[, , i], phi[i - 1])
+ }
```

No terceiro bloco da programação, executamos o método de estimação proposto na Seção 3.2. Isto é, geramos 300000 amostras das quantidades desconhecidas. Note que optamos por não armazenar, mas sim reciclar, as amostras das observações faltantes $Y_{\mathcal{A}^c}$.

A matriz W é amostrada de forma particionada. Ou seja, amostramos cada uma das sub-matrizes W_i separadamente. Dessa forma, garantimos que os componentes do modelo (polinomial, sazonal e regressivo) sejam tratados como independentes uns dos outros. Em outras palavras, a evolução do comportamento sazonal não interfere na dinâmica da relação entre X_t e Y_t, por exemplo. Como mencionado na Subseção 3.2.1.4, adotamos como priori para Φ_i os parâmetros $(\nu_i, S_i) = (k/2, 10^{-3}\ \mathbf{I})$, onde k é o número de linhas (e colunas) de Φ_i. Vale dizer que, a rigor, não há qualquer impedimento em se fazer de W uma matriz genérica, sem qualquer partição.

Como se sabe, o algoritmo MCMC produz, após o alcance da convergência, amostras correlacionadas. O exame descritivo, tanto dos traços da cadeia quanto das médias ergódicas, indicou o início do estado de convergência por volta da iteração 150000. Dessa forma, descartamos a primeira metade da cadeia como período de *burn-in*. Outra medida útil foi manter apenas 1 a cada 100 amostras geradas. O objetivo sendo a redução da correlação entre as amostras.

Análise dos dados simulados

A Figura 4.1 apresenta, em gráficos distintos, o ajuste de cada um dos elementos envolvidos na análise. O primeiro gráfico mostra a série Y_t (linha cinza), enquanto o segundo sobrepõe as médias reais μ_t (linha vermelha) e suas estimativas (linha preta), com os respectivos intervalos de credibilidade de 95% (linhas pretas pontilhadas), baseados nos quantis empíricos da distribuição a posteriori. Os 3 últimos gráficos exibem os valores reais (linhas vermelhas) e estimados (linhas pretas) para os efeitos de nível, de sazonalidade e da regressão. A leitura global da Figura 4.1 mostra que o procedimento inferencial foi bem sucedido em decompor a informação de y_t em termos dos componentes que o definem.

Embora somente tenhamos tratado de uma única série temporal y_t, foi possível aferir, com certa precisão, a contribuição de cada componente latente no resultado observado. Note, por exemplo, que ao mesmo tempo em que estimamos a relação

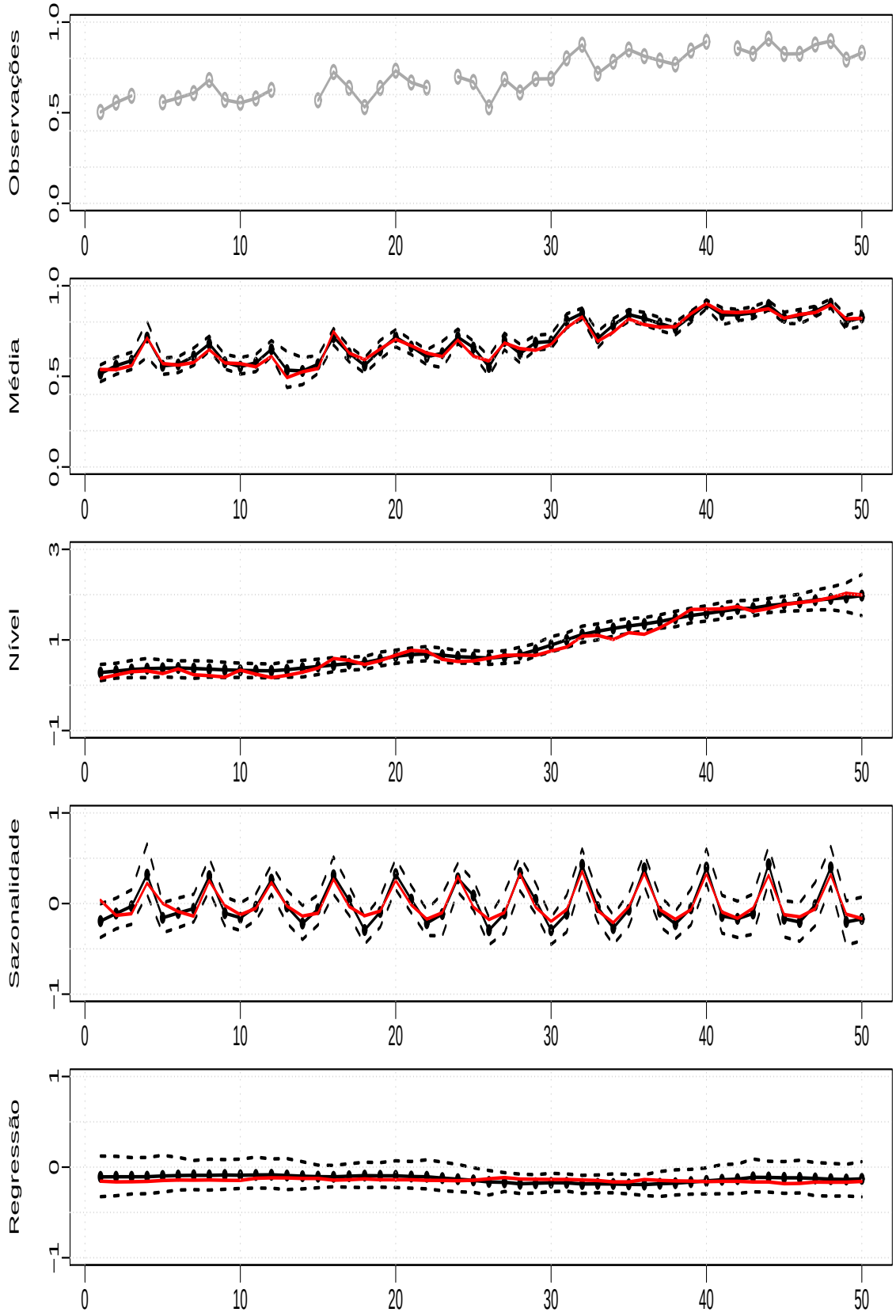

Figura 4.1: (a) Serie histórica y_t (linhas cinzas); (b) Médias μ_t; (c) Níveis θ_{1t}; (d) Efeitos de sazonalidade $\theta_{3t} + \theta_{5t}$; (e) Efeito da regressão θ_{6t}. Valores reais (linhas vermelhas), valores estimados (linhas pretas) e intervalos de credibilidade de 95% (linhas pretas tracejadas).

entre Y_t e X_t, atividade comum em modelos de regressão, descrevemos a dinâmica do nível da série, tarefa normalmente desempenhada por modelos de séries temporais. O fato de haver observações faltantes não causou degeneração nos valores estimados dos parâmetros.

Quanto ao componente regressivo, pode-se dizer, pela Figura 4.1, que $\theta_{6t} < 0$, para $i = 1, \ldots, N$. Ou seja, para um dado instante t, um aumento de m unidades no valor de X_t culmina numa redução, da ordem de $m\,\theta_{6t}$, no valor de λ_t. Dessa forma, pela função de ligação, tal variação em λ_t resulta, em última instância, numa redução da média μ_t. Em resumo, o parâmetro θ_{6t} determina em que medida variações do valor da variável regressora X_t impactam na média μ_t. Recorde que a relação entre essas quantidades não é linear e esta definida pela função de ligação logito aditiva.

Cabe destacar que a *moda* e *mediana* da distribuição a posteriori de ϕ foram, respectivamente, 260.99 e 339.30 (recorde que o valor real de ϕ na geração dos dados é 300). Então, podemos dizer que, ao menos neste caso, o método inferencial proposto conseguiu recuperar os parâmetros do modelo.

4.1.2 Regressão Beta - comparação com a abordagem clássica

O modelo dinâmico Dirichlet para dados univariados de proporções pode ser usado em um contexto estático, funcionando como um modelo de regressão. Nesse sentido, a regressão Beta proposta por Ferrari e Cribari-Neto (2004) pode ser entendida, quando ϕ é constante e a função de ligação logito é empregada, como um modelo clássico alternativo ao MDD no caso univariado.

Nesta subseção aplicaremos um MDD aos dados do segundo exemplo apresentado na ajuda da função *R betareg*, do pacote *betareg* (Cribari-Neto e Zeileis, 2010). Os autores usam um conjunto de dados retirados de Griffiths et al. (1993) para ajustar um modelo de regressão Beta em que a variável resposta é o percentual da renda domiciliar gasto com alimentação, enquanto as variáveis regressoras são a renda e o número de moradores do domicílio. A comparação dos resultados nos permitirá visualizar como os modelos, caracterizados por métodos de estimação absolutamente distintos, conduzem, sob certas condições, a estimativas virtualmente idênticas. Antes de prosseguir à discussão do Código 4.2, sugerimos a releitura da Subseção 3.2.1.3.

No primeiro bloco do Código 4.2, definimos o modelo MDD e formatamos as proporções y_t. Observe que adotamos $W = \mathbf{0}$, $F_t = (1, x_{1t}, x_{2t})'$, $G_t = \mathbf{I}$ e prioris vagas para o sistema $\theta \sim N(m_0, C_0)$ e para o parâmetro de precisão $\phi \sim \text{Gama}(0.001, 0.001)$, com $m_0 = (0, 0, 0)'$ e $C_0 = 10\,\mathbf{I}$.

Código 4.2: MDD estático e modelo clássico de regressão Beta

```
> # Bloco 1: Geração dos dados e definição do modelo
> data("FoodExpenditure", package = "betareg")
> X <- cbind(FoodExpenditure$income, FoodExpenditure$persons)
> mod <- dlmModReg(X)
> aux <- as.matrix(FoodExpenditure$food / FoodExpenditure$income)
> y <- cbind(aux, 1 - aux)
> N <- nrow(y)
> dim.sys <- length(mod$m0)
> mod$C0 <- diag(10, dim.sys)
> phi0 <- 100
>
> # Bloco 2: Estimação do modelo
> mcmc <- 20000
> filtro <- mddFilter(y, mod, phi0)
> theta <- rbind(filtro$m[N + 1, ], matrix(nr = mcmc, nc = dim.sys))
> phi <- c(phi0, rep(0, mcmc))
> sigma <- dlmSvd2var(filtro$U.C, filtro$D.C)[[N + 1]]
> for (i in 2:(mcmc + 1)) {
+   aux <- mddStaticThetaSample(y, mod,
+     phi[i - 1], theta[i - 1, ], Sigma = sigma)
+   theta[i, ] <- aux[[1]]
+   phi[i] <- mddPhiSample(y, mod, aux[[2]], phi[i - 1])
> }
>
> # Bloco 3: Organização das amostras e apresentação dos resultados
> burnin <- 2000
> thin <- 10
> nAmostras <- (mcmc - burnin) / thin
> aux <- burnin + (1:nAmostras) * thin
> theta.final <- theta[aux, ]
> phi.final <- phi[aux]
> fe_beta <- betareg(I(food/income) ~ income +
```

```
+   persons, data = FoodExpenditure)
> aux <- cbind(theta.final, phi.final)
> quadro.resumo <- rbind(apply(aux, 2, mean), coef(fe_beta),
+   t(c(filtro$m[N + 1,], 100)), sqrt(diag(cov(aux))),
+   sqrt(diag(vcov(fe_beta))), t(c(sqrt(diag(sigma)), NA)))
> print(quadro.resumo, digits=3)
                          Intercepto    Renda  Pessoas     phi
Estimativa MDD offline        -0.619   -0.012    0.115   32.78
Estimativa betareg            -0.623   -0.012    0.118   35.61
Estimativa MDD online         -0.649   -0.012    0.127  100.00
-------------------------------------------------------------
Erro padrão MDD offline        0.236    0.003    0.040    7.67
Erro padrão betareg            0.224    0.003    0.035    8.08
Erro padrão MDD online         0.141    0.002    0.022      NA
```

Observe que os dados da série temporal que funcionam como "variável resposta" são armazenados em uma matriz `y`, na qual a primeira coluna representa as proporções e a segunda os complementos das proporções. Esse artifício é necessário para a compatibilidade com o caso multivariado. Note que, apesar de tal procedimento de armazenamento dos dados parecer um tanto burocrático, ele é útil no caso multivariado do modelo dinâmico Dirichlet.

No segundo bloco da programação, fazemos uso simbiótico dos MDDs formulados para a abordagem *online* e *offline*. Isto é, assumimos que $\phi = 100$ e executamos o processo de filtragem descrito na Seção 3.1. Os momentos m_N e C_N da distribuição a posteriori $(\theta_N|D_N)$ representam, respectivamente, uma estimativa do vetor de parâmetros da regressão e da sua matriz de covariância. Essa matriz C_N, armazenada no objeto `sigma`, é adotada como a matriz de covariância da função geradora de candidatos Gaussiana do parâmetro θ. Essa maneira de definir tal matriz é bastante interessante por considerar a estrutura de covariância do vetor θ. Note que adotar uma função geradora de candidatos Gaussiana com matriz de covariância diagonal seria bem menos eficiente.

O terceiro bloco destina-se a organizar a amostra final produzida pelo algoritmo de Gibbs e apresentar os resultados obtidos. De acordo com os procedimentos amostrais normalmente empregados na coleta dos dados gerados pelas cadeias de Markov,

optamos por descartar as primeiras 2000 amostras e tomar uma a cada 10 das restantes. As últimas 7 linhas do código apresentam as estimativas dos parâmetros da regressão, juntamente com os respectivos desvios padrões, para cada uma das opções consideradas: modelo clássico de regressão Beta, MDD *online* e MDD *offline*.

Observe que os modelos de regressão Beta e MDD *offline* produziram resultados muito parecidos. Já o modelo dinâmico Dirichlet recursivo (MDD *online*), apesar de ter gerado estimativas pontuais razoáveis, subestimou a incerteza relativa ao parâmetro θ. Isso era de se esperar, visto que assumimos como conhecido o parâmetro de precisão $\phi = 100$, enquanto o valor real, muito provavelmente, está entre 20 e 50: quanto mais próximo de 35 fosse o valor escolhido de ϕ, mais parecidos seriam os modelos. Da mesma forma, se fosse assumido $\phi = 300$, por exemplo, o modelo já seria muito inadequado.

Cabe destacar que a escolha de ϕ define a matriz de covariância da função geradora de candidatos na estimação via MCMC. Por isso, se o valor atribuído a ele for muito inadequado, é possível que a taxa de aceitação se torne ruim (muito baixa ou muito alta) ou os dados simulados sejam muito correlacionados. Alternativamente, como mencionado anteriormente, podemos *varrer* o espaço paramétrico de ϕ em busca do valor que minimiza, por exemplo, a soma dos resíduos absolutos. Ao se adotar esse procedimento *ad-hoc*, obtivemos um valor $\widetilde{\phi} = 16$, que proporcionou estimativas razoáveis para os parâmetros da regressão.

4.1.3 MDD: Uma aplicação a dados de internações no DF

Nesta subseção será ajustado um modelo dinâmico Dirichlet a uma série temporal univariada. O conjunto de dados em questão é relativo ao sistema de saúde do Distrito Federal. Brasília é uma cidade caracterizada por um longo período sem chuvas. A seca, definida pela baixa umidade relativa do ar, costuma se mostrar mais forte justamente no período mais frio. Entre os efeitos dessa condição climática no organismo humano, podemos citar a maior incidência de doenças do aparelho respiratório.

Estamos interessados na modelagem da proporção das internações no distrito Federal decorrentes de doenças do aparelho respiratório. Cada internação é registrada

de acordo com a atual *Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde*, designada pela sigla CID-10. Essa classificação é subdividida em capítulos e, nessa aplicação, será avaliada a evolução temporal da participação do capítulo X (Doenças do aparelho respiratório) no total de internação. Pelo site do DATASUS (*http://www2.datasus.gov.br*), coletamos os dados mensais de internações e geramos uma série histórica trimestral das mencionadas proporções no período de 1998 a 2009. O ano de 2010 foi descartado do ajuste, a fim de permitir uma melhor visualização do processo de previsão desenvolvido posteriormente.

Adotamos um modelo polinomial de ordem 2 com 4 períodos sazonais e sem covariáveis. Mantendo a notação da Subseção 4.1.1, o modelo dinâmico Dirichlet é descrito através das seguintes matrizes:

$$F_t = (E_2, E_2, 1)' \quad \text{e} \quad G = \begin{pmatrix} J_2(1) & \mathbf{0} \\ \mathbf{0} & \mathcal{G}_{par} \end{pmatrix}.$$

Código 4.3: Ajuste de dados de internações no Distrito Federal

```
> # Bloco 1: Leitura dos dados e especificação do modelo
> aux <- read.table("Dados - Internações.txt")
> y <- ts(cbind(aux, 1 - aux), start = c(1998, 1), frequency = 4)
> y <- window(y, start = c(1998, 1), end = c(2009, 4))
> N <- nrow(y)
> mod <- dlmModPoly(2) + dlmModTrig(4)
> mod$GG[3, 3] <- mod$GG[4, 4] <- 0
> disc <- c(0.9, 0.9)
> dim.disc <- c(2, 3)
> dim.sys <- length(mod$m0)
> mod$C0 <- diag(10, dim.sys)
> phi0 <- 1000
>
> # Bloco 2: Geração, via Gibbs, de amostras da posteriori
> mcmc <- 100000
> Theta <- array(dim = c(N, dim.sys, mcmc + 1))
> W <- array(0, dim = c(dim.sys, dim.sys, mcmc + 1))
> theta0 <- rbind(mod$m0, matrix(nr = mcmc, nc = dim.sys))
> phi <- c(phi0, rep(0, mcmc))
> aux <- mddFilter(y, mod, phi0 , disc, dim.disc)
```

```
> W[, , 1] <- mod$W <- dlmSvd2var(aux$U.W, aux$D.W)[[N]]
> filtro <- mddFilter(y, mod, phi0)
> Theta[, , 1] <- dropFirst(mddSmooth(filtro)$s)
> Si <- diag(1e-3, dim.sys)
> for (i in 2:(mcmc + 1)) {
+   filtro <- mddFilter(y, mod, phi[i - 1])
+   Theta[, , i] <- mddThetaSample(filtro, y, mod,
+     Theta[, , i - 1], theta0[i - 1, ], phi[i - 1])
+   W[1:2, 1:2, i] <- mddWSample(mod, Theta[, , i],
+     theta0[i - 1, ], 2 / 2, Si[1:2, 1:2], 1, 2)
+   W[3:5, 3:5, i] <- mddWSample(mod, Theta[, , i],
+     theta0[i - 1, ], 3 / 2, Si[3:5, 3:5], 3, 5)
+   mod$W <- W[, , i]
+   theta0[i, ] <- mddTheta0Sample(mod, Theta[, , i])
+   phi[i] <- mddPhiSample(y, mod, Theta[, , i], phi[i - 1])
> }
```

A estratégia sugerida na Subseção 3.2.1, de especificação da distribuição a priori conjunta dos parâmetros desconhecidos, foi aqui adotada. A forma de definição dos parâmetros iniciais $\Theta^{(0)}$ e $W^{(0)}$ e organização das amostras geradas é a mesma do caso simulado na Subseção 4.1.1. Dessa forma, os Códigos 4.2 e 4.3 são bastante similares, e podemos seguir à análise dos resultados.

A Figura 4.2 apresenta, nessa ordem, a série observada y_t, e as estimativas das médias μ_t, dos níveis θ_{1t} e dos efeitos de sazonalidade $\theta_{3t} + \theta_{5t}$. O ajuste do modelo nos permite descrever o resultado observado em termos dos fatores que o determinam. Essa representação do fenômeno pode ser útil na definição e avaliação de políticas públicas de saúde.

A título de ilustração, uma ação estruturada de combate ao tabagismo estaria mais facilmente ligada a mudanças do nível do que do comportamento cíclico da série. Isso é, trata-se de uma medida que busca a diminuição da população pertencente ao grupo de risco, e não uma intervenção relativa aos períodos mais críticos do ano. Nesse sentido, se o foco principal da análise fosse a avaliação do impacto da campanha no nível não observável da série, estaríamos particularmente interessados no acompanhamento do quarto gráfico da Figura 4.2. Desse modo, o modelo representaria um diagnóstico capaz de destacar o aspecto ao qual estamos interessados.

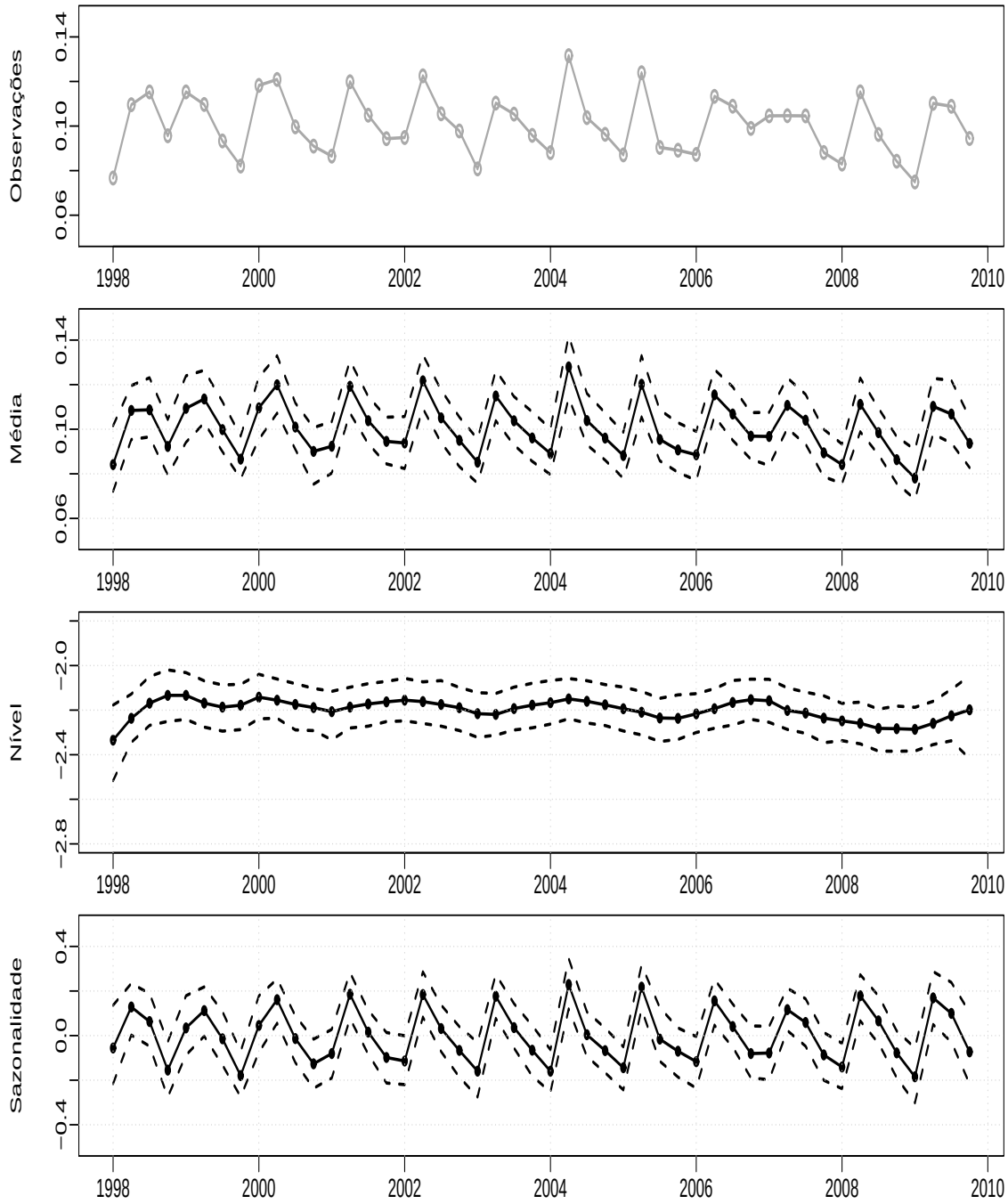

Figura 4.2: (a) Serie histórica y_t (linhas cinzas); (b) Médias μ_t; (c) Níveis θ_{1t}; (d) Efeitos de sazonalidade $\theta_{3t} + \theta_{5t}$. Valores estimados (linhas pretas) e intervalos de credibilidade de 95% (linhas pretas tracejadas).

Suponha que a baixa umidade relativa do ar fosse o principal motivo da variação periódica da série. Neste caso, se o governo decidisse determinar a obrigatoriedade do uso de umidificadores de ar em escolas e órgãos públicos, o efeito, se fosse significativo, poderia influenciar o comportamento cíclico da série. Esperaríamos observar um achatamento da amplitude da curva presente no terceiro gráfico da Figura 4.2. Seria uma evidência de que a determinação governamental foi útil em reduzir os efeitos relativos ao período de seca. Note que a leitura direta dos dados dificulta a visualização da mudança do padrão sazonal.

O modelo poderia ser aprimorado para incorporar o efeito de covariáveis, como os níveis de poluição do ar, por exemplo. Para lidar com o efeito cumulativo da poluição, provavelmente o uso de funções de transferência seria recomendável.

Por fim, usar o modelo para fazer previsões pode ser útil no planejamento governamental. Cada ponto simulado da posteriori $(\Theta, \psi|D_t)$ nos permite gerar valores da distribuição a priori expressa por

$$p(\theta_{T+1}, \ldots, \theta_{T+m}, y_{T+1}, \ldots, y_{T+m}|D_T).$$

Para tanto, usamos a seguinte relação:

$$p(\theta_{T+1}, \ldots, \theta_{T+m}, y_{T+1}, \ldots, y_{T+m}, \psi, \theta_t|D_T) =$$
$$p(\theta_{T+1}, \ldots, \theta_{T+m}, y_{T+1}, \ldots, y_{T+m}|\psi, \theta_T)p(\theta_t, \psi|D_t).$$

A Figura 4.3 apresenta a previsão feita para cada trimestre de 2010. No intuito de destacar o período de previsão, apresentamos apenas os dados a partir do ano de 2007. Vale salientar que a previsão foi feita com o ajuste baseado em toda a série histórica (de 1998 a 2009). Observe que o intervalo de credibilidade de 90% se torna cada vez mais impreciso na medida em que aumenta a distância temporal entre o momento atual (fim de 2009) e o momento previsto. As estimativas são um tanto imprecisas. Futuros aprimoramentos, como a inclusão de covariáveis, podem tornar a distribuição preditiva mais precisa, potencializando os benefícios do processo de previsão.

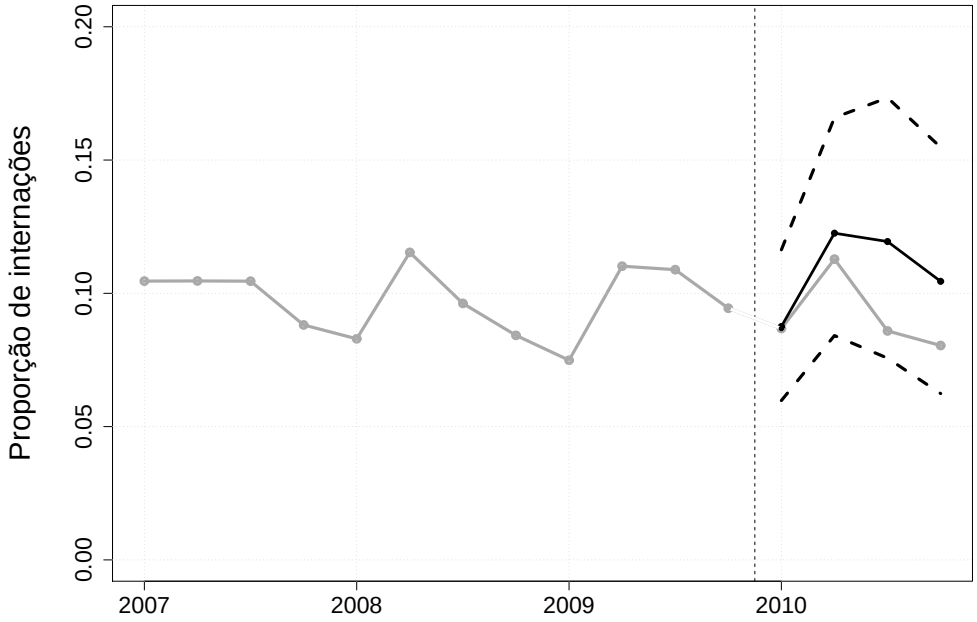

Figura 4.3: Proporção de internações por doenças do aparelho respiratório no DF (linhas cinzas) e valores previstos para o ano de 2010 (linha preta) com respectivos intervalos de credibilidade de 90% (linhas pretas tracejadas).

4.2 Modelos multivariados - Caso Dirichlet

Nesta seção retornamos ao tema central dessa dissertação, que trata do modelo dinâmico Dirichlet no caso geral multivariado. Iniciamos a exposição ajustando um conjunto de dados simulados. Tal procedimento nos auxilia a discutir certas particularidades dos modelos dinâmicos Dirichlet no contexto multivariado. Em seguida, abordamos o caso estático, correspondente ao *modelo de regressão Dirichlet*. Por fim, na Subseção 4.2.3 modelamos dados relativos ao perfil das taxas de óbitos por causas externas no Brasil.

4.2.1 Ajuste de dados simulados

Nesta seção ajustamos um conjunto de dados simulados a partir de um modelo dinâmico Dirichlet envolvendo dados composicionais com 3 categorias. Isso nos permite avaliar a qualidade do ajuste baseado na proposta de estimação via cadeias de Markov. Optamos por utilizar um MDD caracterizado por uma tendência de cresci-

mento polinomial, definido pelas seguintes matrizes:

$$\left\{ F = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},\ G = \begin{pmatrix} J_2(1) & \mathbf{0} \\ \mathbf{0} & J_2(1) \end{pmatrix},\ W = \text{diag}(W_1, \ldots, W_4),\ \phi = 500 \right\},$$

com $W_1 = W_3 = 10^{-3}$ e $W_2 = W_4 = 10^{-4}$. Os valores de ϕ e W foram estrategicamente especificados de forma a garantir certa similaridade entre o modelo simulado nesta seção e o MDD aplicado a dados reais na Subseção 4.2.3.

A Equação de evolução do referido modelo pode, então, ser expressa, para $t = 1, \ldots, N$, por:

$$\begin{aligned} \theta_{1t} &= \theta_{1,t-1} + \theta_{2,t-1} + \omega_{1,t}, \\ \theta_{2t} &= \qquad\quad \theta_{2,t-1} + \omega_{2,t}, \\ \theta_{3t} &= \theta_{3,t-1} + \theta_{4,t-1} + \omega_{3,t}, \\ \theta_{4t} &= \qquad\quad \theta_{4,t-1} + \omega_{4,t}. \end{aligned}$$

Dessa forma, os parâmetros de tendência θ_{2t} e θ_{4t} seguem um passeio aleatório, enquanto os parâmetros de nível θ_{1t} e θ_{3t} estão sujeitos a uma evolução localmente linear.

A partir da aplicação da matriz F ao vetor θ_t, obtemos $\lambda_t = (\lambda_{1t}, \lambda_{2t})' = (\theta_{1t}, \theta_{3t})'$, e pela função de ligação logito aditiva, $\mu_t = (\mu_{1t}, \mu_{2t})'$, com

$$\mu_{1t} = \frac{e^{\theta_{1t}}}{1 + e^{\theta_{1t}} + e^{\theta_{3t}}} \quad \text{e} \quad \mu_{2t} = \frac{e^{\theta_{3t}}}{1 + e^{\theta_{1t}} + e^{\theta_{3t}}}.$$

Portanto, os parâmetros de nível θ_{1t} e θ_{3t} definem a média da distribuição das observações.

Como cada observação y_t corresponde a um vetor tridimensional, tal que a soma de suas entradas é igual a 1, modelamos diretamente apenas duas das três categorias. De todo modo, o comportamento da terceira categoria também é descrito, ainda que indiretamente, à medida que as médias se relacionam entre si. Se, para um dado t, θ_{1t} e θ_{3t} são ambos negativos, as médias μ_{1t} e μ_{2t} são, necessariamente, inferiores a $1/3$, e consequentemente, a participação da terceira categoria será expressiva.

Inclusão de covariáveis no modelo

Embora não tenhamos, aqui, incluído qualquer covariável ao modelo, nossa proposta do MDD, como mencionado anteriormente, foi feita de forma a permitir tal inclusão. Há, contudo, uma particularidade que deve ser salientada. A título de ilustração, considere os processos judiciais protocolados em um Tribunal de Justiça estadual. Tais processos podem ser classificados como de *primeira instância*, de *segunda instância* ou de *juizado especial*. Suponha que haja interesse no monitoramento da participação de cada uma dessas 3 categorias no volume de processos novos. Neste caso, como decidir qual delas será modelada indiretamente? Se o interesse em cada uma dessas classes de processos é o mesmo, é possível *separar* aquela em que há maior dificuldade em se descrever.

Ainda considerando o exemplo do Tribunal, poder-se-ia adotar, como variável explicativa para os processos novos de segunda instância, o número de processos novos de primeira instância no período anterior. Para os casos novos de juizado especial, por sua vez, uma covariável candidata seria o índice de acordos feitos no período anterior. Nessas condições, a categoria *primeira instância* não seria modelada diretamente por se considerar mais difícil sua descrição, em comparação às demais classes de processos. Essa escolha poderia, também, ser justificada por um possível interesse acerca da relação entre as covariáveis citadas e as médias μ_t.

Observe que, a depender do cenário, uma mesma variável pode ser usada para explicar mais de uma categoria. No contexto multivariado, há incontáveis possibilidades de especificação das matrizes F e G. A sensibilidade e compreensão do fenômeno, por parte do analista, são determinantes para a qualidade e utilidade do modelo.

Análise dos dados simulados

Retornando ao modelo em estudo nessa seção, com o primeiro bloco do Código 4.4, geramos observações do modelo em questão. Assim como no caso Beta simulado, definimos m_0 aleatoriamente e fizemos $\theta_0 \approx m_0$. Para tanto, definimos $C_0 = 10^{-5}\,\mathbf{I}$. Novamente, observe que não estamos adotando uma priori precisa para θ_0 na estimação dos parâmetros do modelo. Estamos, simplesmente, definindo o estado inicial do sistema no processo de geração de uma amostra do modelo. Na última linha do Bloco 1, removemos, aleatoriamente, 20% das observações y_t para ilustrar como lidar com

observações faltantes.

Código 4.4: Ajuste de dados simulados multivariados

```
> # Bloco 1: Geração dos dados e definição do modelo
> set.seed(111)
> N1 <- 50
> FF <- matrix(0, 2, 4)
> FF[1, 1] <- FF[2, 3] <- 1
> aux <- dlmModPoly(2)$GG
> GG <- bdiag(aux, aux)
> W.real <- diag(rep(c(1e-3, 1e-4), 2))
> m0 <- rnorm(4, sd = rep(c(0.5, 0.01), 2))
> C0 <- diag(1e-5, 4)
> mod <- list(FF = FF, GG = GG,
+   W = W.real, m0 = m0, C0 = C0)
> phi.real <- 500
> modObs <- rMdd(mod, phi.real, N1)
> N <- 40
> y <- modObs$y[1:N, ]
> mu.real <- modObs$mu[1:N, ]
> theta.real <- modObs$theta[1:N, ]
> y[sample(2:(N - 1), N / 10), ] <- NA
>
> # Bloco 2: Especificação dos valores iniciais do Gibbs
> disc <- c(0.8, 0.8)
> dim.disc <- c(2, 2)
> dim.sys <- length(mod$m0)
> mod$m0 <- rep(0, dim.sys)
> mod$C0 <- diag(10, 4)
> phi0 <- 250
> mcmc <- 200000
> Theta <- array(dim = c(N, dim.sys, mcmc + 1))
> W <- array(0, dim = c(dim.sys, dim.sys, mcmc + 1))
> theta0 <- rbind(mod$m0, matrix(nr = mcmc, nc = dim.sys))
> phi <- c(phi0, rep(0, mcmc))
> aux <- mddFilter(y, mod, phi0 , disc, dim.disc)
> W[, , 1] <- mod$W <- dlmSvd2var(aux$U.W, aux$D.W)[[N]]
> filtro <- mddFilter(y, mod, phi0)
```

```
> Theta[, , 1] <- dropFirst(mddSmooth(filtro)$s)
> linhasNA <- order(y[, 1], na.last = F)[1:sum(is.na(y[, 1]))]
> Si <- diag(1e-3, dim.sys)
> controle <- 0
>
> # Bloco 3: Geração, via Gibbs, de amostras da posteriori
> for (i in 2:(mcmc + 1)) {
+   if(controle >= 100) {
+     filtro <- mddFilter(y, mod, phi[i - 1])
+     controle <- 0
+   }
+   y[linhasNA, ] <- mddYSample(mod,
+     Theta[, , i - 1], phi[i - 1], linhasNA)
+   Theta[, , i] <- mddThetaSample(filtro, y, mod,
+     Theta[, , i - 1], theta0[i - 1, ], phi[i - 1])
+   mod$W <- W[, , i] <- mddWSample(mod, Theta[, , i],
+     theta0[i - 1, ], 4 / 2, Si, 1, 4)
+   theta0[i, ] <- mddTheta0Sample(mod, Theta[, , i])
+   phi[i] <- mddPhiSample(y, mod, Theta[, , i], phi[i - 1])
+   tmp <- identical(Theta[, , i], Theta[, , i - 1])
+   controle <- ifelse(tmp, controle + 1, 0)
+ }
```

No segundo bloco do Código 4.4, definimos, com base no processo de filtragem descrito na Subseção 3.1.1, os valores iniciais $\Theta^{(0)}$ e $W^{(0)}$ do algoritmo MCMC.

O terceiro bloco do código foi construído para executar o algoritmo de estimação *offline*, amostrando das distribuições condicionais completas. É instrutivo notar que adotamos uma estratégia extremamente útil no caso multivariado: o procedimento de filtragem relativo ao método *CUBS* não é feito a cada iteração. Em outras palavras, não atualizamos a função geradora de candidatos a cada passo, e sim quando nenhum dos 100 (número escolhido arbitrariamente) primeiros candidatos gerados por ela é aceito. Observe a função desempenhada pelo objeto `controle` no código.

Recorde que a cada iteração, o processo de filtragem é responsável por praticamente todo o tempo computacional demandado. Felizmente, o ganho em qualidade da função geradora de candidatos, decorrente da atualização de seus momentos, é muito baixo entre iterações próximas na cadeia de Markov. Dessa forma, podemos,

tanto do ponto de vista prático como teórico, evitar o gasto de tempo desnecessário com tal atividade. Por outro lado, se a atualização for demasiadamente rara, a função geradora de candidatos se tornará obsoleta, ineficiente e incompatível com o estado corrente da cadeia. Em resumo, caso não tivéssemos adotado tal medida, o tempo computacional do algoritmo MCMC seria proibitivo, podendo demorar mais de um mês para ser finalizado.

Visto que não há covariáveis ou sazonalidade no modelo aqui considerado, o interesse retrospectivo principal do ajuste refere-se às estimativas das médias das categorias. Neste ajuste, optamos por descartar as primeiras 75000 amostras e utilizar apenas uma a cada 100 das restantes. A Figura 4.4 exibe, além dos valores simulados para a variável resposta em cada uma das categorias (linhas cinzas), as médias reais (linhas vermelhas) e as estimativas das médias (linhas pretas). Mais uma vez, o modelo foi capaz de recuperar as médias não observáveis que geraram os dados. A moda e a mediana da distribuição a posteriori de ϕ foram, respectivamente, 602.22 e 623.73 (o valor real é 500). Ressaltamos que a moda foi obtida pela maximização da densidade empírica, suavizada pela função *R density*, das amostras MCMC do parâmetro ϕ.

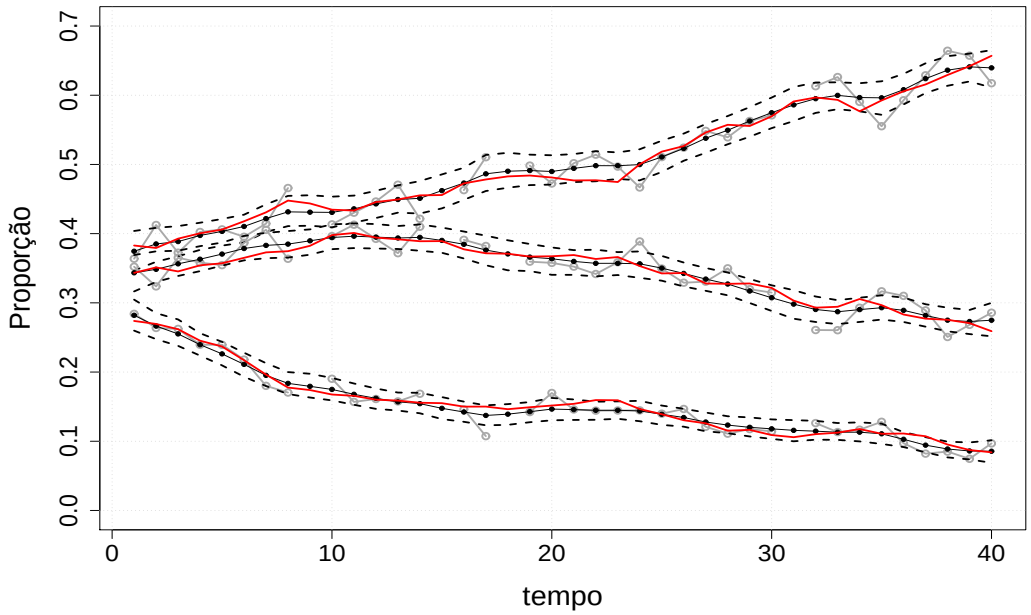

Figura 4.4: Proporções observadas (linhas cinzas), médias reais (linhas vermelhas) e médias estimadas (linha preta) com intervalos de credibilidade de 95% (linhas pretas tracejadas).

Cabe salientar que, pela maneira como foi formulado o modelo, só há dois parâmetros de média a cada tempo, μ_{1t} e μ_{2t}, um para cada uma das duas primeiras categorias. Por outro lado, podemos definir, como o fizemos na formulação do modelo, $\mu_{3t} = 1 - (\mu_{1t} + \mu_{2t})$. Dessa forma, cada amostra MCMC de μ_t está associada um valor μ_{3t}, e a partir daí, construímos um intervalo de credibilidade quantílico para μ_{3t}.

Representação da sazonalidade no modelo

Com base nos procedimentos de especificação das matrizes F e G, descritos na Seção 1.4, é possível incorporar o efeito da sazonalidade no comportamento de cada categoria modelada. Considerando o modelo ajustado nesta subseção, caso o fenômeno apresente 4 períodos sazonais, é possível incorporar a sazonalidade definindo-se as seguintes matrizes:

$$F = \begin{pmatrix} E_2 & E_2 & 1 & \mathbf{0} & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{0} & 0 & E_2 & E_2 & 1 \end{pmatrix} \quad \text{e} \quad G = \begin{pmatrix} G^* & \mathbf{0} \\ \mathbf{0} & G^* \end{pmatrix}, \text{ com } \quad G^* = \begin{pmatrix} J_2(1) & \mathbf{0} \\ \mathbf{0} & \mathcal{G}_{par} \end{pmatrix}.$$

Recorde que adotamos a representação trigonométrica da sazonalidade. Dessa forma, incluímos, a cada uma das duas categorias modeladas, dois parâmetros de sazonalidade (um para cada harmônico). As matrizes F e G apresentadas acima implicam em $\lambda_t = (\lambda_{1t}, \lambda_{2t})$, com:

$$\lambda_{1t} = \theta_{1t} + \theta_{3t} + \theta_{5t},$$
$$\lambda_{2t} = \theta_{6t} + \theta_{8t} + \theta_{10t}.$$

Os parâmetros θ_{1t} e θ_{6t} representam, respectivamente, o nível das categorias A e B. θ_{3t} e θ_{5t} representam o estado de cada um dos harmônicos referentes a A. De maneira análoga, θ_{8t} e θ_{10t} retratam o estado dos harmônicos de B. A extensão para cenários onde o período sazonal é diferente de 4 é imediata adotando-se a estratégia de definição de F e G já discutida.

4.2.2 Regressão Dirichlet - estimação num contexto simulado

Nesta subseção mostraremos como o modelo dinâmico Dirichlet pode ser especificado de modo a se apresentar como um modelo de regressão Dirichlet. Em seguida, ajustaremos um conjunto de dados simulados a fim de mostrar a capacidade do método inferencial em estimar os parâmetros da regressão.

O modelo de regressão Dirichlet pode ser obtido fazendo-se $G_t = \mathbf{I}$ e $W = \mathbf{0}$. Na regressão Dirichlet, o índice t não mais denota "tempo"e sim "observação". Dessa forma, obtém-se a seguinte estrutura:

- **Equação das observações:**

$$(Y_t|\mu_t, \phi) \sim \text{Dirichlet}(\mu_{1t}\phi, \mu_{2t}\phi, \ldots, \mu_{kt}\phi) \text{ , onde:}$$
$$\mu_{kt} = 1 - \sum_{i=1}^{k-1} \mu_{it}, \quad y_{kt} = 1 - \sum_{i=1}^{k-1} y_{it}; \quad 0 < y_{it}, \mu_{it} < 1 \quad \forall i;$$
$$\phi > 0 \quad \text{e} \quad \mu_t = (\mu_{1t}, \ldots, \mu_{k,t-1}).$$

- **Função de ligação: logito aditiva**

$$\lambda_t = F_t'\theta = \left[\log\left(\frac{\mu_{1t}}{\mu_{kt}}\right), \ldots, \log\left(\frac{\mu_{k-1,t}}{\mu_{kt}}\right)\right]'.$$

- **Distribuição a priori:**

$$(\theta, \phi) \sim p(\theta)p(\phi) = \text{N}(m_0, C_0)\ \text{Gama}(\alpha, \beta).$$

Para dar prosseguimento ao estudo do modelo de regressão Dirichlet, vamos considerar o caso em que as observações retratam a participação (taxa) de cada uma das 3 categorias (A, B e C) no fenômeno. Além disso, cada uma das categorias *livres* (A e B) são modeladas por duas covariáveis e um intercepto. Isto é, o vetor do sistema, θ, tem dimensão igual a 6 e se relaciona ao preditor linear $\lambda_t = (\lambda_{1t}, \lambda_{2t})'$ pelas seguintes

equações:

$$\lambda_{1t} = \theta_1 + x_{1A,t}\theta_2 + x_{2A,t}\theta_3,$$
$$\lambda_{2t} = \theta_4 + x_{1B,t}\theta_5 + x_{2B,t}\theta_6,$$

onde $x_{1A,t}$ é o valor da 1ª covariável, no tempo t, relativa à categoria A. A mesma regra se aplica às demais covariáveis. Observe que estamos usando informações auxiliares para descrever a média μ_t, que em conjunto com o parâmetro ϕ, define a distribuição das observações. Observe, ainda, que estamos adequadamente considerando as restrições impostas pelo fenômeno. Isto é, a média de uma categoria só pode aumentar em detrimento de pelo menos uma das restantes. Esse resultado não seria válido se estivéssemos modelando as proporções em cada categoria separadamente, em uma abordagem univariada.

A estrutura acima obtida é naturalmente descrita definindo-se os seguintes componentes do modelo dinâmico Dirichlet

$$\left\{ F_t = \begin{pmatrix} 1 & x_{1A,t} & x_{2A,t} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_{1B,t} & x_{2B,t} \end{pmatrix}, \; G = \mathbf{I}, \; W = \mathbf{0}, \; \phi = 500 \right\}.$$

Seguindo o tratamento dado aos casos discutidos anteriormente, usamos o primeiro bloco do Código 4.5 para definir o modelo e dele gerar uma observação. Ressaltamos o fato de termos de criar o objeto `JFF` para lidar com o armazenamento das matrizes F_t. Recorde que adotamos a mesma estratégia de especificação de modelos do pacote *dlm*. Assim como no caso univariado, utilizamos o processo de filtragem para definir a matriz de covariância da função geradora de candidatos Gaussiana.

Código 4.5: Modelo de regressão Dirichlet

```
> # Bloco 1: Geração dos dados e definição do modelo
> set.seed(111)
> N <- 60
> FF <- diag(2) %x% matrix(c(1, 1, 1), nr = 1)
> GG <- diag(6)
> W <- diag(0, 6)
> JFF <- matrix(c(0:2, rep(0, 7), 3:4), 2, 6, byrow = T)
```

```
> X <- matrix(rnorm(4 * N), nc = 4)
> m0 <- rnorm(6, sd = c(0.7, 0.2, 0.2))
> C0 <- diag(1e-5, 6)
> mod <- list(FF = FF, GG = GG,
+   W = W, JFF = JFF, X = X, m0 = m0, C0 = C0)
> phi.real <- 300
> modObs <- rMdd(mod, phi.real, N)
> y <- modObs$y
> mu.real <- modObs$mu
> theta.real <- modObs$theta
>
> # Bloco 2: Estimação do modelo
> mcmc <- 20000
> dim.sys <- length(mod$m0)
> mod$m0 <- rep(0, dim.sys)
> mod$C0 <- diag(10, 6)
> phi0 <- 100
> filtro <- mddFilter(y, mod, phi0)
> theta <- rbind(filtro$m[N + 1, ], matrix(nr = mcmc, nc = dim.sys))
> phi <- c(phi0, rep(0, mcmc))
> sigma <- dlmSvd2var(filtro$U.C, filtro$D.C)[[N + 1]]
> for (i in 2:(mcmc + 1)) {
+   aux <- mddStaticThetaSample(y, mod,
+     phi[i - 1], theta[i - 1, ], Sigma = sigma)
+   theta[i, ] <- aux[[1]]
+   phi[i] <- mddPhiSample(y, mod, aux[[2]], phi[i - 1])
+ }
>
> # Bloco 3: Organização das amostras e apresentação dos resultados
> burnin <- 100
> thin <- 10
> nAmostras <- (mcmc - burnin) / thin
> aux <- burnin + (1:nAmostras) * thin
> theta.final <- theta[aux, ]
> phi.final <- phi[aux]
> aux <- cbind(theta.final, phi.final)
> parm.hat <- apply(aux, 2, mean)
> nomes <- c("IntA", "Reg1A", "Reg2A",
```

```
+   "IntB", "Reg1B", "Reg2B", "phi")
> quadro.resumo <- round(rbind(c(m0, phi.real),
+   parm.hat, sqrt(diag(cov(aux)))), 4)
> colnames(quadro.resumo) <- nomes
> rownames(quadro.resumo) <- c("Valores reais",
+   "Valores estimados", "Erro padrão")
> print(quadro.resumo, digits=2)
                     IntA  Reg1A Reg2A  IntB Reg1B  Reg2B     phi
Valores reais       0.665 0.0047 0.349 0.482 0.067 -0.037 300.000
Valores estimados   0.624 0.0165 0.351 0.451 0.042 -0.028 296.465
Erro padrão         0.020 0.0138 0.017 0.020 0.015  0.018  39.442
```

No segundo bloco do Código 4.5, usamos as funções descritas anteriormente para gerar amostras da distribuição a posteriori de (θ, ϕ). O terceiro bloco do código, por sua vez, organiza as amostras e apresenta uma tabela contendo os valores reais e estimados para os parâmetros da regressão $\theta_1, \ldots, \theta_6$.

No intuito de facilitar a comparação dos valores reais e estimados, exibimos, na Figura 4.5, os valores reais (pontos vermelhos) e os respectivos intervalos *HPD* de 95%. O eixo das abscissas representa o índice i do vetor θ_i. Da Figura 4.5, é imediato perceber a boa capacidade do modelo em recuperar os valores dos parâmetros θ que geraram a série observada y_t. Note que aqui estamos comparando as estimativas com os valores reais dos parâmetros. Na Subseção 4.1.2, a comparação era entre as estimativas do MDD e as estimativas do modelo clássico de regressão Beta.

Há, neste momento, um projeto formal de desenvolvimento de um pacote *R* relacionado ao modelo clássico de regressão Dirichlet, denominado *DirichletReg*. Quando os autores disponibilizarem este pacote, as rotinas nele implementadas poderão ser usadas na comparação do modelo clássico de regressão Dirichlet com o MDD (Bayesiano) introduzido nesta dissertação.

O caso estático, aqui considerado, é extremamente mais simples, do ponto de vista operacional, que o modelo dinâmico equivalente. Não é preciso lidar com as matrizes W, além de só ser preciso executar o passo de filtragem uma única vez. O tempo computacional necessário para estimar um modelo com $W = \mathbf{0}$ é muitas vezes inferior ao requerido pelo modelo equivalente com $W \neq \mathbf{0}$.

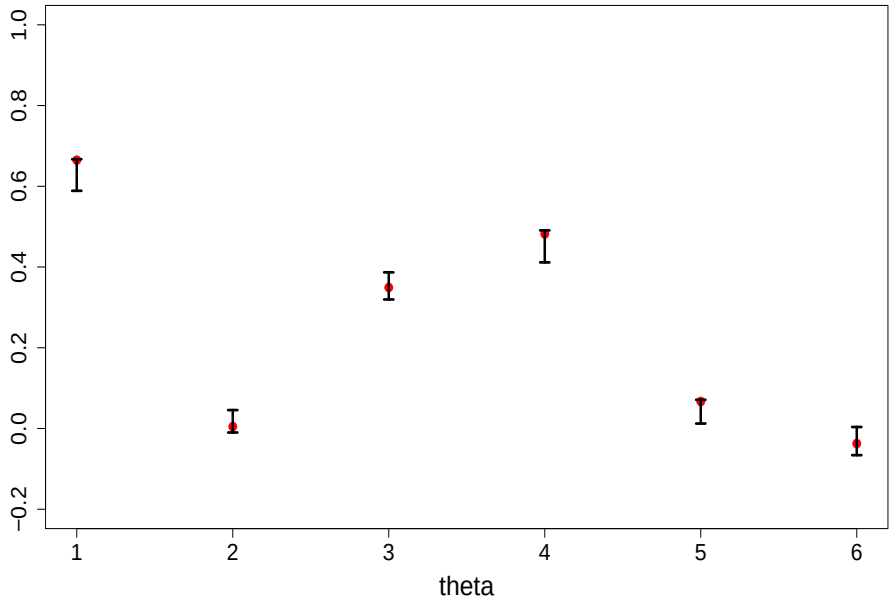

Figura 4.5: Valores reais (pontos vermelhos) e estimativas intervalares de credibilidade de 95% (linhas pretas) dos parâmetros da regressão θ_i.

4.2.3 MDD: Uma aplicação a dados de óbitos no Brasil

Finalizaremos este capítulo ajustando um Modelo Dinâmico Dirichlet a um conjunto de dados reais, também relacionados à saúde pública. O capítulo XX: *causas externas de morbidade e de mortalidade*; da atual *Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde* - CID-10 - está seccionado de maneira que nos permite estabelecer as seguintes categorias:

A - Acidentes de transporte;

B - Agressões e lesões autoprovocadas intencionalmente;

C - Outras causas externas de mortalidade.

Do ponto de vista social, a redução dos acidentes de trânsito fatais e da violência, seja ela a terceiros ou a si próprio, são objetivos fundamentais de qualquer governo. Dessa maneira, busca-se o aumento da contribuição de C (outras causas externas de mortalidade), em relação ao volume total de óbitos. Note, de todo modo, que é somente desejável o aumento da contribuição de C, ou seja, a mudança do perfil dos óbitos no país. A redução absoluta, tanto dessa quanto das demais categorias, é obviamente apreciável.

Pelo site do DATASUS (*http://www2.datasus.gov.br*), extraímos os dados mensais, em nível nacional, de óbitos por causas externas e geramos uma série histórica trimestral das proporções de cada uma das categorias A, B e C no período de 1998 a 2010. Ajustamos aos dados um modelo polinomial equivalente ao utilizado na Subseção 4.2.1. Isto é,

$$F = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{e} \quad G = \begin{pmatrix} J_2(1) & \mathbf{0} \\ \mathbf{0} & J_2(1) \end{pmatrix}.$$

O primeiro bloco do Código 4.6 trata da leitura dos dados armazenados em um documento de texto (*.txt*) e da especificação do modelo. As funções de especificação do pacote *dlm* já não são suficientes para se construir o objeto `mod`. Ou seja, é preciso definir, separadamente, cada objeto `F`, `G`, dentre outros, e agregá-los usando a função *list*.

Código 4.6: Ajuste de dados sobre óbitos por causas externas

```
> # Bloco 1: Leitura dos dados e especificação do modelo
> aux <- read.table("Dados - Óbitos por causa.txt")
> y <- ts(aux, start = c(1998, 1), frequency = 4)
> N <- nrow(y)
> FF <- matrix(0, 2, 4)
> FF[1, 1] <- FF[2, 3] <- 1
> aux <- dlmModPoly(2)$GG
> GG <- bdiag(aux, aux)
> m0 <- rep(0, 4)
> C0 <- diag(10, 4)
> mod <- list(FF = FF, GG = GG, W = W, m0 = m0, C0 = C0)
> dim.sys <- length(mod$m0)
>
> # Bloco 2: Atribuição dos valores iniciais do Gibbs
> mcmc <- 20000
> phi0 <- 500
> disc <- c(0.85, 0.85)
> dim.disc <- c(2, 2)
> Theta <- array(dim = c(N, dim.sys, mcmc + 1))
> W <- array(0, dim = c(dim.sys, dim.sys, mcmc + 1))
```

```
> theta0 <- rbind(mod$m0, matrix(nr = mcmc, nc = dim.sys))
> phi <- c(phi0, rep(0, mcmc))
> aux <- mddFilter(y, mod, phi0 , disc, dim.disc)
> W[, , 1] <- mod$W <- dlmSvd2var(aux$U.W, aux$D.W)[[N]]
> filtro <- mddFilter(y, mod, phi0)
> Theta[, , 1] <- dropFirst(mddSmooth(filtro)$s)
> Si <- diag(1e-3, dim.sys)
>
> # Bloco 3: Geração, via Gibbs, de amostras da posteriori
> for (i in 2:(mcmc + 1)) {
+   if(controle >= 100) {
+     filtro <- mddFilter(y, mod, phi[i - 1])
+     controle <- 0
+   }
+   Theta[, , i] <- mddThetaSample(filtro, y, mod,
+     Theta[, , i - 1], theta0[i - 1, ], phi[i - 1])
+   mod$W <- W[, , i] <- mddWSample(mod, Theta[, , i],
+     theta0[i - 1, ], 4 / 2, Si, 1, 4)
+   theta0[i, ] <- mddTheta0Sample(mod, Theta[, , i])
+   phi[i] <- mddPhiSample(y, mod, Theta[, , i], phi[i - 1])
+   tmp <- identical(Theta[, , i], Theta[, , i - 1])
+   controle <- ifelse(tmp, controle + 1, 0)
+ }
```

O bloco 2 do Código 4.6 foi desenvolvido para criar os objetos que receberão as amostras geradas e definir os valores iniciais da cadeia de Markov. O bloco 3 executa o processo de amostragem propriamente dito. Ressaltamos que, novamente, a atualização dos parâmetros da função geradora de candidatos só ocorre eventualmente.

A categoria C foi modelada indiretamente. Note que, entre as 3, ela é, do ponto de vista prático, de menor importância e representa o saldo residual (caracterizado pelo termo *outros*) dos óbitos.

Neste ajuste, utilizamos apenas uma a cada 100 amostras geradas. A Figura 4.6 exibe, sobrepostas, as proporções observadas e as estimativas da média de cada categoria. As proporções de A, B e C foram apresentadas, respectivamente, com as cores cinza, azul e marrom. As linha pretas, como de costume, denotam as estimativas intervalares.

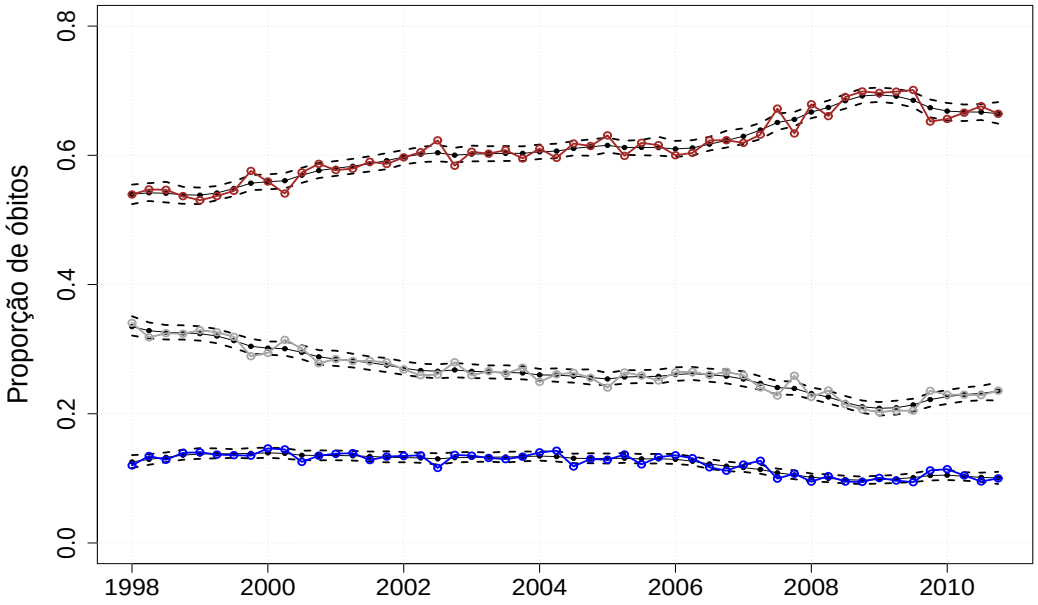

Figura 4.6: Proporções observadas de óbitos por A - acidentes de transporte (linha cinza), B - agressões e lesões autoprovocadas intencionalmente (linha azul) e C - outras causas externas de mortalidade (linha marrom). Estimativas das médias das categorias (linhas pretas) e intervalos de credibilidade de 95% (linhas pretas tracejadas).

Do início da série, em 1998, até o ano de 2009, houve uma redução contundente e sistemática da contribuição dos acidentes de transporte no total de óbitos por causas externas. Tal redução se refletiu no aumento da participação de C (outras causas externas de mortalidade). Essa constatação é motivo de satisfação, considerados os argumentos apresentados na introdução desta subseção. Não obstante, a partir do ano de 2009, a curva assumiu a direção oposta, evidenciando um retrocesso no que se refere aos acidentes de trânsito. Os óbitos decorrentes de agressões e lesões autoprovocadas intencionalmente estiveram estáveis até o ano de 2006, quando mostraram, até 2009, uma redução percentual significativa. A partir de então, estabilizaram-se nesse novo patamar.

No que se refere ao ajuste do parâmetro de precisão das observações, a moda e a mediana da distribuição a posteriori de ϕ foram, respectivamente, 2057.22 e 2043.87. O intervalo HPD, de 95%, para ϕ é (1337.86, 2819.19). Esse alto valor estimado para ϕ não surpreende. A série y_t, de fato, apresenta um ruído muito baixo. Do ponto de vista prático, dados de abrangência nacional correspondem a situações onde o erro

amostral é pouco significativo. Em outras palavras, y_t é uma medida precisa da média não observável μ_t.

A Figura 4.7 apresenta as proporções observadas a partir do ano de 2009, bem como as predições pontuais e intervalares (linhas tracejadas) para cada trimestre dos anos de 2011 e 2012. Vale salientar que as previsões foram construídas a partir do ajuste de toda a série disponível (desde 1998), ainda que, para facilitar a visualização do período de previsão, tenhamos omitido os anos iniciais. O gráfico fortalece o argumento de haver uma tendência atual de crescimento da proporção associada aos acidentes de transporte.

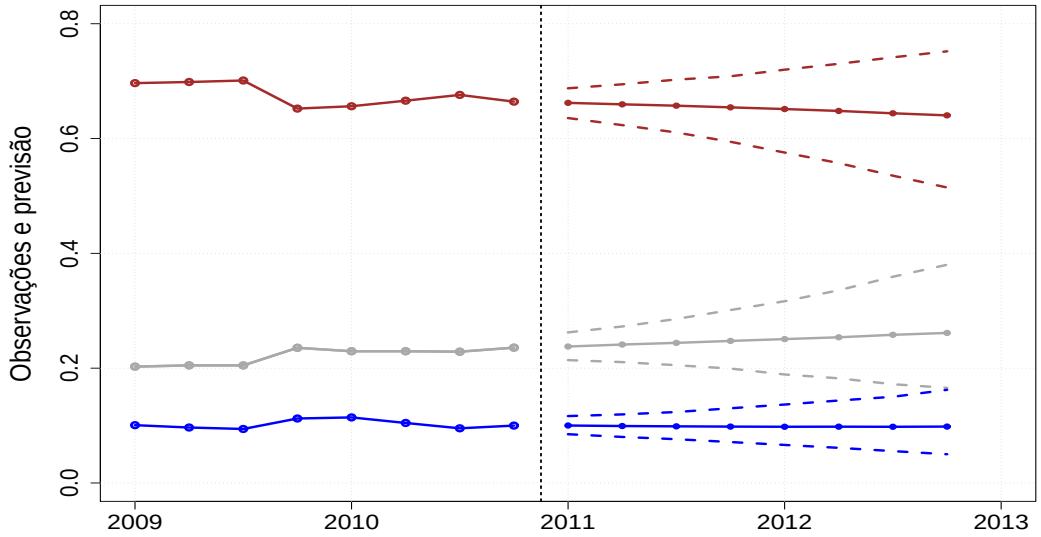

Figura 4.7: Proporções observadas de óbitos, em 2009 e 2010, por A - acidentes de transporte (linha cinza), B - agressões e lesões autoprovocadas intencionalmente (linha azul) e C - outras causas externas de mortalidade (linha marrom). Previsões das proporções para 2011 e 2012 com intervalos de credibilidade de 90% (linhas tracejadas).

O modelo, como foi construído, pode ser útil tanto no monitoramento contínuo desse importante fenômeno de saúde pública, que também envolve aspectos de violência e de transportes, quanto no estabelecimento de políticas públicas. Poderia se pensar, por exemplo, em metas baseadas na redução em relação ao cenário previsto, a exemplo do que normalmente é feito na definição de metas ambientais.

Outra possibilidade é avaliar os resultados obtidos com base nos valores ajustados, e não nos observados. O ganho está em evitar que possíveis comportamentos atípicos

prejudiquem a qualidade do diagnóstico. Ou seja, essa estratégia afasta a possibilidade da análise ser ludibriada por interferências não sustentáveis que modifiquem somente o resultado do período (y_t), sem que seja alterada de maneira significativa a média não observável do sistema (μ_t).

Capítulo 5

Conclusão e trabalhos futuros

Iniciamos este trabalho de pesquisa com a aplicação da abordagem de estimação via MCMC, em especial o método CUBS (Ravines et al., 2007), ao modelo proposto por da-Silva et al. (2011). Esta atividade foi acompanhada de um árduo trabalho de implementação computacional. Foi nesse momento que desenvolvemos a base de todas as rotinas R apresentadas nesta dissertação.

No momento em que conseguimos finalizar essa etapa, tendo observado bons resultados, nos propusemos a tentar estender a metodologia ao caso multivariado. Mais especificamente, ao caso em que as observações têm distribuição Dirichlet, uma generalização da distribuição Beta. Sabíamos, contudo, que por se tratar de um modelo inédito, havia a possibilidade clara de não obtermos êxito nessa empreitada. De fato, percorremos caminhos infrutíferos antes de propor um modelo coerente, sob a ótica probabilística, e viável, no que se refere à possibilidade de estimação de seus parâmetros.

Deve-se ressaltar que se por um lado a distribuição Dirichlet é muito indicada ao papel de distribuição das observações, sua utilização como distribuição a priori $(\mu_t|D_{t-1})$ inviabiliza o processo de filtragem. As duas afirmações são consequência da estrutura de covariância rígida dessa distribuição, expressa como função somente do vetor de médias e do parâmetro (escalar) ϕ.

A fim de tornar possível, especificamente, a elicitação da priori $(\mu_t|D_{t-1})$, respeitando-se a relação $\lambda_t = \text{alz}(\mu_t)$ e os momentos a priori de λ_t, f_t e Q_t, fomos obrigados a buscar outras distribuições, mais flexíveis, para dados composicionais, alternativas

à Dirichlet. Foi nesse contexto que adotamos como priori a distribuição Logística-Normal e, finalmente, conseguimos viabilizar a conclusão de todo ciclo inferencial recursivo para a estimação dos parâmetros do sistema. Cabe destacar que a adoção da distribuição Logística-Normal como distribuição das observações nos obriga a lidar com a especificação (ou amostragem) de uma matriz Φ, em substituição ao parâmetro escalar ϕ. Do ponto de vista prático e operacional, tal alteração aumenta significativamente a dificuldade da estimação e utilização do modelo.

Considerando o cenário acima descrito, o fato de termos alcançado o desafio lançado, propor o modelo dinâmico Dirichlet (MDD) baseado nas idéias que amparam os modelos dinâmicos lineares generalizados, nos deixa muito satisfeitos. O capítulo 4 evidencia a grande versatilidade do MDD apresentado. Recorde que com ele, é possível descrever dados univariados de proporções e dados composicionais multivariados, quando há mais de uma categoria de resposta possível. Além disso, o MDD pode ser usado tanto no contexto estático, funcionando como um modelo de regressão Beta ou Dirichlet, quanto no contexto dinâmico de séries temporais.

Diversas características do fenômeno em estudo podem ser retratadas de maneira extremamente simples e imediata. Em especial, citamos efeito de covariáveis, efeito de sazonalidade, comportamento auto-regressivo e tendência de crescimento do nível da série observável.

Os modelos dinâmicos Dirichlet são aplicáveis a incontáveis situações práticas e podem ser usados no processo de previsão, na descrição da trajetória dos parâmetros do sistema, no diagnóstico da relação entre variáveis, na decomposição de um resultado observado em termos dos fatores que o determinam, na identificação de unidades amostrais atípicas, etc.

No que se refere ao procedimento de estimação, foram propostas duas abordagens, *online* e *offline*. A primeira, de custo computacional despresível, foi desenhada para o processamento instantâneo de novos dados. Por outro lado, a segunda abordagem, mais poderosa e abrangente, exige um esforço computacional significativo, principalmente no caso dinâmico multivariado. Felizmente, as duas proposta apresentaram resultados muito satisfatórios e proporcionam ao analista o poder de escolher, em função de suas necessidades, o método de estimação que melhor o atende.

Naturalmente, não é possível esgotar as propriedades e características de um modelo desse porte em uma dissertação de mestrado. Há diversas atividades que podem ser desenvolvidas em trabalhos futuros. Entre as principais, citamos as seguintes:

• Como mencionado na Subseção 3.2.1.4, tivemos certa dificuldade em definir parâmetros que tornem vaga a distribuição a priori Wishart, adotada para a matriz de covariância dos erros de evolução do sistema W. É possível que a má escolha da informação inicial ocasione certa perda de qualidade das estimativas. Diante disso, a exploração mais profunda dos aspectos relacionados à adoção da distribuição a priori Wishart pode ser de grande valia. Sabe-se que há, na literatura, formas alternativas de definir prioris para matrizes dessa natureza. Talvez alguma delas seja mais adequada no contexto do MDD.

• O método *CUBS* amostra, em um único movimento, todos os parâmetros desconhecidos $\Theta = (\theta_1, \ldots, \theta_T)$. Em certos casos, tal estratégia resulta em taxas de aceitação proibitivamente baixas, comprometendo, portanto, a eficiência do método. Desse modo, imaginamos que outras propostas de funções geradoras de candidatos podem trazer ganhos de eficiência em alguns cenários.

• Pela grande flexibilidade dos modelos dinâmicos Dirichlet, alguns casos particulares importantes não foram devidamente explorados. É preciso detalhar outros procedimentos de especificação das matrizes F_t e G_t. Sabe-se, por exemplo, que é simples, pelas propriedades dos modelos dinâmicos, acomodar componentes autoregressivos ao modelo, ainda que não tenhamos nos referenciado diretamente a esse caso nesta dissertação.

• O processo computacional envolvido na aplicação dos modelos dinâmicos Dirichlet não é trivial. Mesmo tendo desenvolvido funções, de certo modo, robustas para a implementação da metodologia proposta, ainda há, evidentemente, pontos passíveis de melhoramento na programação. O aprimoramento das rotinas R desenvolvidas nesse projeto poderia tanto tornar as função menos suscetíveis a dificuldades numéricas quanto aumentar a velocidade de processamento dos dados. Executar parte do processo em C, em especial as integrais no hipercubo envolvidas no processo de filtragem, poderia representar ganhos significativos de eficiência.

• Embora tenhamos observado, com o auxílio de técnicas descritivas, bons ajustes no capítulo 4, não fomos tão rigorosos na fase de diagnóstico do modelo. Dessa

maneira, ainda está pendente a atividade de detalhamento de técnicas formais de diagnostico. Nesse mesmo sentido, métodos de seleção de modelos também precisam ser investigados.

• Uma forte restrição do modelo proposto, é ser este aplicável somente a cenários nos quais o parâmetro de precisão é constante no tempo. A adição de dinâmica a esse parâmetro ou descrevê-lo em termos de covariáveis, permitiria ao modelo dinâmico Dirichlet um ganho adicional de flexibilidade.

• Por fim, é preciso simular o procedimento proposto de estimação milhares de vezes, e em diversos cenários diferentes, a fim de investigar suas propriedades estatísticas.

Referências Bibliográficas

Aitchison, J. & Shen, S. M. (1980). Logistic-Normal Distributions: Some Properties and Uses. *Biometrika*, 67(2):261–272.

Carter, C. & Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, 81:541–553.

Cribari-Neto, F. & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, 34(2):1–24.

da-Silva, C. Q., Migon, H. S., & Correia, L. T. (2011). Dynamic bayesian beta models. *Computational Statistics and Data Analysis*, 55(6):2074–2089.

Ferrari, S. L. P. & Cribari-Neto, F. (2004). Beta regression for modeling rates and proportions. *Journal of Applied Statistics*, 31:799–815.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15:183–202.

Gamerman, D. (1998). Markov chain monte carlo for dynamic generalized linear models. *Biometrika*, 85:215–227.

Gamerman, D. & Lopes, H. F. (2006). *Markov Chain Monte Carlo*, (2nd edition ed.). Chapman and Hall/CRC.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2011). *mvtnorm: Multivariate Normal and t Distributions*. R package version 0.9-96.

Geweke, J. & Tanizaki, H. (2001). Bayesian estimation of state space models using metropolis-hastings algorithm within gibbs sampling. *Computation Statistics and Data Analysis*, 37:151–170.

Griffiths, W. E., Hill, R. C., & Judge, G. G. (1993). *Learning and Practicing Econometrics*. John Wiley and Sons, New York.

Grunwald, G., Raftery, A., & Guttorp, P. (1993). Time series for continuous proportions. *Journal of the Royal Statistical Society*, B, 55:103–116.

Hankin, R. K. S. (2010). A generalization of the dirichlet distribution. *Journal of Statistical Software*, 33(11):1–18.

Harrison, P. J. & Scott, F. A. (1965). A development system for use in short-term forecasting (original draft 1965, re-issued 1982). Research Report 26, Department of Statistics, University of Warwick.

Johnson, S. G. & Narasimhan, B. (2009). *cubature: Adaptive multivariate integration over hypercubes*. R package version 1.0.

Lange, K. (1999). *Numerical Analysis for Statisticians*. Springer.

Lindley, D. V. & Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society*, B, 34:1–41.

Martin, A. D., Quinn, K. M., & Park, J. H. (2011). *MCMCpack: Markov chain Monte Carlo (MCMC) Package*. R package version 1.0-9.

McCullagh, P. & Nelder, J. A. (1994). *Generalized Linear Models*, (2nd edition ed.). The University Press, Cambridge: Monographs on Statistics and Applied Probability 37.

Migon, H. & Gamerman, D. (1993). Dynamic Hierachical Models. *Journal of the Royal Statistical Society*, B, 55:629–642.

Nelder, J. A. & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, A, 135:370–384.

Petris, G. (2010). An R package for dynamic linear models. *Journal of Statistical Software*, 36(12):1–16.

Petris, G., Petrone, S., & Campagnoli, P. (2009). *Dynamic Linear Models with R*. Springer.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2010). *coda: Output analysis and diagnostics for MCMC*. R package version 0.14-2.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Ravines, R., Migon, H., & Schmidt, A. (2007). An efficient sampling scheme for dynamic generalized models. Technical Report 201/2007, Departamento de Métodos Estatísticos - IM - UFRJ.

Roberts, G. O., Gelman, A., & Gilks, W. R. (1994). Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. Technical report, University of Cambridge.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate bayesian iference for latent gaussian models by using integrated nested laplace appoximations. *Journal of the Royal Statistical Society*, B, 71(2):319–392–346.

Shephard, N. & Pitt, M. (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika*, 84:653–667.

Smith, J. Q. (1979). A Generalization of the Bayesian Steady Forecasting Model. *Journal of the Royal Statistical Society*, B, 41:375–387.

Tierney, L. & Kadane, J. B. (1986). Accurate approximation for the posterior moments and marginal densities. *Journal of the American Statistical Association*, 81:82–6.

West, M. & Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Models*, (2nd edition ed.). Springer.

West, M., Harrison, P. J., & Migon, H. S. (1985). Dynamic generalized linear models and bayesian forecasting (with discussion). *Journal of the American Statistical Association*, 80:73–97.