

Universidade Federal do Amazonas - UFAM
Instituto de Ciências Exatas - ICE
Programa de Pós-Graduação em Matemática - PPGM

**UM MODELO LINEAR COM AMORTECEDOR DE
TENDÊNCIA DINÂMICO PARA PREVISÃO
BAYESIANA DE SÉRIES TEMPORAIS**

Autor: Guilherme Santos Silva

Manaus
2014

UM MODELO LINEAR COM AMORTECEDOR DE TENDÊNCIA DINÂMICO
PARA PREVISÃO BAYESIANA DE SÉRIES TEMPORAIS

Guilherme Santos Silva

Dissertação de Mestrado apresentada ao
Programa de Pós-graduação em Matemática,
da Universidade Federal do Amazonas, como
parte dos requisitos necessários à obtenção do
título de Mestre em Matemática

Orientador: Prof. Dr. James Dean Oliveira dos
Santos Júnior

Manaus

Abril de 2014

Ficha Catalográfica
(Catalogação realizada pela Biblioteca Central da UFAM)

Silva, Guilherme Santos

S586m Um modelo linear com amortecedor de tendência dinâmico para
previsão Bayesiana de séries temporais / Guilherme Santos Silva -
Manaus, 2014.

64f. il. color.

Dissertação (mestrado em Matemática) – Universidade Federal do
Amazonas.

Orientador: Prof. Dr. James Dean dos Santos Júnior

1. Modelos lineares 2. Sistemas lineares 3. Parâmetro de
amortecimento I. Santos Júnior, James dean (Orient.) II.
Universidade Federal do Amazonas III. Título


CDU 2007 624.15(043.3)

UM MODELO LINEAR COM ABAFADOR DE TENDÊNCIA DINÂMICO PARA
PREVISÃO BAYESIANA DE SÉRIES TEMPORAIS

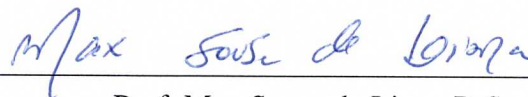
Guilherme Santos Silva

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PROGRAMA DE
PÓS-GRADUAÇÃO EM MATEMÁTICA, DA UNIVERSIDADE FEDERAL DO
AMAZONAS, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A
OBTENÇÃO DO GRAU DE MESTRE EM MATEMÁTICA.

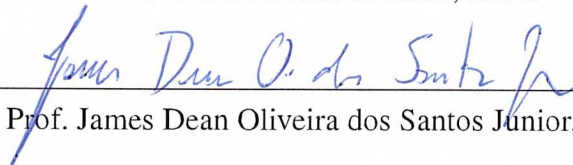
Examinada por:



Prof. Cibele Queiroz da Silva, Ph.D.



Prof. Max Sousa de Lima, D.Sc.



Prof. James Dean Oliveira dos Santos Junior, D.Sc.

MANAUS, AM – BRASIL

ABRIL DE 2014

*À Deus,
À banda de rock U2,
A todos aqueles que defendem os
fracos e oprimidos, em todo o
mundo.*

Agradecimentos

Gostaria de agradecer a CAPES pelo apoio financeiro. Ao professor James Dean de Oliveira dos Santos Jr. pelo trabalho de orientação. Aos professores do departamento de estatística da UFAM, Max Souza de Lima e José Raimundo Gomes pelo apoio prestado. A minha mãe, Tânia Eloisa Silva, por toda a dedicação e paciência ao longo dos anos. A Ana Ruth Ferreira Monteiro pelo amor e apoio ao longo dos anos que estivemos juntos. A Bill Gates, Dilma Rousseff, Sarah Palin, George W. Bush e os Papas Benedicto XVI e Francisco pelo apoio em momentos difíceis. Ao Professor Rob J. Hyndman pelas contribuições dos pacotes R forecast e Mcomp, usados neste trabalho. A toda comunidade R pelo trabalho e dedicação em oferecer software livre de qualidade. A todos aqueles que me ajudaram e me apoiaram ao longo dos últimos quatro anos, cujos nomes nunca tive chance de conhecer.

Resumo da Dissertação apresentada ao Programa de Pós-Graduação em Matemática, da Universidade Federal do Amazonas, como parte dos requisitos necessários para a obtenção do grau de Mestre em Matemática. (M.Sc.)

UM MODELO LINEAR COM AMORTECEDOR DE TENDÊNCIA DINÂMICO
PARA PREVISÃO BAYESIANA DE SÉRIES TEMPORAIS

Guilherme Santos Silva

Abril/2014

Orientador: Prof. Dr. James Dean Oliveira dos Santos Júnior

Área de Concentração : Estatística

Investigamos a performance de um novo modelo linear dinâmico onde o parâmetro de amortecimento da tendência possui uma evolução temporal dada por uma distribuição de probabilidade normal. O objetivo é determinar se a inclusão de uma dinâmica na evolução do parâmetro de amortecimento melhora a performance preditiva do modelo em relação aos modelos polinomiais existentes, em particular o modelo de tendência aditiva *Damped Holt*. Para avaliar esta nova proposta, foi desenvolvido um estudo de simulação e de aplicações em dados da competição internacional M3, analisando a performance do modelo. Foram simuladas séries polinomiais de ordem dois com diferentes valores para a variância observacional e a variância de evolução dos estados. Os resultados do estudo sugerem que o novo modelo proposto consegue obter um ganho preditivo marginal em relação aos modelos polinomiais existentes em boa parte do espaço paramétrico. Com os dados da competição internacional M3, várias séries com diferentes características foram analisadas. A função preditiva k passos a frente foi avaliada após um período de ajuste do modelo aos dados. Para a estimação dos parâmetros dos modelos já existentes, foi empregada a técnica de multiprocesso classe I e para a estimação dos parâmetros do novo modelo foi empregada a minimização da medida de erro SMAPE. O novo modelo tem toda a evolução dos estados em forma analítica e nenhum tipo de simulação é necessária para a estimação dos parâmetros. A limitação para o novo modelo surge quando o

parâmetro de estado referente a inclinação se aproxima de zero. Neste caso o modelo é considerado inapropriado e novos estudos são necessários para contornar este problema.

Palavras-chave: Séries Temporais, Modelos Lineares Dinâmicos, *Damped Holt* e Normal.

Abstract of Dissertation presented to Postgraduate in Mathematics, of the Federal University of Amazonas, as a partial fulfillment of the requirements for the degree of Master of Mathematics. (M.Sc.)

A DYNAMIC DAMPED HOLT MODEL FOR BAYESIAN TIME SERIES FORECASTING

Guilherme Santos Silva

April/2014

Advisor: Prof. Dr. James Dean Oliveira dos Santos Júnior

Research lines: Statistics

A novel dynamic linear model is proposed and its forecasting performance was investigated. This model allows the trend dampening parameter to evolve with time according to a normal distribution. The main goal is verify whether the inclusion of a temporal dynamic in the trend dampening parameter improves the model's performance when compared to the existing polynomial models, in particular the Damped Holt model. To assess this new model, a simulation study was carried out and the model's forecasts were compared with other models using the M3 competition monthly data. Time series of second order polynomial models were simulated with different values for the observational variance and the state evolution variance. The results of both studies suggest that the proposed model achieves a marginal predictive gain when compared with the existing polynomial models in part of the parametric space. Several series from the M3 competition data with different characteristics were analysed. The k-step ahead predictive function was evaluated after a training period. Multi-process class I and mean squared error minimization were employed for the parameter estimation of the existing models. For the novel model, minimization of the Symmetric Mean Absolute Percentage Error (SMAPE) metric was used. The proposed model has the state evolution in analytical form and there is no need for any kind of simulation for parameter estimation. A limitation of the proposed model is its inability in dealing with the second component of the state vector when its value is

near zero. In this case, the model is deemed inappropriate and new studies are necessary to solve this problem. **keyword:** Time Series, Dynamic Linear Models, *Damped Holt* and Normal Distribution.

Sumário

| | |
|---|----------|
| Lista de Figuras | h |
| Lista de Tabelas | i |
| 1 Introdução | 1 |
| 1.1 A Análise Estatística de Séries Temporais | 1 |
| 1.2 Inferência Bayesiana | 2 |
| 1.3 Proposta do Trabalho | 5 |
| 1.4 Organização do Trabalho | 6 |
| 2 Modelos Lineares Dinâmicos | 7 |
| 2.1 O Modelo Linear Dinâmico | 7 |
| 2.1.1 Modelos de Espaço de Estados | 8 |
| 2.2 Modelos Lineares Dinâmicos | 10 |
| 2.2.1 Modelo Polinomial de Ordem 1 | 12 |
| 2.2.2 Modelo Polinomial de Ordem 2 | 13 |
| 2.2.3 Modelos Sazonais | 14 |
| 2.2.4 Estimação dos Estados e Previsão | 15 |

| | |
|---|-----------|
| <i>SUMÁRIO</i> | h |
| 2.2.5 Distribuições de Previsão | 17 |
| 2.2.6 Distribuições de <i>Suavização</i> | 18 |
| 2.2.7 Variâncias Observacionais | 19 |
| 2.2.8 Variâncias Observacionais Dinâmicas | 21 |
| 2.2.9 Fatores de Desconto para a Especificação da Matriz de Evolução . | 22 |
| 3 Modelos <i>Damped Holt</i> e Estimação Bayesiana via Multiprocessos | 24 |
| 3.1 O Modelo Aditivo <i>Damped Holt</i> | 24 |
| 3.2 O Modelo Multiplicativo <i>Damped Pegels</i> | 26 |
| 3.3 Estimação Bayesiana via Multiprocessos | 27 |
| 3.4 Métricas de Performance Preditiva | 30 |
| 4 O Modelo <i>Damped Holt</i> com Amortecedor Dinâmico | 33 |
| 4.1 Motivação | 33 |
| 4.2 O Modelo <i>Damped Holt</i> Dinâmico | 34 |
| 4.3 A Evolução do Amortecedor | 45 |
| 4.4 Estimação de Parâmetros | 49 |
| 4.5 Limitações do Modelo | 50 |
| 4.6 Estudo de uma Série | 53 |
| 5 Análise dos Resultados do novo Modelo | 54 |
| 5.1 Simulação de Séries Temporais | 54 |
| 5.1.1 Resultados para as Séries com $V_t = 100, W_{11,t} = 1, W_{22,t} = 0.1$. . | 55 |
| 5.1.2 Resultados para as Séries com $V_t = 100, W_{11,t} = 10, W_{22,t} = 1$. . | 58 |

| | | |
|----------|--|-----------|
| 5.1.3 | Resultados para as Séries com $V_t = 10, W_{11,t} = 10, W_{22,t} = 1$ | 60 |
| 5.1.4 | Resultados para as Séries com $V_t = 1, W_{11,t} = 10, W_{22,t} = 1$ | 62 |
| 5.1.5 | Resultados para as Séries com $V_t = 1, W_{11,t} = 100, W_{22,t} = 10$. . . | 63 |
| 5.2 | Análise da Competição Internacional M3 | 66 |
| 5.2.1 | Resultados para todas as Séries | 69 |
| 5.2.2 | Resultados para todas as Séries usando Multiprocessos | 72 |
| 5.2.3 | Resultados para todas as Séries onde a Tendência é forte | 73 |
| 5.2.4 | Resultados para todas as Séries onde AIC seleciona Modelos Or- dem Dois | 75 |
| 6 | Conclusão e Trabalhos Futuros | 78 |
| 6.1 | Conclusão | 78 |
| 6.2 | Trabalhos Futuros | 80 |
| A | Lemas empregados | 82 |
| A.1 | Teoria de Distribuição Normal | 82 |
| | Referências Bibliográficas | 84 |

Lista de Figuras

| | | |
|-----|---|----|
| 4.1 | Exemplo de série da competição M3 (N2523) onde a tendência possui dinâmica temporal. | 34 |
| 4.2 | Série da competição M3 (N2533) ilustrando o problema da dinâmica do amortecedor. | 45 |
| 4.3 | Série da competição M3 (N2525) ilustrando a influência de diferentes valores de u_t | 48 |
| 4.4 | Valores da média de ψ_t para a série da competição M3 (N2525) ilustrando a influência de u_t | 48 |
| 4.5 | Valores da média de $\theta_{2,t}$ para a série da competição M3 (N2525) ilustrando a influência de u_t | 49 |
| 4.6 | Série da competição M3 (N2527) ilustrando o problema da dinâmica do amortecedor com valores próximos de zero. | 52 |
| 4.7 | Série da competição M3 (N2528) ilustrando a vantagem da dinâmica do amortecedor para previsões. | 53 |
| 5.1 | SMAPE das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 1, W_{22,t} = 0.1$ | 56 |
| 5.2 | MedAPE das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 1, W_{22,t} = 0.1$ | 57 |
| 5.3 | MASE das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 1, W_{22,t} = 0.1$ | 57 |

| | | |
|------|---|----|
| 5.4 | <i>Percent Better</i> das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 1, W_{22,t} = 0.1$ | 58 |
| 5.5 | SMAPE das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 10, W_{22,t} = 1$ | 58 |
| 5.6 | MedAPE das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 10, W_{22,t} = 1$ | 59 |
| 5.7 | MASE das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 10, W_{22,t} = 1$ | 59 |
| 5.8 | <i>Percent Better</i> das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 10, W_{22,t} = 1$ | 60 |
| 5.9 | SMAPE das séries simuladas com parâmetros $V_t = 10, W_{11,t} = 10, W_{22,t} = 1$ | 60 |
| 5.10 | MedAPE das séries simuladas com parâmetros $V_t = 10, W_{11,t} = 10, W_{22,t} = 1$ | 61 |
| 5.11 | MASE das séries simuladas com parâmetros $V_t = 10, W_{11,t} = 10, W_{22,t} = 1$ | 61 |
| 5.12 | SMAPE das séries simuladas com parâmetros $V_t = 1, W_{11,t} = 10, W_{22,t} = 1$ | 62 |
| 5.13 | MedAPE das séries simuladas com parâmetros $V_t = 1, W_{11,t} = 10, W_{22,t} = 1$ | 62 |
| 5.14 | MASE das séries simuladas com parâmetros $V_t = 1, W_{11,t} = 10, W_{22,t} = 1$ | 63 |
| 5.15 | SMAPE das séries simuladas com parâmetros $V_t = 1, W_{11,t} = 100, W_{22,t} = 10$ | 63 |
| 5.16 | MedAPE das séries simuladas com parâmetros $V_t = 1, W_{11,t} = 100, W_{22,t} = 10$ | 64 |
| 5.17 | MASE das séries simuladas com parâmetros $V_t = 1, W_{11,t} = 100, W_{22,t} = 10$ | 64 |

Lista de Tabelas

| | | |
|-----|---|----|
| 4.1 | Resumo das Equações do <i>Damped Holt</i> Dinâmico | 44 |
| 5.1 | SMAPE, MedAPE e MASE de horizontes 1-6 para as 1428 séries mensais da competição M3. | 70 |
| 5.2 | SMAPE, MedAPE e MASE de horizontes curto, médio e longo prazo para as 1428 séries mensais da competição M3. | 71 |
| 5.3 | SMAPE, MedAPE e MASE de horizontes 1-6 para as 1428 séries mensais da competição M3, com multiprocessos classe I. | 73 |
| 5.4 | SMAPE, MedAPE e MASE de horizontes curto, médio e longo prazo para as 1428 séries mensais da competição M3, com multiprocessos classe I. | 73 |
| 5.5 | SMAPE, MedAPE e MASE de horizontes 1-6 para as 153 séries mensais da competição M3 onde $\psi > 1$ | 74 |
| 5.6 | SMAPE, MedAPE e MASE de horizontes curto, médio e longo prazo para as 153 séries mensais da competição M3 onde $\psi > 1$ | 75 |
| 5.7 | SMAPE, MedAPE e MASE de horizontes 1-6 para as 644 séries mensais da competição M3 onde o AICc seleciona o modelo de ordem dois. | 76 |
| 5.8 | SMAPE, MedAPE e MASE de horizontes curto, médio e longo prazo para as 644 séries mensais da competição M3 onde o AICc seleciona o modelo de ordem dois. | 77 |

Capítulo 1

Introdução

Neste capítulo apresenta-se o enfoque de previsão bayesiana de séries temporais, seus objetivos e exemplos de aplicações. Também é descrita a proposta do trabalho e sua organização.

1.1 A Análise Estatística de Séries Temporais

As aplicações de análise de séries temporais ganharam um crescimento enorme nas últimas décadas e podem ser usadas em diversos campos de estudo, tais como finanças, engenharia, demografia, economia, medicina, etc. Como exemplos de alguns problemas que são resolvidos por esta técnica temos:

- Previsão de vendas de produtos;
- Análise e previsão de efeitos promocionais em vendas e *marketing*;
- Detecção de anormalidades em monitoramento de pacientes pós-operatório;
- Previsão de preços de ativos financeiros para gerenciamento de *portfólio*;
- Previsão de volatilidade para precificação de instrumentos financeiros;
- Previsão de demanda de partes em gerenciamento de inventário;

Devido ao recente desenvolvimento computacional, muitos destes procedimentos já são feitos em processos automatizados e *on-line*, muitas vezes em séries de alta frequência (Hang [16]).

O problema abordado em análise de séries temporais é oferecer uma *previsão* do futuro desenvolvimento de uma série com *precisão* aceitável. A previsão de dados é um componente essencial na estratégia de muitas empresas. Existem duas principais abordagens para análise estatística de séries temporais: a frequentista (por exemplo, a metodologia ARIMA de Box-Jenkins (Carmona [7])), e a bayesiana (por exemplo, os modelos dinâmicos de Harrison-Stevens (West & Harrison [39])). Neste trabalho adotou-se a abordagem bayesiana.

A previsão bayesiana de séries temporais é uma técnica que permite resolver os problemas acima de forma intuitiva e simples. Ela oferece uma formulação muito flexível para analisar fenômenos dinâmicos e sistemas que evoluem no tempo. Além disso, ela contribuiu para estender os domínios de aplicação para processos não estacionários e dados discretos e multivariados. Os modelos bayesianos mais comuns são os modelos de Espaço de Estados. Eles usam um vetor ou matriz de estados para definir o modelo. Caso os estados e as observações tenham distribuição normal e o relacionamento entre as observações e os estados seja linear, tem-se o caso particular, chamado de Modelos Lineares Dinâmicos (Durbin & Koopman [10], West & Harrison [39]).

1.2 Inferência Bayesiana

Este trabalho emprega a modelagem bayesiana na previsão de séries temporais usando Modelos Lineares Dinâmicos (DLM). Estes modelos foram desenvolvidos pelo campo da engenharia na década de sessenta e receberam um enorme impulso devido ao desenvolvimento de recursos computacionais na década de noventa (Robert & Casella [31], Venables & Ripley [37], Tanner [33], Wasserman [38]).

Em um sistema estocástico, a incerteza é devida principalmente à variáveis que não foram consideradas no modelo e erros de medição. Esta incerteza é descrita através de uma distribuição de probabilidade. Por consequência, a estimação dos estados do sistema e outras quantidades de interesse são resolvidas calculando a distribuição condicional, dadas as informações disponíveis. Este é um conceito básico em inferência bayesiana.

DLMs são baseados na idéia de descrever a saída (denominado \mathbf{Y}_t no tempo t) de um sistema dinâmico como uma função de estados não observáveis que possuem uma evolução *Markoviana*. Este sistema é naturalmente afetado por erros aleatórios. Ao condicionar a variáveis latentes à dependência temporal dos dados consegue-se uma dinâmica que se encaixa de maneira natural no enfoque bayesiano. Outra vantagem de DLMs é que os cálculos podem ser feitos de maneira recursiva. As distribuições condicionais de interesse

podem ser atualizadas, incorporando os novos dados, sem que seja necessário guardar todo o histórico de informações. Isto é muito vantajoso quando os dados tem natureza sequencial e a inferência *on-line* é necessária. A natureza recursiva dos cálculos é uma consequência da fórmula de Bayes na formulação dos DLMS.

Na estatística bayesiana toda a incerteza que se tem sobre os parâmetros de um processo ou fenômeno é descrita por meio de uma distribuição de probabilidade. Esta distribuição tem uma interpretação subjetiva, sendo uma maneira de formalizar a informação incompleta que o pesquisador tem sobre o processo que está sendo analisado. O processo de aprendizado consiste da aplicação de regras de probabilidade, onde basta calcular a probabilidade condicional do evento de interesse, dado a informação disponível. O Teorema de Bayes é a regra básica a ser usada. Este teorema explica como atualizar crenças depois que nova evidência sobre o processo é considerada. A distribuição *posteriori* do parâmetro de interesse é proporcional a verossimilhança observada vezes a probabilidade *a priori* do parâmetro. Em geral, isto é representado como:

$$\text{Posteriori} \propto \text{Verossimilhança observada} \times \text{Priori} . \quad (1.1)$$

O evento de interesse e o resultado do sistema dependem do problema.

Antes de seguir adiante, vamos introduzir a notação e as convenções usadas. Variáveis aleatórias ou vetores aleatórios serão denotados por letras maiúsculas e um valor realizado será denotado por letras minúsculas. Não existe distinção de notação entre vetores e quantidades multivariadas e a distinção será clara dentro do contexto. Uma série temporal univariada ou multivariada é uma sequência de variáveis ou vetores aleatórios e será denotada por $(\mathbf{Y}_t : t = 1, 2, \dots)$, $(\mathbf{Y}_t)_{t \geq 1}$ ou simplesmente (\mathbf{Y}_t) . Quando se considera uma sequência finita de observações consecutivas entre o i -ésimo e j -ésimo valor, usa-se a notação $\mathbf{Y}_{i:j}$. Densidades de probabilidade serão denotadas por $f(\cdot)$. Adota-se a convenção de usar o mesmo simbolo f para a distribuição de diferentes variáveis aleatórias, onde o argumento torna claro a qual distribuição se refere. Para um vetor aleatório p -dimensional $\boldsymbol{\theta}$, com média \mathbf{m} e matrix de variância \mathbf{C} , adota-se a sua distribuição multivariada como $\boldsymbol{\theta} \sim N_p[\mathbf{m}, \mathbf{C}]$. Quantidades multivariadas serão denotadas em negrito.

Em inferência estatística, o experimento é geralmente o resultado de uma amostragem, e este é descrito por um vetor aleatório \mathbf{Y} . É comum usar um modelo paramétrico para atribuir uma densidade de probabilidade de \mathbf{Y} . A quantidade de interesse é o vetor $\boldsymbol{\theta}$ dos parâmetros do modelo. A inferência bayesiana de $\boldsymbol{\theta}$ consiste em calcular a sua distribuição condicional dado os resultados da amostragem. Baseado no conhecimento do processo sendo analisado, o pesquisador pode atribuir uma distribuição condicional $f(\mathbf{Y}|\boldsymbol{\theta})$ para \mathbf{Y} dado $\boldsymbol{\theta}$. Esta é a função verossimilhança, como definido acima. É também atribuída

uma distribuição *priori* $f(\boldsymbol{\theta})$ que expressa a incerteza sobre o vetor de parâmetros $\boldsymbol{\theta}$. Ao observar $\mathbf{Y} = \mathbf{y}$, pode-se usar uma generalização do Teorema de Bayes para calcular a densidade condicional de $\boldsymbol{\theta}$ dado \mathbf{y} :

$$f(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{\int f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}}. \quad (1.2)$$

Na maioria das vezes, o principal objetivo da análise estatística de séries temporais é a previsão. Logo tem-se interesse pela distribuição preditiva de \mathbf{Y} , ou seja, $f(\mathbf{y}_{t+k}|\mathbf{y}_{1:t})$, onde $k \geq 1$. Em modelos paramétricos isto pode ser calculado como

$$f(\mathbf{y}_{t+k}|\mathbf{y}_{1:t}) = \int f(\mathbf{y}_{t+k}, \boldsymbol{\theta}|\mathbf{y}_{1:t})d\boldsymbol{\theta} = \int f(\mathbf{y}_{t+k}|\boldsymbol{\theta}, \mathbf{y}_{1:t})f(\boldsymbol{\theta}|\mathbf{y}_{1:t})d\boldsymbol{\theta}. \quad (1.3)$$

Geralmente é mais simples especificar a densidade condicional $f(\mathbf{y}_{1:n}|\boldsymbol{\theta})$ de $\mathbf{Y}_{1:n}$ dado $\boldsymbol{\theta}$ e a densidade $f(\boldsymbol{\theta})$ do que obter diretamente $f(\mathbf{y}_{1:n})$ usando $f(\mathbf{y}_{1:n}) = \int f(\mathbf{y}_{1:n}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}$.

A estrutura de dependência mais simples é a independência condicional (Berger [4]). Nas aplicações que serão tratadas, é possível assumir que Y_1, \dots, Y_n são condicionalmente independentes e identicamente distribuídas (i.i.d.) dado $\boldsymbol{\theta}$, ou seja, $f(\mathbf{y}_{1:n}|\boldsymbol{\theta}) = \prod_{i=1}^n f(\mathbf{y}_i|\boldsymbol{\theta})$. Neste caso, a densidade preditiva é dada por:

$$f(\mathbf{y}_{t+1}|\mathbf{y}_{1:t}) = \int f(\mathbf{y}_{t+1}|\boldsymbol{\theta})f(\boldsymbol{\theta}|\mathbf{y}_{1:t})d\boldsymbol{\theta}. \quad (1.4)$$

A densidade *posteriori* pode ser calculada pela fórmula de Bayes:

$$f(\boldsymbol{\theta}|\mathbf{y}_{1:n}) = \frac{f(\mathbf{y}_{1:n}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\mathbf{y}_{1:n})} \propto \prod_{t=1}^n f(\mathbf{y}_t|\boldsymbol{\theta})f(\boldsymbol{\theta}). \quad (1.5)$$

Note que a distribuição a *posteriori* pode ser calculada recursivamente, de forma que a *posteriori* no tempo $t = n - 1$ se torna a *priori* no tempo $t = n$. Assim que a nova observação \mathbf{y}_n é feita, basta calcular a verossimilhança, que é $f(\mathbf{y}_n|\boldsymbol{\theta}, \mathbf{y}_{1:n-1}) = f(\mathbf{y}_n|\boldsymbol{\theta})$ assumindo a independência condicional, e atualizar a *priori* $f(\boldsymbol{\theta}|\mathbf{y}_{1:n-1})$ pela regra de Bayes, obtendo

$$f(\boldsymbol{\theta}|\mathbf{y}_{1:n-1}, \mathbf{y}_n) \propto f(\boldsymbol{\theta}|\mathbf{y}_{1:n-1})f(\mathbf{y}_n|\boldsymbol{\theta}) \propto \prod_{t=1}^{n-1} f(\mathbf{y}_t|\boldsymbol{\theta})f(\boldsymbol{\theta})f(\mathbf{y}_n|\boldsymbol{\theta}). \quad (1.6)$$

A estrutura recursiva da *posteriori* será fundamental ao introduzir os modelos dinâmicos no próximo capítulo.

Por conveniência, é comum usar *prioris* conjugadas, i.e., se a *priori* pertence à uma certa família de distribuições, então a *posteriori* também pertencerá à mesma família.

Para a família exponencial, tem-se uma noção de *priori* natural conjugada, que é definida a partir da densidade da estatística suficiente (Berger [4]). Na prática, usa-se *prioris* não informativas para expressar ignorância ou informação vaga sobre o parâmetro. Estender a noção de *priori* vaga ou uniforme quando o espaço paramétrico é infinito leva a distribuições impróprias. Estas distribuições não podem ser consideradas distribuições de probabilidade, pois a integração sobre o seu domínio retorna o valor infinito. Contudo, muitas vezes a densidade *posteriori* de *prioris* impróprias são próprias, o que justifica o seu uso.

1.3 Proposta do Trabalho

Modelos dinâmicos são uma alternativa eficiente para análise de séries temporais, pois permitem a inclusão de conhecimento específico do especialista de maneira natural no modelo (West & Harrison [39]). Dentro desta classe, vários tipos de modelos foram desenvolvidos ao longo dos anos, cada um lidando com determinadas características da série a ser ajustada. Neste contexto, foram desenvolvidos modelos polinomais, para lidar com diferentes taxas de crescimento e modelos sazonais para lidar com os ciclos que a série pode apresentar. Dentre os vários modelos existentes, os modelos com amortecedores de tendência se apresentam como um dos mais promissores para previsões (Taylor [35]). Estes modelos foram desenvolvidos numa classe mais ampla, denominada *exponential smoothing* (Hyndman *et al.* [19]).

O modelo aditivo com amortecimento de tendência, também conhecido como *Damped Holt*, se destacou em competições internacionais de previsão. Nestes modelos, o parâmetro de amortecimento de tendência é constante ao longo do desenvolvimento da série. Geralmente a estimação deste parâmetro ocorre por minimização dos erros quadrados ou maximização da função de verossimilhança (Hyndman & Khandakar [20]).

O objetivo geral deste trabalho é investigar se a inclusão de uma dinâmica temporal no parâmetro de amortecimento traz ganhos de performance preditiva aos modelos *Damped Holt*.

Os objetivos específicos deste trabalho são:

- Revisão de literatura sobre todos os conceitos envolvidos no desenvolvimento desse trabalho;
- Propor um novo Modelo Linear Dinâmico com parâmetro de amortecimento de

tendência dinâmico;

- Avaliar o novo modelo proposto em diversos cenários de simulação, com diferentes valores para os parâmetros;
- Aplicar a nova metodologia proposta em dados reais da competição internacional M3 e, em particular, quantificar a diferença de performance entre o novo modelo e os modelos existentes.

1.4 Organização do Trabalho

Esta dissertação está dividida em seis capítulos. Este Capítulo 1 com a Introdução sobre inferência bayesiana e motivação. No Capítulo 2 apresenta-se a definição de Modelos Lineares Dinâmicos e os métodos de estimação de parâmetros para esse modelo. Aborda-se ainda algumas particularidades desses modelos, como a incorporação de uma evolução temporal para a variância observacional e o uso de fatores de desconto ao invés da especificação da variância de evolução dos estados.

No Capítulo 3, apresenta-se os modelos aditivos *Damped Holt* e a estimação de parâmetros por misturas de densidades via *multiprocessos classe I*. Além disso, são apresentados as diferentes medidas de comparação de performance de modelos, assim como as suas principais características.

No Capítulo 4 é apresentado o novo modelo *Damped Holt* dinâmico, assim como as suas propriedades. Neste contexto, é discutida a estimação de parâmetros e algumas limitações do modelo, e também situações onde o modelo é adequado. A evolução do amortecedor é demonstrada e a sua dinâmica é discutida.

No Capítulo 5 são apresentados os resultados obtidos por meio dos estudos de simulações e aplicações com dados da competição internacional M3. Já houveram vários estudos para comparar a capacidade preditiva de vários métodos para análise de séries temporais. As competições M1, M2 e M3, conhecidas como as competições M, são as mais reconhecidas na literatura. Estes estudos comparam a performance de um grande número de métodos usando especialistas renomados, cujas previsões são analisadas em um conjunto amplo de séries com variadas características e de diferentes áreas de conhecimento. Os resultados da simulação são discutidos e a análise da performance do modelo na competição é apresentada.

As conclusões e sugestões são apresentadas no Capítulo 6.

No Apêndice A são apresentados os dois lemas usados na demonstração da evolução do amortecedor dinâmico de tendência e nos principais teoremas do Capítulo 2 e 4.

Capítulo 2

Modelos Lineares Dinâmicos

Neste capítulo são apresentados modelos lineares dinâmicos e suas principais propriedades. Discute-se a estimação dos seus parâmetros por meio de aprendizagem da variância observacional e a especificação da variância de evolução dos estados por meio de fatores de desconto. Os principais resultados apresentados na literatura são mostrados e discutidos no contexto deste trabalho. Os modelos lineares dinâmicos são apresentados como um caso especial dos modelos de Espaço de Estados, sendo lineares na relação com a variável resposta e com distribuições gaussianas.

2.1 O Modelo Linear Dinâmico

Devido ao acesso facilitado de poder computacional, tem havido um interesse crescente na aplicação de modelos de Espaço de Estados na análise de séries temporais; veja (Harvey [17], West & Harrison [39], Durbin & Koopman [10], Migon *et al.* [28], Petris *et al.* [29] e Pole *et al.* [30]) e todas as referências citadas nestas obras. Modelos de Espaço de Estados consideram uma série temporal como uma saída de um sistema dinâmico perturbado por efeitos aleatórios. Com esta formulação, é possível uma interpretação intuitiva de uma série temporal como sendo uma combinação de blocos estruturais mais simples, tais como tendência, sazonalidade ou componentes de regressão. Além disso, devido à natureza da formulação e ao Teorema de Bayes, todos os cálculos podem ser implementados de maneira recursiva, sem a necessidade de armazenar todo o histórico da série. A estimação de parâmetros e a previsão é feita de maneira recursiva através do cálculo da distribuição condicional dos estados e da variável resposta, dada a informação disponível.

Modelos de Espaço de Estados podem ser usados para modelar séries temporais univariadas ou multivariadas, além de séries não estacionárias, séries com mudanças estruturais e/ou séries com padrões irregulares. Toda análise de séries temporais se baseia na possibilidade de encontrar padrões que se reproduzem no desenvolvimento da série. Existe também a demanda por modelos que sejam flexíveis que não assumem padrões regulares ou possuem mudanças estruturais, tais como uma mudança repentina de nível. A possibilidade de analisar a série sem a necessidade de pré-transformar os dados, como é o caso dos modelos ARIMA, também pode ser citada como uma vantagem desta classe de modelos. Em várias situações, é muito desejável poder incorporar no modelo variáveis regressoras com o objetivo de auxiliar a previsão de mudanças bruscas nas séries. A modelagem de séries multivariadas também pode ser realizada de maneira simples e intuitiva, desde que cada série do conjunto atenda aos requisitos de normalidade e as séries sejam correlacionadas.

Nas próximas seções será discutido a formulação básica dos modelos de Espaço de Estados e seu caso particular, modelos lineares dinâmicos, assim como a estrutura recursiva do filtro de Kalman, usado para obter uma estimativa dos estados do sistema.

2.1.1 Modelos de Espaço de Estados

Seja uma série temporal $(\mathbf{Y}_t)_{t \geq 1}$. Neste caso não é possível assumir independência das observações, já que geralmente a autocorrelação entre elementos da série é alta pelo fato do tempo exercer influência sobre os dados. Dentre as formas de dependência, a *Markoviana* é a mais simples entre os \mathbf{Y}_t 's no qual o índice de tempo t tem influência. Diz-se que a série $(\boldsymbol{\theta}_t)_{t \geq 1}$ é uma cadeia de *Markov* se, para qualquer $t \geq 1$,

$$f(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{1:t-1}) = f(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}). \quad (2.1)$$

Isto significa que a informação sobre $\boldsymbol{\theta}_t$ vinda de todas as observações até o tempo $t - 1$ é exatamente a mesma informação vinda apenas de $\boldsymbol{\theta}_{t-1}$. Para uma cadeia de Markov, a distribuição conjunta pode ser escrita de maneira simples como:

$$f(\mathbf{y}_{1:t}) = f(\mathbf{y}_1) \prod_{j=2}^t f(\mathbf{y}_j | \mathbf{y}_{j-1}). \quad (2.2)$$

Considere no tempo $t = 0$ a informação inicial *a priori* como \mathbf{D}_0 . Em qualquer tempo

t futuro o conjunto de informações disponível é dado por:

$$\mathbf{D}_t = \{\mathbf{Y}_t, \mathbf{D}_{t-1}\}.$$

No entanto, assumir uma estrutura markoviana para as observações não é apropriado em muitos casos. Modelos de Espaço de Estados definem relações mais complexas para as observações com base na estrutura de dependência de uma cadeia de *Markov*. Em um modelo de Espaço de Estados tem-se uma cadeia de *Markov* não observável (θ_t), chamada de processo de estados. \mathbf{Y}_t é definido como uma medida imprecisa de θ_t , podendo ser uni- ou multivariado. Em análise de séries temporais, θ_t é uma variável latente que caracteriza o processo estocástico sendo analisado. Pode-se enxergar (θ_t) como uma série temporal que auxilia na especificação da distribuição de probabilidade da série observável (\mathbf{Y}_t). De maneira formal, um modelo de Espaço de Estados consiste de uma série temporal p -dimensional ($\theta_t : t = 0, 1, \dots$) e uma série temporal m -dimensional ($\mathbf{Y}_t : t = 1, 2, \dots$) onde as seguintes condições são satisfeitas:

1. A série temporal dos estados θ_t é uma cadeia de Markov;
2. O valor de \mathbf{Y}_t depende apenas de seu respectivo θ_t e, condicionado a θ_t , os \mathbf{Y}_t são independentes entre si.

Como consequência das condições listadas acima, o modelo de Espaço de Estados é completamente especificado pela:

1. Distribuição inicial a priori dos estados $f(\theta_0)$;
2. Densidade condicional da evolução dos estados $f(\theta_t | \theta_{t-1})$;
3. Densidade condicional $f(\mathbf{y}_t | \theta_t)$;

de forma que a distribuição conjunta das observações e dos estados, para qualquer $t \geq 1$ seja dada por

$$f(\theta_{0:t}, \mathbf{y}_{1:t}) = f(\theta_0) \prod_{j=1}^t f(\theta_j | \theta_{j-1}) f(\mathbf{y}_j | \theta_j). \quad (2.3)$$

De 2.3, pode-se obter qualquer outra distribuição de interesse, seja por condicionamento ou por marginalização. Note também que esta é a densidade usada para obter as condicionais completas para o amostrador de *Gibbs* (veja Robert & Casella [31]), se esta

for necessária para inferência sobre os estados. Modelos de Espaço de Estados nos quais os estados são variáveis aleatórias discretas são comumente chamados de Modelos de *Markov Ocultos* (*Hidden Markov Models*).

No caso de modelos de Espaço de Estados, a sua execução funciona de acordo com o princípio de "Gerenciamento por Exceção" (*Management by Exception*). Isto envolve o uso do modelo até que circunstâncias excepcionais aconteçam. Tais circunstâncias ocorrem de duas maneiras distintas. A primeira acontece quando informação de especialistas no domínio é recebida. Por exemplo, pode-se obter informações sobre nova legislação a ser adota, greves, dificuldades de fornecimento, etc. Esta informação é de caráter antecipatória (*feed-forward*) e é incluída no modelo através de mudanças nas densidades dos estados que sejam afetados. A segunda possui natureza corretiva (*feedback*) e ocorre quando o sistema de monitoramento usado para avaliar a performance preditiva da série sinaliza erros de previsão fora do usual. A reação pode ser a introdução de procedimentos de correção, automáticos ou manuais. Estes procedimentos podem ser semelhantes aos descritos acima ou podem ser projetados para corrigir as deficiências detectadas.

2.2 Modelos Lineares Dinâmicos

Modelos Lineares Dinâmicos são uma classe importante dos modelos de Espaço de Estados. Para um vetor m dimensional de observações \mathbf{Y}_t , onde as observações são recebidas sequencialmente ao longo do tempo, e um vetor p -dimensional de estados $\boldsymbol{\theta}_t$, com $t = 1, 2, \dots$, temos que:

Definição 2.1. *O Modelo Linear Normal é construído pela seguinte quádrupla:*

$$\{\mathbf{F}, \mathbf{G}, \mathbf{V}, \mathbf{W}\}_t = \{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$$

para cada tempo t , onde

1. \mathbf{F}_t é conhecida como a matriz de projeto, com dimensões $(p \times m)$;
2. \mathbf{G}_t é conhecida como a matrix de evolução, com dimensões $(p \times p)$;
3. \mathbf{V}_t é conhecida como a matrix de variância observacional, com dimensões $(m \times m)$;
4. \mathbf{W}_t é conhecida como a matrix de variância dos estados, com dimensões $(p \times p)$;

A quádrupla define o modelo que relaciona a variável resposta \mathbf{Y}_t com o vetor de estados $\boldsymbol{\theta}_t$ no tempo t através das seguintes distribuições de probabilidade:

$$(\mathbf{Y}_t | \boldsymbol{\theta}_t) \sim N[\mathbf{F}'_t \boldsymbol{\theta}_t, \mathbf{V}_t] \quad (2.4)$$

$$(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}) \sim N[\mathbf{G}_t \boldsymbol{\theta}_{t-1}, \mathbf{W}_t], \quad (2.5)$$

onde assume-se implicitamente que as distribuições são condicionais a \mathbf{D}_{t-1} , o conjunto de informações disponível antes do tempo t . Este conjunto tipicamente se compõe dos valores passados da série $\mathbf{Y}_{1:t-1}$ e os valores das variâncias \mathbf{V}_t e \mathbf{W}_t , além da informação inicial \mathbf{D}_0 . Também pode-se representar as probabilidades acima com as seguintes equações

$$\mathbf{Y}_t = \mathbf{F}'_t \boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \boldsymbol{\nu}_t \sim N[\mathbf{0}, \mathbf{V}_t] \quad (2.6)$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \boldsymbol{\omega}_t \sim N[\mathbf{0}, \mathbf{W}_t], \quad (2.7)$$

onde as sequências de erros $\boldsymbol{\nu}_t$ e $\boldsymbol{\omega}_t$ são mutuamente independentes entre si e dentro de cada série. Denomina-se $\boldsymbol{\nu}_t$ como o erro observacional e $\boldsymbol{\omega}_t$ como o erro de evolução ou erro de estados. A Equação 2.6 é denominada a equação de observação do modelo e define a distribuição de \mathbf{Y}_t condicional ao vetor de estados $\boldsymbol{\theta}_t$. A independência condicional é válida aqui e os \mathbf{Y}_t são independentes entre si, dado o vetor $\boldsymbol{\theta}_t$. Esta equação relaciona a variável resposta \mathbf{Y}_t ao vetor de estados $\boldsymbol{\theta}_t$ através de uma regressão linear com erros que tem uma distribuição normal multivariada se a variável resposta for multivariada também. A resposta média (*mean response, mean level*) é dada por $\boldsymbol{\mu}_t = \mathbf{F}'_t \boldsymbol{\theta}_t$.

A Equação 2.7 é denominada a equação de evolução ou equação dos estados e define a evolução do vetor de estados. A evolução é dada por uma cadeia de *Markov*, como descrito anteriormente, de forma que dado $\boldsymbol{\theta}_{t-1}$ e os valores de \mathbf{G}_t e \mathbf{W}_t , $\boldsymbol{\theta}_t$ é independente da informação anterior ao tempo t .

Agora pode-se introduzir a definição geral do DLM:

Definição 2.2. Para cada índice de tempo t , o Modelo Linear Normal multivariado é definido por

$$\text{Equação de Observação: } \mathbf{Y}_t = \mathbf{F}'_t \boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \boldsymbol{\nu}_t \sim N[\mathbf{0}, \mathbf{V}_t]$$

$$\text{Equação de Sistema: } \boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \boldsymbol{\omega}_t \sim N[\mathbf{0}, \mathbf{W}_t]$$

$$\text{Priori Inicial: } (\boldsymbol{\theta}_0 | \mathbf{D}_0) \sim N[\mathbf{m}_0, \mathbf{C}_0],$$

onde assume-se que as sequências de erros observacionais $\boldsymbol{\nu}_t$ e de evolução $\boldsymbol{\omega}_t$ são independentes ao longo do tempo e entre si, e independentes da priori $(\boldsymbol{\theta}_0 | \mathbf{D}_0)$.

Note que o erro ν_t é simplesmente uma perturbação aleatória no processo de medida das observações \mathbf{Y}_t . O erro de evolução ω_t , por outro lado, influencia no desenvolvimento do sistema ao longo do tempo. A suposição de que estes erros são independentes entre si e mutuamente, claramente separa estas duas fontes de variação estocástica e torna mais nítido o papel que cada uma representa. Se algum componente é dado como conhecido, basta assumir que a sua respectiva variância/covariância é zero.

O modelo de Espaço de Estados pode ser especificado de modo geral de maneira semelhante a acima, com uma distribuição à *priori* para o vetor de estados, mas com equações de observação e evolução dadas por

$$\begin{aligned}\mathbf{Y}_t &= h_t(\boldsymbol{\theta}_t, \nu_t) \\ \boldsymbol{\theta}_t &= g_t(\boldsymbol{\theta}_{t-1}, \omega_t),\end{aligned}$$

para funções arbitrárias g_t e h_t . Neste caso mais geral pode-se assumir distribuições diferentes da normal e relações não lineares entre os estados e entre as observações e os estados. Um exemplo importante são os modelos lineares generalizados, usados para tratar observações com distribuição pertencente à família exponencial (West & Harrison [39]). Tais modelos são apropriados para diversos problemas, tais como séries temporais discretas e séries com erros com distribuição diferente da normal.

2.2.1 Modelo Polinomial de Ordem 1

O DLM mais usado é o modelo polinomial de ordem 1, também chamado de passeio aleatório (*random walk*). Ele é usado sobretudo para previsões de curto prazo, onde neste horizonte de tempo assume-se que a série permanece constante. Este modelo é definido pela quádrupla

$$\{1, 1, \mathbf{V}_t, \mathbf{W}_t\}$$

Ele é introduzido pela seguinte definição

Definição 2.3. Para cada índice de tempo t , o Modelo Linear Dinâmico de ordem 1 é definido por

$$\begin{aligned}\text{Equação de Observação: } \mathbf{Y}_t &= \boldsymbol{\mu}_t + \nu_t, \nu_t \sim N[0, \mathbf{V}_t] \\ \text{Equação de Sistema: } \boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + \omega_t, \omega_t \sim N[0, \mathbf{W}_t] \\ \text{Priori Inicial: } (\boldsymbol{\mu}_0 | \mathbf{D}_0) &\sim N[\mathbf{m}_0, \mathbf{C}_0],\end{aligned}$$

onde assume-se que as sequências de erros observacionais ν_t e de evolução ω_t são independentes ao longo do tempo e entre si, e independentes da priori ($\mu_0|D_0$).

Por exemplo, ao modelar a demanda por um produto, μ_t representa a demanda real no tempo t e ν_t descreve uma flutuação aleatória, que naturalmente surge ao se postar as ordens dos clientes. De forma geral a demanda μ_t é considerada aproximadamente constante no curto prazo. É possível que mudanças significativas aconteçam no longo prazo, mas a variável aleatória ω_t implica que o pesquisador não deseja antecipar a direção da mudança no longo prazo.

2.2.2 Modelo Polinomial de Ordem 2

Modelos polinomiais de ordem 2 são usados para descrever processos estocásticos que apresentam tendência linear. São também chamados de modelos de crescimento linear (*linear growth models*). São caracterizados pela quádrupla

$$\left\{ \left(\begin{array}{c} 1 \\ 0 \end{array} \right), \left(\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \right), \mathbf{V}_t, \left(\begin{array}{cc} \sigma_\mu^2 & 0 \\ 0 & \sigma_\beta^2 \end{array} \right) \right\}, \quad (2.8)$$

A sua definição é dada por

Definição 2.4. Para cada índice de tempo t , o Modelo Linear Dinâmico de ordem dois é definido por

$$\begin{aligned} \text{Equação de Observação: } \mathbf{Y}_t &= \boldsymbol{\theta}_{1,t} + \nu_t, \nu_t \sim N[0, \mathbf{V}_t] \\ \text{Equação de Sistema: } \boldsymbol{\theta}_{1,t} &= \boldsymbol{\theta}_{1,t-1} + \boldsymbol{\theta}_{2,t-1} + \mathbf{w}_{1,t}, \mathbf{w}_{1,t} \sim N[0, \sigma_\mu^2] \\ \boldsymbol{\theta}_{2,t} &= \boldsymbol{\theta}_{2,t-1} + \mathbf{w}_{2,t}, \mathbf{w}_{2,t} \sim N[0, \sigma_\beta^2] \\ \text{Priori Inicial: } (\boldsymbol{\theta}_0|D_0) &\sim N[\mathbf{m}_0, \mathbf{C}_0], \end{aligned}$$

onde assume-se que as sequências de erros observacionais ν_t e de evolução \mathbf{w}_t são independentes ao longo do tempo e entre si, e independentes da priori ($\boldsymbol{\theta}_0|D_0$). onde

$$\boldsymbol{\theta}_t = \left(\begin{array}{c} \boldsymbol{\theta}_{1,t} \\ \boldsymbol{\theta}_{2,t} \end{array} \right), \mathbf{w}_t = \left(\begin{array}{c} \mathbf{w}_{1,t} \\ \mathbf{w}_{2,t} \end{array} \right). \quad (2.9)$$

Note que as matrizes \mathbf{G}_t e \mathbf{F}_t e as matrizes de covariância \mathbf{V}_t and \mathbf{W}_t são constantes. Neste caso, o modelo é invariante no tempo, também é chamado de DLM de séries temporais (TSDLM). Como usual, $\boldsymbol{\theta}_{1,t}$ é o nível da série e $\boldsymbol{\theta}_{2,t}$ representa o crescimento incremental.

2.2.3 Modelos Sazonais

A sazonalidade é usada para descrever flutuações periódicas no nível da série temporal. Assume-se que estas flutuações possuem um padrão regular, de forma que podem ser incorporadas no modelo. Como exemplo pode-se citar as alterações mensais que ocorrem na demanda por passagens aéreas. Este fator deve ser levado em consideração ao realizar as previsões, pois ele pode ter implicações importantes no desenvolvimento da série.

Em DLMS, existem duas formas usuais de se incorporar sazonalidade nos modelos: a de fatores de forma livre (*form-free seasonal factors*) e a de coeficientes de *fourier* (*fourier form*). Os fatores de forma livre modelam as flutuações como componentes do vetor de estado, um componente para cada flutuação existente. Os modelos de coeficiente de *fourier* tentam aproximar as flutuações como uma combinação linear de senos e cossenos. Neste trabalho optou-se por utilizar somente a sazonalidade com fatores de forma livre. As flutuações a serem modeladas possuem caráter mensal e podem ser bem aproximadas por doze fatores, um para cada mês do ano. Geralmente se aplica uma restrição nos fatores de forma livre. A mais usual é tornar a soma dos fatores zero, de forma que a tendência seja incorporada no modelo pelo componente polinomial, apresentado acima.

Antes de apresentar formalmente este tipo de modelo, é necessário a seguinte definição.

Definição 2.5. E_p ($p \times 1$) e a matriz de permutação P ($p \times p$) são definidas por:

$$E_p = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, P = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 1 \\ 1 & \dots & & 0 & 0 \end{pmatrix}. \quad (2.10)$$

Note que P é p -cíclica, ou seja, para $n \geq 0$, $P^{np} = I_p$.

Agora pode-se introduzir o modelo sazonal de forma livre:

Definição 2.6. Para cada índice de tempo t e variâncias observacional V_t , de evolução W_t e vetor de estados θ_t , o Modelo Sazonal de Forma Livre de período p é definido por

$$\{E_p, P, V_t, W_t\}. \quad (2.11)$$

Com o vetor de parâmetros $\boldsymbol{\theta}_t$, este DLM pode ser escrito como:

$$\begin{aligned} \text{Observação: } \mathbf{Y}_t &= \mathbf{E}'_p \boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \\ \text{Sistema: } \boldsymbol{\theta}_t &= \mathbf{P} \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t. \end{aligned}$$

Com $E[\boldsymbol{\theta}_t | \mathbf{D}_t] = \mathbf{m}_t = (\boldsymbol{\theta}_{t,0}, \dots, \boldsymbol{\theta}_{t,p-1})'$, a função de previsão é dada por:

$$\mathbf{f}_t(k) = \mathbf{E}'_p \mathbf{P}^k \mathbf{m}_t = \mathbf{m}_{t,j}, j = p|k. \quad (2.12)$$

Com a restrição $\mathbf{1}' \boldsymbol{\theta}_t = 0$, o modelo é chamado também *form-free seasonal effects*.

Para um ciclo trimestral, por exemplo, toma-se \mathbf{G}_t como sendo uma matriz $p \times p$, onde $p = 4$, ou seja, cada fator representa um trimestre. Ao decompor um conjunto de p fatores sazonais em um nível desazonalizado e p desvios sazonais, os desvios são chamados efeitos sazonais (*seasonal effects*). Por fim, impõe-se uma restrição que permite a sua identificabilidade. Geralmente a restrição é tal que a soma dos efeitos é zero, para sazonalidade aditiva. Outros tipos de restrições são possíveis, mas não são discutidos aqui. Para mais detalhes, veja (West & Harrison [39]) e (Harvey [17]). É muito comum sobrepor um DLM de efeito sazonal com DLMs polinomiais, de forma a obter um componente de efeitos sazonais com nível diferente de zero. Note que a restrição de soma de efeitos igual a zero é a mais usada para DLMs de forma livre.

2.2.4 Estimação dos Estados e Previsão

Nesta seção assume-se que o modelo é dado, i.e., as densidades $(\mathbf{Y}_t | \boldsymbol{\theta}_t)$ e $(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$ são conhecidas. O problema de construção de modelos não será abordado neste trabalho, mas pode-se consultar as referências (West & Harrison [39], Durbin & Koopman [10] e Petris *et al.* [29]). Para um dado modelo de Espaço de Estados, o mais importante é fazer inferência sobre os estados não observáveis e prever futuros valores da série com base nos valores observados até o presente. A estimação e a previsão são resolvidos calculando as densidades condicionais dos estados e das observações, dado a informação disponível. Neste caso, pode-se usar os resultados padrões da teoria de distribuições multivariadas normais, veja (Johnson & Wichern [22]). Pode-se provar que o vetor aleatório $(\boldsymbol{\theta}_{1:t}, \mathbf{Y}_{1:t})$ tem distribuição normal para todo índice de tempo t , $t \geq 1$. Então segue, a partir dos lemas (A.1, A.2) do apêndice que as distribuições marginais e as condicionais são também normais, sendo completamente determinadas pela sua média e variância. A solução ao problema de filtragem é dada pelo filtro de Kalman. O próximo teorema é chave para as seções seguintes:

Teorema 2.1. Para cada tempo t , no DLM multivariado da definição 2.2, as distribuições dos estados e as preditivas um passo a frente são dadas por:

(a) *Posteriori no tempo $t - 1$:*

$$(\boldsymbol{\theta}_{t-1} | \mathbf{D}_{t-1}) \sim N[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}].$$

(b) *Priori no tempo t :*

$$(\boldsymbol{\theta}_t | \mathbf{D}_{t-1}) \sim N[\mathbf{a}_t, \mathbf{R}_t],$$

onde

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}, \quad \mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t.$$

(c) *Previsão um passo a frente:*

$$(\mathbf{Y}_t | \mathbf{D}_{t-1}) \sim N[\mathbf{f}_t, \mathbf{Q}_t],$$

onde

$$\mathbf{f}_t = \mathbf{F}_t' \mathbf{a}_t, \quad \mathbf{Q}_t = \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + \mathbf{V}_t.$$

(d) *Posteriori no tempo t :*

$$(\boldsymbol{\theta}_t | \mathbf{D}_t) \sim N[\mathbf{m}_t, \mathbf{C}_t],$$

onde

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t \mathbf{e}_t, \quad \mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{Q}_t \mathbf{A}_t', \quad \mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t \mathbf{Q}_t^{-1}, \quad \mathbf{e}_t = \mathbf{Y}_t - \mathbf{f}_t.$$

Prova: A prova é omitida, mas pode ser encontrada em West & Harrison [39], por exemplo.

O erro de previsão um passo a frente é \mathbf{e}_t e a quantidade \mathbf{A}_t chama-se coeficiente adaptivo, no tempo t . Uma identidade útil, obtida acima, é $\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t \mathbf{Q}_t^{-1} = \mathbf{C}_t \mathbf{F}_t \mathbf{V}_t^{-1}$.

A expressão da média do filtro \mathbf{m}_t tem a forma intuitiva de erro-correção, onde a correção depende do peso dado a ela. A importância da inovação (diferença entre o valor observado e a previsão) é de acordo com o valor do coeficiente adaptativo \mathbf{A}_t . Note que o peso depende da variância observacional \mathbf{V}_t e de \mathbf{R}_t .

Na prática, ao avaliar a matriz de variância do filtro \mathbf{C}_t (e consequentemente também

\mathbf{R}_t e \mathbf{Q}_t) usando a formulação do teorema acima pode trazer instabilidade numérica que pode levar a matrizes de variância não simétricas e que sejam negativa definidas, i.e., matrizes de variância inválidas. Durante o desenvolvimento do trabalho notou-se as instabilidades numéricas descritas acima com certa frequência. Algoritmos mais estáveis que solucionam este problema de forma satisfatória já foram propostos na literatura. Neste trabalho adotou-se a técnica de decomposição por valor singular (SVD ou *singular value decomposition*). Alguns detalhes podem ser encontrados em Petris *et al.* [29]. Basicamente, o filtro SVD pode ser visto como um filtro de raiz quadrada usual, por exemplo, se \mathbf{A} é a matriz de variância, a decomposição leva a $\mathbf{A} = \mathbf{U}\mathbf{D}^2\mathbf{U}'$, onde \mathbf{D} é uma matriz diagonal e \mathbf{U} é uma matriz ortogonal, i.e., $\mathbf{U}^{-1} = \mathbf{U}'$.

2.2.5 Distribuições de Previsão

Definição 2.7. A função de previsão $f_t(k)$, no tempo t , é definida para todos os inteiros $k \geq 1$ como

$$f_t(k) = E[\boldsymbol{\mu}_{t+k}|\mathbf{D}_t] = E[\mathbf{F}'_{t+k}\boldsymbol{\theta}_{t+k}|\mathbf{D}_t] = E[\mathbf{Y}_{t+k}|\mathbf{D}_t], \quad (2.13)$$

onde

$$\boldsymbol{\mu}_t = \mathbf{F}'_t\boldsymbol{\theta}_t, \quad (2.14)$$

é a função de resposta média para qualquer tempo $t \geq 0$.

A função de previsão é muito importante ao se projetar DLMs, como veremos nos capítulos seguintes. Os resultados seguintes oferecem as distribuições de previsão dos estados e das observações.

Teorema 2.2. Em cada tempo t , as distribuições marginais preditivas k -passos à frente, $k \geq 0$, são dadas por

(a) Distribuição dos Estados: $(\boldsymbol{\theta}_{t+k}|\mathbf{D}_t) \sim N[\mathbf{a}_t(k), \mathbf{R}_t(k)]$

(b) Distribuição de Previsão: $(\mathbf{Y}_{t+k}|\mathbf{D}_t) \sim N[f_t(k), \mathbf{Q}_t(k)]$

onde

$$f_t(k) = \mathbf{F}'_{t+k}\mathbf{a}_t(k), \quad \mathbf{Q}_t(k) = \mathbf{F}'_{t+k}\mathbf{R}_t(k)\mathbf{F}_{t+k} + \mathbf{V}_{t+k}. \quad (2.15)$$

As quantidades acima são calculadas recursivamente usando

$$\mathbf{a}_t(k) = \mathbf{G}_{t+k}\mathbf{a}_t(k-1) \quad (2.16)$$

$$\mathbf{R}_t(k) = \mathbf{G}_{t+k}\mathbf{R}_t(k-1)\mathbf{G}'_{t+k} + \mathbf{W}_{t+k}, \quad (2.17)$$

junto com os valores iniciais $\mathbf{a}_t(0) = \mathbf{m}_t, \mathbf{R}_t(0) = \mathbf{C}_t$. As expressões para as covariâncias entre os estados e as observações podem ser vistas em West & Harrison [39].

Prova: A prova é por indução, assumamos que $(\boldsymbol{\theta}_{t+k-1} | \mathbf{D}_t) \sim N[\mathbf{a}_t(k-1), \mathbf{R}_t(k-1)]$ é verdade. Usando a equação de evolução dos estados $(\boldsymbol{\theta}_{t+k} | \boldsymbol{\theta}_{t+k-1}, \mathbf{D}_t) \sim N[\mathbf{G}_{t+k} \boldsymbol{\theta}_{t+k-1}, \mathbf{W}_{t+k}]$ e o lema A.1 do apêndice, temos que $(\boldsymbol{\theta}_{t+k} | \mathbf{D}_t) \sim N[\mathbf{a}_t(k), \mathbf{R}_t(k)]$ com média e variância dadas como acima. Para a previsão das observações, use a equação de observação padrão k passos à frente $(\mathbf{Y}_{t+k} | \boldsymbol{\theta}_{t+k}) \sim N[\mathbf{F}'_{t+k} \boldsymbol{\theta}_{t+k}, \mathbf{V}_{t+k}]$ e o resultado (a) acima, junto com o lema A.1. Isto mostra $(\mathbf{Y}_{t+k} | \mathbf{D}_t) \sim N[\mathbf{f}_t(k), \mathbf{Q}_t(k)]$, com média e variância dados acima.

Nos casos especiais onde a matriz de evolução \mathbf{G}_t e \mathbf{F}_t são constantes, os resultados acima simplificam consideravelmente, e são detalhados abaixo.

Corolário 2.1. *Se a matriz de evolução $\mathbf{G}_t = \mathbf{G}$ é constante para todo t , então para $k \geq 0$,*

$$\mathbf{a}_t(k) = \mathbf{G}^k \mathbf{m}_t, \quad \mathbf{f}_t(k) = \mathbf{F}'_{t+k} \mathbf{G}^k \mathbf{m}_t. \quad (2.18)$$

Se \mathbf{F}_t for também invariante no tempo, de forma que o modelo seja um TSDLM, então a função de previsão fica $\mathbf{f}_t(k) = \mathbf{F}' \mathbf{G}^k \mathbf{m}_t$.

2.2.6 Distribuições de Suavização

A inferência sobre os estados passados, dado todas as observações, é feita pelas equações de suavização, ou análise retrospectiva. Geralmente é empregada para melhorar o entendimento do processo estocástico e auxiliar no processo de decisões quanto ao desenvolvimento futuro do processo. O teorema abaixo demonstra a distribuição condicional de estados passados.

Teorema 2.3. *No DLM definido por $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$, defina $\mathbf{B}_t = \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1}$. Então para todo $k, 1 \leq k < t$, as distribuições marginais de suavização são*

$$(\boldsymbol{\theta}_{t-k} | \mathbf{D}_t) \sim N[\mathbf{a}_t(-k), \mathbf{R}_t(-k)],$$

onde

$$\mathbf{a}_t(-k) = \mathbf{m}_{t-k} + \mathbf{B}_{t-k} [\mathbf{a}_t(-k+1) - \mathbf{a}_{t-k+1}] \quad (2.19)$$

$$\mathbf{R}_t(-k) = \mathbf{C}_{t-k} + \mathbf{B}_{t-k} [\mathbf{R}_t(-k+1) - \mathbf{R}_{t-k+1}] \mathbf{B}'_{t-k}, \quad (2.20)$$

com valores iniciais $\mathbf{a}_t(0) = \mathbf{m}_t, \mathbf{R}_t(0) = \mathbf{C}_t, \mathbf{a}_{t-k}(1) = \mathbf{a}_{t-k+1}, \mathbf{R}_{t-k}(1) = \mathbf{R}_{t-k+1}$

Prova: A prova é feita por indução. Note que $(\boldsymbol{\theta}_{t-k+1} | \boldsymbol{\theta}_{t-k}, \mathbf{D}_{t-k}) \sim N[\mathbf{G}_{t-k+1} \boldsymbol{\theta}_{t-k}, \mathbf{W}_{t-k+1}]$

e $(\boldsymbol{\theta}_{t-k}|\mathbf{D}_{t-k}) \sim N[\mathbf{m}_{t-k}, \mathbf{C}_{t-k}]$. Aplicando o lema A.2, reconstrói-se a distribuição conjunta de $(\boldsymbol{\theta}_{t-k+1}, \boldsymbol{\theta}_{t-k}|\mathbf{D}_{t-k})$:

$$\begin{pmatrix} \boldsymbol{\theta}_{t-k+1} \\ \boldsymbol{\theta}_{t-k} \end{pmatrix} | \mathbf{D}_{t-k} \sim N \left[\begin{pmatrix} \mathbf{a}_{t-k+1} \\ \mathbf{m}_{t-k} \end{pmatrix}, \begin{pmatrix} \mathbf{R}_{t-k+1} & \mathbf{G}_{t-k+1} \mathbf{C}_{t-k} \\ \mathbf{C}_{t-k} \mathbf{G}'_{t-k+1} & \mathbf{C}_{t-k} \end{pmatrix} \right],$$

onde

$$\mathbf{R}_{t-k+1} = \mathbf{W}_{t-k+1} + \mathbf{G}_{t-k+1} \mathbf{C}_{t-k} \mathbf{G}'_{t-k+1}, \quad (2.21)$$

e

$$(\boldsymbol{\theta}_{t-k}|\boldsymbol{\theta}_{t-k+1}, \mathbf{D}_t) \sim N_d[\mathbf{h}_t(-k), \mathbf{H}_t(-k)], \quad (2.22)$$

onde

$$\begin{aligned} \mathbf{h}_t(-k) &= \mathbf{m}_{t-k} + \mathbf{B}_{t-k}[\boldsymbol{\theta}_{t-k+1} - \mathbf{a}_{t-k+1}] \\ \mathbf{H}_t(-k) &= \mathbf{C}_{t-k} - \mathbf{B}_{t-k} \mathbf{R}_{t-k+1} \mathbf{B}'_{t-k}. \end{aligned}$$

Agora aplicando o lema A.1 com $(\boldsymbol{\theta}_{t-k}|\boldsymbol{\theta}_{t-k+1}, \mathbf{D}_t)$ e $(\boldsymbol{\theta}_{t-k+1}|\mathbf{D}_t) \sim N[\mathbf{a}_t(-k+1), \mathbf{R}_t(-k+1)]$ temos que

$$(\boldsymbol{\theta}_{t-k}|\mathbf{D}_t) \sim N[\mathbf{a}_t(-k), \mathbf{R}_t(-k)], \quad (2.23)$$

onde

$$\begin{aligned} \mathbf{a}_t(-k) &= \mathbf{m}_{t-k} + \mathbf{B}_{t-k}[\mathbf{a}_t(-k+1) - \mathbf{a}_{t-k+1}] \\ \mathbf{R}_t(-k) &= \mathbf{C}_{t-k} - \mathbf{B}_{t-k}[\mathbf{R}_{t-k+1} - \mathbf{R}_t(-k+1)] \mathbf{B}'_{t-k}, \end{aligned}$$

e

$$\mathbf{a}_t(0) = \mathbf{m}_t, \quad \mathbf{R}_t(0) = \mathbf{C}_t. \quad (2.24)$$

2.2.7 Variâncias Observacionais

Na quádrupla que define um DLM, $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$, geralmente a matriz de variância observacional \mathbf{V}_t é desconhecida e precisa ser estimada. Como esta variância é geralmente a principal fonte de incerteza no processo estocástico sendo modelado, foram

desenvolvidos procedimentos de aprendizagem usando o enfoque bayesiano no caso da variância ser desconhecida mas constante, i.e., $\mathbf{V}_t = \mathbf{V}$ para todo t . É mais conveniente trabalhar com a sua inversa, chamada de precisão ϕ , onde $\phi = 1/\mathbf{V}$. A próxima seção trata de variâncias com pequenas mudanças estocásticas, onde a análise introduzida aqui é ligeiramente modificada para tratar do caso onde a variância possui uma evolução temporal. Segue abaixo a definição geral.

Definição 2.8. Para cada t , o DLM univariado com aprendizagem de variância observacional é definido por

$$\begin{aligned} \text{Equação de Observação} & : Y_t = \mathbf{F}_t' \boldsymbol{\theta}_t + v_t, v_t \sim N[0, V] \\ \text{Equação de Sistema} & : \boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \boldsymbol{\omega}_t \sim N[\mathbf{0}, \mathbf{V} \mathbf{W}_t^*] \\ \text{Priori Inicial} & : (\boldsymbol{\theta}_0 | \mathbf{D}_0, \phi) \sim N[\mathbf{m}_0, \mathbf{V} \mathbf{C}_0^*] \\ & : (\phi | \mathbf{D}_0) \sim G \left[\frac{n_0}{2}, \frac{n_0 S_0}{2} \right], \end{aligned}$$

onde assume-se que as sequências de erros observacionais v_t e de evolução $\boldsymbol{\omega}_t$ são independentes ao longo do tempo e entre si, e independentes da priori $(\boldsymbol{\theta}_0 | \mathbf{D}_0, \phi)$.

O teorema abaixo mostra as equações de evolução. Note que as distribuições de previsão agora são t -Student ao invés de gaussianas.

Teorema 2.4. O DLM definido acima possui as seguintes distribuições condicionais (a) Condicionado a V :

$$\begin{aligned} (\boldsymbol{\theta}_{t-1} | \mathbf{D}_{t-1}, V) & \sim N[\mathbf{m}_{t-1}, \mathbf{V} \mathbf{C}_{t-1}^*] \\ (\boldsymbol{\theta}_t | \mathbf{D}_{t-1}, V) & \sim N[\mathbf{a}_t, \mathbf{V} \mathbf{R}_t^*] \\ (Y_t | \mathbf{D}_{t-1}, V) & \sim N[f_t, \mathbf{V} \mathbf{Q}_t^*] \\ (\boldsymbol{\theta}_t | \mathbf{D}_t, V) & \sim N[\mathbf{m}_t, \mathbf{V} \mathbf{C}_t^*], \end{aligned}$$

com

$$\begin{aligned} \mathbf{a}_t & = \mathbf{G}_t \mathbf{m}_{t-1}, & \mathbf{R}_t^* & = \mathbf{G}_t \mathbf{C}_{t-1}^* \mathbf{G}_t' + \mathbf{W}_t^* \\ f_t & = \mathbf{F}_t' \mathbf{a}_t, & \mathbf{Q}_t^* & = 1 + \mathbf{F}_t' \mathbf{R}_t^* \mathbf{F}_t \\ e_t & = Y_t - f_t, & \mathbf{A}_t & = \mathbf{R}_t^* \mathbf{F}_t / \mathbf{Q}_t^* \\ \mathbf{m}_t & = \mathbf{a}_t + \mathbf{A}_t e_t, & \mathbf{C}_t^* & = \mathbf{R}_t^* - \mathbf{A}_t \mathbf{A}_t' \mathbf{Q}_t^*. \end{aligned}$$

(b) Para a precisão $\phi = 1/V$, temos:

$$\begin{aligned}(\phi|\mathbf{D}_{t-1}) &\sim G\left[\frac{n_{t-1}}{2}, \frac{n_{t-1}S_{t-1}}{2}\right], \\(\phi|\mathbf{D}_t) &\sim G\left[\frac{n_t}{2}, \frac{n_tS_t}{2}\right],\end{aligned}$$

onde $n_t = n_{t-1} + 1$ e $S_t = S_{t-1} + \frac{S_{t-1}}{n_t} \left(\frac{e_t^2}{Q_t} - 1\right)$,

(c) Marginalizando as distribuições em relação a \mathbf{V} , temos:

$$\begin{aligned}(\boldsymbol{\theta}_{t-1}|\mathbf{D}_{t-1}) &\sim T_{n_{t-1}}[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}], \\(\boldsymbol{\theta}_t|\mathbf{D}_{t-1}) &\sim T_{n_{t-1}}[\mathbf{a}_t, \mathbf{R}_t], \\(Y_t|\mathbf{D}_{t-1}) &\sim T_{n_{t-1}}[f_t, Q_t], \\(\boldsymbol{\theta}_t|\mathbf{D}_t) &\sim T_{n_t}[\mathbf{m}_t, \mathbf{C}_t],\end{aligned}$$

onde $\mathbf{R}_t = S_{t-1}\mathbf{R}_t^*$, $Q_t = S_{t-1}Q_t^*$ e $\mathbf{C}_t = S_t\mathbf{C}_t^*$

(d) As equações de atualização são dadas abaixo, onde $Q_t = \mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t + S_{t-1}$ e $\mathbf{A}_t = \mathbf{R}_t\mathbf{F}_t/Q_t$:

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t e_t, \quad \mathbf{C}_t = \frac{S_t}{S_{t-1}}(\mathbf{R}_t - \mathbf{A}_t\mathbf{A}_t'Q_t). \quad (2.25)$$

Prova: A demonstração usa conceitos da teoria da distribuição normal-gama e será omitida por usar resultados padrões (ver West & Harrison [39]).

2.2.8 Variâncias Observacionais Dinâmicas

Assumir que a variância observacional \mathbf{V}_t é invariante no tempo é uma hipótese muito restritiva em muitas aplicações. Podem haver elementos adicionais que não foram incorporados no modelo. Estes elementos são fontes extras de incerteza e podem ser modelados através de uma variância observacional dinâmica. A dinâmica da variância é dada por um distúrbio aleatório γ_t com distribuição beta, de forma que esta variável tenha esperança igual a δ e oscile entre zero e um.

Definição 2.9. Para cada t , o DLM univariado com dinâmica na variância observacional

é definido por

$$\begin{aligned}
\text{Equação de Observação} & : Y_t = \mathbf{F}_t' \boldsymbol{\theta}_t + v_t, v_t \sim N[0, V_t] \\
\text{Equação de Sistema} & : \boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \boldsymbol{\omega}_t \sim T_{n_{t-1}}[\mathbf{0}, \mathbf{W}_t] \\
\text{Equação de Precisão} & : \phi_t = \gamma_t \phi_{t-1} / \delta \\
& \text{com} : \gamma_t \sim \text{Beta}[\delta n_{t-1} / 2, (1 - \delta) n_{t-1} / 2] \\
\text{Priori Inicial} & : (\boldsymbol{\theta}_0 | \mathbf{D}_0, \phi_0) \sim N[\mathbf{m}_0, \mathbf{V}_0 / \phi_0] \\
& : (\phi_0 | \mathbf{D}_0) \sim G\left[\frac{n_0}{2}, \frac{n_0 S_0}{2}\right],
\end{aligned}$$

onde assume-se que as sequências de erros observacional v_t e de evolução $\boldsymbol{\omega}_t$ são independentes ao longo do tempo e entre si, e independentes da priori $(\boldsymbol{\theta}_0 | \mathbf{D}_0, \phi_0)$.

O teorema abaixo mostra os principais resultados da evolução temporal da variância.

Teorema 2.5. *O DLM definido acima possui as seguintes distribuições condicionais*
(a) *Evolução:*

$$\begin{aligned}
(\boldsymbol{\theta}_{t-1} | \mathbf{D}_{t-1}) & \sim T_{n_{t-1}}[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}] \\
(\boldsymbol{\theta}_t | \mathbf{D}_{t-1}) & \sim T_{n_{t-1}}[\mathbf{a}_t, \mathbf{R}_t] \\
(Y_t | \mathbf{D}_{t-1}) & \sim T_{\delta n_{t-1}}[f_t, Q_t] \\
(\boldsymbol{\theta}_t | \mathbf{D}_t) & \sim T_{\delta n_t}[\mathbf{m}_t, \mathbf{C}_t] \\
(\phi_{t-1} | \mathbf{D}_{t-1}) & \sim G[n_{t-1} / 2, d_{t-1} / 2] \\
(\phi_t | \mathbf{D}_{t-1}) & \sim G[\delta n_{t-1} / 2, \delta d_{t-1} / 2],
\end{aligned}$$

com

$$\begin{aligned}
\mathbf{a}_t & = \mathbf{G}_t \mathbf{m}_{t-1}, & \mathbf{R}_t & = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t \\
f_t & = \mathbf{F}_t' \mathbf{a}_t, & Q_t & = S_{t-1} + \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t \\
e_t & = Y_t - f_t, & \mathbf{A}_t & = \mathbf{R}_t \mathbf{F}_t / Q_t \\
\mathbf{m}_t & = \mathbf{a}_t + \mathbf{A}_t e_t, & \mathbf{C}_t & = (S_t / S_{t-1}) (\mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t' Q_t),
\end{aligned}$$

onde $n_t = \delta n_{t-1} + 1$, $d_t = \delta d_{t-1} + S_{t-1} e_t^2 / Q_t$ e $S_{t-1} = d_{t-1} / n_{t-1}$

Prova: A demonstração usa conceitos padrões de teoria da probabilidade e será omitida, mais detalhes veja (West & Harrison [39]) ou (James [21]).

Note que δ é uma espécie de fator de desconto e tipicamente tem valores na faixa (0.8, 0.99). A definição de fatores de desconto é dada na subseção seguinte.

2.2.9 Fatores de Desconto para a Especificação da Matriz de Evolução

Ao determinar a influência de observações passadas, a estrutura e a magnitude da matriz de covariância \mathbf{W}_t tem grande importância. Valores grandes de \mathbf{W}_t implicam que muita informação é perdida ao se passar do tempo $t - 1$ ao tempo t . Defina $\mathbf{P}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}'_t$, ou seja, \mathbf{P}_t seria a variância caso não haja incerteza sobre a evolução dos estados, i.e., $\mathbf{W}_t = \mathbf{0}$. Tem-se que $\mathbf{R}_t = \mathbf{P}_t + \mathbf{W}_t$, logo \mathbf{W}_t expressa a perda de informação em cada transição de tempo, a magnitude da perda relacionada as proporções de \mathbf{W}_t em relação a \mathbf{P}_t . Desta maneira, pode-se conceber que \mathbf{R}_t seja uma proporção de \mathbf{P}_t , i.e.,

$$\mathbf{R}_t = \frac{1}{\delta} \mathbf{P}_t, \quad (2.26)$$

de modo que

$$\mathbf{W}_t = \frac{1 - \delta}{\delta} \mathbf{P}_t, \quad (2.27)$$

onde δ é denominado fator de desconto e $\delta \in (0, 1)$. Tipicamente δ fica na faixa $(0.8, 0.99)$. Existem vários critérios para a escolha do valor deste parâmetro. Os mais usados são a maximização da função de verossimilhança e a minimização de alguma estatística preditiva, tal como o MSE (*Mean Square Error*) ou o SMAPE (*Symmetric Mean Absolute Percentage Error*). É possível também especificar um vetor de fatores de desconto $\delta_1, \dots, \delta_p$ para um DLM com p estados, onde cada fator seria relacionado ao respectivo elemento do vetor de estados. Neste caso o modelo chama-se DLM de componentes de desconto (*Component Discount DLM*). Para mais detalhes, veja (Solhjell [32]).

Capítulo 3

Modelos *Damped Holt* e Estimação Bayesiana via Multiprocessos

Neste capítulo apresenta-se os modelos polinomiais de ordem dois com abafamento (*damping*) da tendência, conhecidos na literatura como *Damped Holt*, caso a tendência seja aditiva, e *Damped Pegels*, caso a tendência seja multiplicativa. O parâmetro de amortecimento serve para atenuar uma curva de tendência muito acentuada, de forma que as previsões sejam mais condizentes com o desenvolvimento posterior da série temporal. Discute-se as principais características destes modelos e a estimação dos seus parâmetros pela técnica bayesiana de multiprocessos classe I, onde emprega-se uma mistura de densidades para criar a densidade preditiva dos estados e das observações. Os principais resultados apresentados na literatura são mostrados e discutidos neste trabalho. Para completar a revisão, são apresentadas as diferentes medidas frequentistas de comparação de performance de modelos, assim como as suas principais características.

3.1 O Modelo Aditivo *Damped Holt*

O modelo *Damped Holt* foi introduzido como uma maneira de melhorar a performance preditiva do modelo aditivo *Holt* tradicional (Makridakis & Hibon [26], McKenzie & Gardner [27], Gooijer & Hyndman [15]). Este modelo é definido como um modelo polinomial de ordem dois univariado com um parâmetro adicional ψ para abafamento da tendência linear (Gardner [12]). A motivação para a inclusão deste parâmetro é a melhoria

na performance preditiva. Neste caso, ele é caracterizado pela quádrupla abaixo.

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \psi \\ 0 & \psi \end{pmatrix}, \mathbf{V}_t, \begin{pmatrix} \omega_{1,t} & 0 \\ 0 & \omega_{2,t} \end{pmatrix} \right\}, \quad (3.1)$$

de forma que a sua definição seja

Definição 3.1. Para cada índice de tempo t , o Modelo Linear Dinâmico de ordem dois Damped Holt é definido por

$$\text{Equação de Observação: } Y_t = \mathbf{F}_t' \boldsymbol{\theta}_t + v_t, \quad v_t \sim N[0, V_t]$$

$$\text{Equação de Sistema: } \boldsymbol{\theta}_{1,t} = \boldsymbol{\theta}_{1,t-1} + \psi \boldsymbol{\theta}_{2,t-1} + \boldsymbol{\omega}_{1,t}, \quad \boldsymbol{\omega}_{1,t} \sim N[0, \boldsymbol{\sigma}_\mu^2]$$

$$\boldsymbol{\theta}_{2,t} = \psi \boldsymbol{\theta}_{2,t-1} + \boldsymbol{\omega}_{2,t}, \quad \boldsymbol{\omega}_{2,t} \sim N[0, \boldsymbol{\sigma}_\beta^2]$$

$$\text{Priori Inicial: } (\boldsymbol{\theta}_0 | D_0) \sim N[\mathbf{m}_0, \mathbf{C}_0],$$

onde assume-se que as sequências de erros observacionais v_t e de evolução $\boldsymbol{\omega}_t$ são independentes ao longo do tempo e entre si, e independentes da priori $(\boldsymbol{\theta}_0 | D_0)$. onde

$$\boldsymbol{\theta}_t = \begin{pmatrix} \boldsymbol{\theta}_{1,t} \\ \boldsymbol{\theta}_{2,t} \end{pmatrix}, \quad \boldsymbol{\omega}_t = \begin{pmatrix} \boldsymbol{\omega}_{1,t} \\ \boldsymbol{\omega}_{2,t} \end{pmatrix}, \quad (3.2)$$

Note que as matrizes \mathbf{G}_t e \mathbf{F}_t são constantes. Como usual, $\boldsymbol{\theta}_{1,t}$ é o nível da série e $\boldsymbol{\theta}_{2,t}$ representa crescimento incremental. O escalar V_t e a matriz de covariância \mathbf{W}_t podem ser constantes no tempo, mas existe a possibilidade de elas terem uma dinâmica temporal. A estimação do parâmetro ψ pode ser feita pela maximização da função de verossimilhança, onde neste caso este parâmetro é estimado junto com os demais parâmetros do modelo de forma conjunta em uma malha de valores. Outra alternativa de estimação é minimizar alguma estatística de performance preditiva, tal como o erro de mínimo quadrado (*mean square error* ou MSE) ou o erro de desvio absoluto (*mean absolute deviation* ou MAD). Outra opção é a minimização da variância dos resíduos (Hyndman *et al.* [19] e Hyndman & Khandakar [20]). O uso de multiprocessos é outra alternativa para estimar o parâmetro, mas não é muito adotado na literatura. Note que este parâmetro é invariante no tempo nesta formulação. A função de previsão k passos à frente para este modelo é dada por

$$\mathbf{f}_t(k) = \mathbf{F}_t' \mathbf{G}^k \mathbf{m}_t = \mathbf{m}_{1,t} + \left(\sum_{j=1}^k \psi^j \right) \mathbf{m}_{2,t}. \quad (3.3)$$

O parâmetro ψ controla a taxa de crescimento das previsões. Esta função se aproxima assintoticamente de uma linha reta dada por $\mathbf{m}_{1,t} + \left(\frac{\psi}{1-\psi} \right) \mathbf{m}_{2,t}$ se $0 \leq \psi \leq 1$. Este é o caso

mais usual na prática. Quando $\psi \geq 1$, a função geralmente é empregada para modelar séries com tendência exponencial. Caso $\psi = 1$ temos o modelo *Holt* tradicional. O último caso acontece se $\psi = 0$, onde temos então a suavização exponencial simples e o modelo se torna polinomial de ordem um, i.e., sem um componente de crescimento no vetor de estados. O princípio da superposição se aplica sem restrições aqui, pois assume-se que as distribuições das fontes de erro são normais. Existe na literatura modelos com uma única fonte de erro (*single source of error models*), mas pode-se provar que alguns desses modelos são apenas uma reparametrização do modelo aqui apresentado, veja por exemplo (Hyndman *et al.* [19]).

3.2 O Modelo Multiplicativo *Damped Pegels*

Outra opção para modelos lineares abafados é o emprego de modelos multiplicativos, também chamados de *Damped Pegels* (Taylor [35]). A aplicação da transformação *log* nestes modelos permite que os mesmos sejam considerados para análise, como discutido abaixo. Este método lida com a tendência de maneira multiplicativa, já que as previsões são formadas pelo produto do nível $\theta_{1,t}^*$ e da taxa de crescimento $\theta_{2,t}^*$. Neste caso a equação de observação é escrita como

$$Y_t = \theta_{1,t}^* \theta_{2,t-1}^{*\psi} e^{v_t}, \quad (3.4)$$

e a equação de evolução dos estados permanece a mesma. Assume-se que $(\log(Y_t)|\theta_t) \sim N[\mathbf{F}'_t \theta_t, v_t]$, i.e., a variável resposta Y_t tem distribuição *lognormal*, onde $\mathbf{F}'_t = (1, 0)$. A aplicação da transformação logarítmica na equação acima faz com que o modelo tenha a formulação aditiva:

$$\log(Y_t) = \theta_{1,t} + \psi \theta_{2,t-1} + v_t, \quad (3.5)$$

onde

$$\theta_{i,t} = \log(\theta_{i,t}^*), \quad i = 1, 2. \quad (3.6)$$

A esperança é expressa como $E[Y_t|\theta_t] = \exp(\mathbf{F}'_t \theta_t + v_t/2)$. A função de previsão k passos à frente agora se torna

$$\mathbf{f}_t(k) = \alpha_t \beta_t^{\sum_{j=1}^k \psi^j}, \quad (3.7)$$

onde

$$\alpha_t = e^{m_{1,t} + v_t/2} \text{ e } \beta_t = e^{m_{2,t}}. \quad (3.8)$$

É importante resaltar que a mediana ($\alpha_t = e^{m_{1,t}}$) é um melhor preditor do que a média neste caso, pois a distribuição *lognormal* possui assimetria.

Note que o princípio da superposição de modelos ainda se aplica, pois a transformação cria variáveis normais e a equação volta a ser aditiva. Um grande campo de aplicação deste modelo é economia e finanças, em especial regressões em termos de taxas de mudança da variável Y_t em função da variável θ_t . A interpretação dos diferentes valores que o amortecedor ψ pode tomar é a mesma do modelo aditivo, como descrito na seção anterior. Uma limitação dos modelos multiplicativos é a sua incapacidade de lidar com séries que possuam zeros. Note que quando a mudança de nível da série ($\theta_{2,t}$) tende a ser muito menor que o próprio nível da série ($\theta_{1,t}$), as previsões dos modelos *Damped Holt* e *Damped Pegels* acabam sendo muito similares. Taylor [35] mostra a prova analítica deste fato. Para mais detalhes sobre o modelo *Damped Pegels*, veja (Taylor [35], Gardner [12]).

3.3 Estimação Bayesiana via Multiprocessos

Pode-se assumir que um DLM definido pelo seu conjunto de parâmetros, denotado por α , seja adequado para um certo processo estocástico. É muito comum, no entanto, que o comportamento da série seja tal que mudanças nos parâmetros do modelo sejam necessárias ao longo do desenvolvimento da série. Esta idéia é também uma das motivações para a introdução de intervenção e monitoramento em DLMs. A intervenção permite mudanças nos parâmetros, além de mudanças estruturais na quádrupla que define o modelo $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$. O uso de técnicas de monitoramento automático reconhece a possibilidade de que talvez um único DLM não seja suficiente para modelar a série em estudo.

Introduz-se a noção de que existem alternativas para o uso de um único modelo padrão ao se considerar classes de DLMs, sendo que a combinação linear de modelos de uma classe cria um supermodelo que seja adequado para a série. Modelos multiprocesso são definidos como uma combinação linear de várias instâncias de DLMs, cada instância com um conjunto distinto de parâmetros α . Multiprocessos são também denominados modelos de mistura (*mixture models*). Denota-se uma instância de DLM, definido pela sua quádrupla, como $M_t : \{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$, condicionado a informação inicial \mathbf{D}_0 .

O conjunto de parâmetros α está sujeito a incertezas e tipicamente inclui fatores de desconto, variâncias, algumas das matrizes que definem a quádrupla do modelo ou parâmetros de transformação para as variáveis independentes. Como estas quantidades agora são consideradas incertas, elas são colocadas no condicionamento de todas as distribuições na análise do modelo. A informação inicial \mathbf{D}_0 representa todas as quantidades conhecidas. A dependência do modelo das quantidades incertas é denotada por $M_t = M_t(\alpha), (t = 1, 2, \dots)$. Logo $M_t(\alpha)$ é um DLM para o tempo t e um dado valor de α .

Seja \mathbf{A} o conjunto de valores possíveis para α , onde assume-se que \mathbf{A} é discreto e finito para fins práticos. A classe de DLMs no tempo t é dada por $\{M_t(\alpha) : \alpha \in \mathbf{A}\}$. Neste contexto, existem duas classes de supermodelos, a saber:

1. Classe I: Para algum $\alpha_0 \in \mathbf{A}$, $M_t(\alpha_0)$ é válido para todo t .
2. Classe II: Para alguma sequência de valores $\alpha_t \in \mathbf{A}, (t = 1, 2, \dots)$, $M_t(\alpha_t)$ é válido para o tempo t .

Os modelos classe II não serão abordados neste texto por se considerar que os modelos classe I já são suficientes para fins de aproximação. Uma exposição sobre este tipo de modelo pode ser vista em (Yelland & Lee [40] e West & Harrison [39]). Agora pode-se introduzir a definição formal de modelos multiprocessos classe I:

Definição 3.2. Se $M_t(\alpha)$ é válido para algum $\alpha \in \mathbf{A}$ e para todo tempo t , onde $\mathbf{A} = \{\alpha_1, \dots, \alpha_N\}$, para $N \geq 1$, então a série $(Y_t)_{t \geq 1}$ segue um modelo de **multiprocasso classe I**.

Defina \mathbf{X}_t como sendo um vetor de variáveis aleatórias de interesse no tempo t , como o vetor de estados θ_t ou uma observação futura Y_{t+k} . Dado um valor particular $\alpha \in \mathbf{A}$, o DLM $M_t(\alpha)$ é analisado usualmente. A inferência sobre \mathbf{X}_t é feita pela densidade $f(\mathbf{X}_t | \alpha, \mathbf{D}_t)$. Inicia-se o processo com uma *priori* $f(\alpha | \mathbf{D}_0)$ e a informação é sequencialmente processada para prover inferências sobre α via a *posteriori* $f(\alpha | \mathbf{D}_t)$ no tempo t . Usa-se o Teorema de Bayes para a atualização:

$$f(\alpha | \mathbf{D}_t) \propto f(\alpha | \mathbf{D}_{t-1}) f(Y_t | \alpha, \mathbf{D}_{t-1}), \quad (3.9)$$

onde $f(Y_t | \alpha, \mathbf{D}_{t-1})$ é a densidade preditiva um passo à frente de $M_t(\alpha)$. A *posteriori* $f(\alpha | \mathbf{D}_t)$ dá informação sobre α , identificando os valores mais prováveis com base nos dados observados. Para se fazer inferência sobre a quantidade de interesse \mathbf{X}_t , basta marginalizar α , i.e., $f(\mathbf{X}_t | \mathbf{D}_t) = \sum_{\alpha \in \mathbf{A}} f(\mathbf{X}_t | \alpha, \mathbf{D}_t) f(\alpha | \mathbf{D}_t)$. Com base nesta definição, o processo de inferência sobre α é composto dos seguintes passos:

- Inicie o processo atribuindo *prioris* iniciais para cada $\alpha \in \mathbf{A}$. Geralmente faz-se $p_0(j) = f(\alpha_j | \mathbf{D}_0) = 1/N, (j = 1, \dots, N)$;
- Calcule a verossimilhança $l_t(j) = f(Y_t | \alpha_j, \mathbf{D}_{t-1})$ no tempo t ;
- Calcule a *posteriori* para o tempo t ,

$$f(\alpha_j | \mathbf{D}_t) = \frac{1}{c_t} f(\alpha_j | \mathbf{D}_{t-1}) f(Y_t | \alpha_j, \mathbf{D}_{t-1}), \quad (3.10)$$

onde

$$c_t = \sum_{j=1}^N f(\alpha_j | \mathbf{D}_{t-1}) f(Y_t | \alpha_j, \mathbf{D}_{t-1}). \quad (3.11)$$

Como se assume que α é discreto, as densidades marginais são

$$f(\mathbf{X}_t | \mathbf{D}_t) = \sum_{j=1}^N f(\mathbf{X}_t | \alpha_j, \mathbf{D}_t) f(\alpha_j | \mathbf{D}_t). \quad (3.12)$$

Note que as distribuições preditivas são misturas de distribuições normais ou *t-Student*, dependendo de como se faz inferência sobre a variância observacional V_t (vide Capítulo 2). Os momentos das densidades de misturas são calculados pela seguinte expressão, onde \mathbf{H}_t pode representar qualquer expressão de variáveis aleatórias de interesse:

$$E[\mathbf{H}_t | \mathbf{D}_t] = \sum_{j=1}^N E[\mathbf{H}_t | \alpha_j, \mathbf{D}_t] p(\alpha_j | \mathbf{D}_t). \quad (3.13)$$

As densidades marginais para os dois primeiros momentos são:

$$E[\mathbf{X}_t | \mathbf{D}_t] = \sum_{j=1}^N E[\mathbf{X}_t | \alpha_j, \mathbf{D}_t] p(\alpha_j | \mathbf{D}_t), \quad (3.14)$$

$$V[\mathbf{X}_t | \mathbf{D}_t] = \sum_{j=1}^N (V[\mathbf{X}_t | \alpha_j, \mathbf{D}_t] + (E[\mathbf{X}_t | \alpha_j, \mathbf{D}_t] - E[\mathbf{X}_t | \mathbf{D}_t])(E[\mathbf{X}_t | \alpha_j, \mathbf{D}_t] - E[\mathbf{X}_t | \mathbf{D}_t])') p(\alpha_j | \mathbf{D}_t). \quad (3.15)$$

O procedimento de mistura descrito acima pode ser usado para a identificação de modelo e estimação de parâmetros. As probabilidades *posterioris* em todo o espaço \mathbf{A} identificam os valores de α que são plausíveis para a série em questão. Isto serve para o propósito de encontrar o DLM, ou conjunto de DLMs, que seja mais apropriado para a série. Note que estas probabilidades podem mudar drasticamente ao longo do tempo, onde um DLM ou um conjunto pequeno de DLMs pode descrever melhor a série em

pontos diferentes do tempo. Note que todos os modelos são marginalizados para fazer a previsão em vez de utilizar o modelo mais provável. No entanto, caso o modelo adequado esteja presente no vetor de parâmetros \mathbf{A} , as probabilidades *a posteriori* podem convergir para zero com excessão de um valor no vetor \mathbf{A} , com a probabilidade deste valor tendendo a um. Isto serve para identificar o valor de α que melhor descreve a série após um certo período de tempo. Os valores finais de α são a estimativa dos parâmetros.

3.4 Métricas de Performance Preditiva

Após a definição do modelo e a estimação dos parâmetros, o próximo passo é analisar a sua performance. Existem na literatura várias métricas que permitem quantificar a precisão do modelo em relação aos dados observados (Flores [11], Armstrong & Collopy [3], Hyndman & Koehler [18], Makridakis & Hibon [26]). Entre as várias propostas, as principais e mais usadas na literatura são listadas abaixo, onde h denota o horizonte de previsão, Y_t o dado no tempo t , f_t o valor da previsão e $t \geq 1$.

- MAD (*Mean Absolute Deviation*) Erro Absoluto Médio, definida por:

$$MAD = \frac{1}{h} \sum_{t=1}^h |Y_t - f_t|. \quad (3.16)$$

- MSE (*Mean Square Error*) Erro Quadrático Médio, definida por:

$$MSE = \frac{1}{h} \sum_{t=1}^h (Y_t - f_t)^2. \quad (3.17)$$

- MAPE (*Mean Average Percentage Error*) Erro Percentual Médio, definida por:

$$MAPE = \frac{1}{h} \sum_{t=1}^h 100 \frac{|Y_t - f_t|}{Y_t}. \quad (3.18)$$

- SMAPE (*Symmetric Mean Absolute Percentage Error*) Erro Simétrico Percentual Médio, definida por:

$$SMAPE = \frac{1}{h} \sum_{t=1}^h 100 \frac{|Y_t - f_t|}{(Y_t + f_t)/2}. \quad (3.19)$$

- MedAPE (*Median Average Percentage Error*) Erro Percentual Mediano, definida por:

$$MedAPE = \text{mediana} \left(\frac{|Y_t - f_t|}{Y_t} \right), 1 \leq t \leq h. \quad (3.20)$$

- MASE (*Mean Average Scaled Error*) Erro Escalonado Médio, definido por:

$$MASE = \frac{1}{h} \sum_{t=1}^h 100 \frac{|Y_t - f_t|}{\frac{1}{h-1} \sum_{j=2}^h |Y_j - Y_{j-1}|}. \quad (3.21)$$

Estas medidas de desempenho podem ser classificadas em quatro grupos distintos (Hyndman & Koehler [18]):

- Métricas Dependentes de Escala (*scale-dependent metrics*), tais como MAD e MSE;
- Métricas de Erro Percentual (*percentage-error metrics*), tal como MAPE, MedAPE e SMAPE;
- Métricas de Erro Relativo (*relative-error metrics*), onde a média de razões entre um erro de um método e um erro de um método simples (*naive*), tal como MASE;
- Métricas de Erro Livre de Escala (*scale-free error metrics*): expressam cada erro como uma razão entre um erro médio de um método e o erro médio de um método base, usado para comparação.

Métricas de Erro Percentual tem a vantagem de serem independentes de escala, logo são frequentemente empregadas para comparar a performance de diferentes métodos com várias séries temporais distintas, tais como as séries da competição internacional M3. Medidas baseadas em erros percentuais tem a desvantagem de serem infinitas ou indefinidas se existem zeros nas séries. No entanto, isto não acontece em nenhuma das séries usadas na competição M3.

Métricas de erro relativo são também independentes de escala. Contudo, quando os erros são pequenos, o uso de métodos simples como *benchmark* não é mais possível, já que isso implicaria numa divisão por valores perto de zero. Métricas como MAPE também tem desvantagens, já que erros positivos são mais penalizados que erros negativos.

Isto levou a introdução da métrica SMAPE. No entanto, o SMAPE também tem limitações, já que ele pode ser negativo e portanto ter uma interpretação ambígua. O SMAPE só assume valores entre -200 e 200 , logo o erro é simétrico com relação a escala dos erros.

Neste trabalho adota-se as duas métricas mais empregadas na literatura e que foram também adotadas pela principal competição internacional do gênero, a competição M3. SMAPE e MedAPE são métricas livres de escala e portanto podem ser empregadas para comparação em um conjunto variado de séries com diferentes valores de escala. Além disso, SMAPE penaliza erros positivos e negativos da mesma forma se o erro for de baixa magnitude, veja (Goodwin & Lawton [14]). MedAPE é uma medida robusta a *outliers*, pois utiliza a mediana ao invés da média, uma vez que muitas vezes a distribuição dos erros absolutos pode ser assimétrica. Estas métricas já foram criticas na literatura, veja (Hyndman & Koehler [18] e Goodwin & Lawton [14]), mas para fins de comparação com os dados já publicados sobre a competição, foi decidido mantê-las, apesar das limitações.

Além de reportar as métricas acima, existem sugestões na literatura que consideram relevante calcular testes de significância estatística nos resultados. Por exemplo (Koninga *et al.* [25]) introduz uma nova metodologia que permite comparar cada método contra o melhor e contra a média dos métodos, de forma a avaliar se a acurácia dos diferentes métodos é significativamente melhor dentro de um certo nível de confiança. Como o enfoque deste trabalho é bayesiano, não foram empregados testes de significancia estatística nos resultados. Além disso, existe também na literatura trabalhos com opiniões contrárias ao uso de tais testes (Armstrong [2]), sendo o principal argumento contra baseado na falta de evidência empírica que suporte o uso destes testes de maneira geral. Cita-se que os testes muitas vezes são usados de maneira imprópria e com um certo viés para a hipótese que o pesquisador tenta provar. Como este assunto ainda está em discussão e nenhuma conclusão clara foi atingida, evitamos realizar testes de significância estatística nos resultados.

Capítulo 4

O Modelo *Damped Holt* com Amortecedor Dinâmico

Neste capítulo é apresentado o novo modelo polinomial de ordem dois com abafamento (*damping*) dinâmico da tendência, denominado *Damped Holt* dinâmico. As suas principais propriedades são apresentadas, assim como a estimação de parâmetros e algumas limitações do modelo, e também situações onde o modelo é adequado. A evolução do amortecedor é demonstrada e a sua dinâmica é discutida. O estudo particular de alguns exemplos será feito para mostrar algumas propriedades do modelo. O capítulo seguinte ilustra a performance preditiva na competição internacional M3 e em séries simuladas.

4.1 Motivação

O modelo *Damped Holt* foi introduzido como uma maneira de melhorar a performance preditiva do modelo aditivo *Holt* tradicional (veja definição 3.1). A inclusão de um amortecedor de tendência melhorou significativamente a performance preditiva do modelo em vários horizontes de previsão (Taylor [35], McKenzie & Gardner [27], Gardner & McKenzie [13]). Os dois principais modelos que possuem abafamento, *Damped Holt* para tendência aditiva e *Damped Pegels* para tendência multiplicativa, assumem que o parâmetro de abafamento é constante. No entanto, existem vários tipos de séries onde a tendência tende a mudar de magnitude ao longo do tempo. Ao se estimar um valor para o amortecedor de tendência, usa-se todo o conjunto de treinamento, e neste caso a natureza da série é tal que a estimação não é adequada. Nestes casos, é razoável supor que o abafamento da tendência, sobretudo para previsões de curto prazo, deva ser dinâmico. A

inclusão de uma dinâmica temporal permite, por exemplo, modelar séries onde a tendência cresce ao longo do tempo ou possui saltos de valor, de maneira similar a modelos com quebras estruturais. Note que existem pelo menos dois tipos de tendência, a linear e a exponencial. Na tendência linear a taxa de crescimento permanece praticamente constante e na tendência exponencial a taxa aumenta com a passagem do tempo.

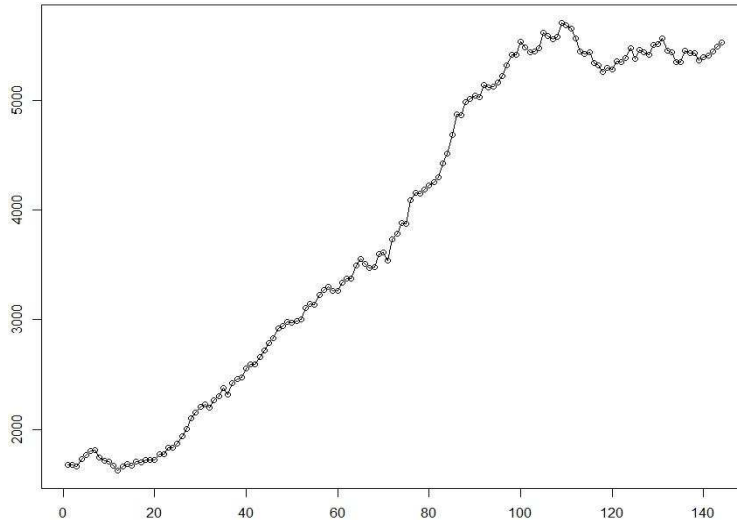


Figura 4.1: Exemplo de série da competição M3 (N2523) onde a tendência possui dinâmica temporal.

O exemplo da Figura 4.1 ilustra o problema de se usar um amortecedor fixo, pois a tendência foi praticamente anulada no final da série.

Várias séries exibem este tipo de comportamento e pode-se perceber que é vantajoso atribuir uma dinâmica no abafamento, principalmente para previsões de curto prazo. Em nosso levantamento bibliográfico, não existe proposta na literatura para atribuir uma dinâmica temporal no coeficiente de abafamento da tendência ψ . A seguir é apresentado a formulação do modelo proposto.

4.2 O Modelo Damped Holt Dinâmico

O modelo com abafamento dinâmico da tendência (em inglês, *Dynamic Damped Holt*) é apresentado na definição abaixo.

Definição 4.1. Para cada tempo t , o DLM univariado com amortecedor dinâmico é defi-

nido por

$$\begin{aligned} (Y_t | \boldsymbol{\theta}_t) &\sim N[\mathbf{F}_t' \boldsymbol{\theta}_t, V_t] \\ (\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \psi_{t-1}) &\sim N[\mathbf{G}_t \boldsymbol{\theta}_{t-1}, \mathbf{W}_t] \\ (\psi_t | \psi_{t-1}, \theta_{2,t}, \theta_{2,t-1}) &\sim N\left[\frac{\theta_{2,t-1}}{\theta_{2,t}} \psi_{t-1}, \frac{1}{\theta_{2,t}^2}\right], \end{aligned}$$

com priori inicial dada por

$$\begin{aligned} (\psi_0 | \boldsymbol{\theta}_0, \mathbf{D}_0) &\sim N\left[p_0, \frac{q_0}{\theta_{2,0}^2}\right] \\ (\boldsymbol{\theta}_0 | \mathbf{D}_0) &\sim N[\mathbf{m}_0, \mathbf{C}_0], \end{aligned}$$

onde

$$\mathbf{F}_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{G}_t = \begin{pmatrix} 1 & \psi_t \\ 0 & \psi_t \end{pmatrix}, \boldsymbol{\theta}_t = \begin{pmatrix} \theta_{1,t} \\ \theta_{2,t} \end{pmatrix},$$

assume-se que $\theta_{2,t} \neq 0$ e $(\psi_t | \theta_{2,t}, \mathbf{D}_t)$ não depende de $\theta_{1,t}$ para todo t .

Nesta definição, como usual, assume-se que os estados tem uma evolução *markoviana* e as observações são independentes entre si, dado o vetor de estados $\boldsymbol{\theta}_t$. Existem várias possibilidades para a evolução do amortecedor ψ , sendo uma das mais intuitivas uma razão entre os estados no tempo $t-1$ e t . Desta maneira obtém-se uma solução analítica, sem a necessidade de realizar simulações para obter uma amostra da *posteriori*. Além disso, a evolução do amortecedor ψ_t depende apenas de $\theta_{2,t-1}$ e $\theta_{2,t}$. Note que $\psi \in \Re$ e possui uma distribuição normal, ao invés da formulação usual onde $0 \leq \psi \leq 2$. A distribuição normal permite modelar a evolução do amortecedor de maneira conveniente e intuitiva. Neste caso, a razão entre os estados indica que conforme a inclinação $\theta_{2,t}$ aumenta, o abafamento da tendência se intensifica, o que é razoável supor. Caso a série decaia de nível o abafamento deve ser minimizado, de forma que o modelo possa acompanhar rapidamente a mudança de nível da série. Se a inclinação não se alterar ao longo do tempo, o valor do amortecedor no tempo t será o seu valor no tempo anterior $t-1$.

A informação sobre ψ_t está contida em $\theta_{2,t}$ e nos dados e não depende de $\theta_{1,t}$. Dado ψ_t , as atualizações do Teorema 2.2.4 continuam válidas.

Antes de apresentar a proposição que demonstra que a evolução dos estados perma-

nece normal, note que condicionado a \mathbf{D}_{t-1} temos que

$$f(\boldsymbol{\theta}_t, \boldsymbol{\psi}_{t-1} | \boldsymbol{\theta}_{t-1}, \mathbf{D}_{t-1}) = f(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \boldsymbol{\psi}_{t-1}, \mathbf{D}_{t-1}) f(\boldsymbol{\psi}_{t-1} | \boldsymbol{\theta}_{t-1}, \mathbf{D}_{t-1}) \quad (4.1)$$

$$= f(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \mathbf{D}_{t-1}) f(\boldsymbol{\psi}_{t-1} | \boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}, \mathbf{D}_{t-1}). \quad (4.2)$$

A variância observacional V_t pode ter uma evolução temporal ou ser fixa no tempo, como descrito na seção 2.2.7. De maneira similar, a especificação de \mathbf{W}_t pode ser feita por fatores de desconto, veja 2.2.9. Valores típicos para os hiperparâmetros de $\boldsymbol{\psi}$ são um para a média p_0 e a variância q_0 . A proposição a seguir mostra que a evolução dos estados ainda é normal após marginalizar $\boldsymbol{\psi}_t$.

Proposição 4.1. *Dado \mathbf{D}_{t-1} e condicionado a $\boldsymbol{\theta}_{t-1}$ temos que*

$$\boldsymbol{\psi}_{t-1} | \boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1} \sim N \left[\frac{d_{t-1}}{\boldsymbol{\theta}_{2,t-1}}, \frac{1}{s_{t-1} \boldsymbol{\theta}_{2,t-1}^2} \right]$$

$$\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1} \sim N [\mathbf{G}_t^* \boldsymbol{\theta}_{t-1}, \mathbf{W}_t^*],$$

onde

$$\mathbf{1}'_2 = (1, 1)$$

$$\mathbf{e}'_2 = (0, 1)$$

$$\mathbf{E}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$g_{t-1} = \frac{p_{t-1}}{s_{t-1} q_{t-1}}$$

$$s_{t-1} = \frac{1}{q_{t-1}} + \mathbf{1}'_2 \mathbf{W}_t^{-1} \mathbf{1}_2$$

$$\mathbf{H}_{t-1} = \mathbf{E}_1 - \frac{p_{t-1}}{s_{t-1} q_{t-1}} \mathbf{e}_2 \mathbf{1}'_2 = \begin{pmatrix} 1 & 0 \\ g_{t-1} & g_{t-1} \end{pmatrix}$$

$$d_{t-1} = \frac{1}{s_{t-1}} (\boldsymbol{\theta}_t - \mathbf{H}_{t-1} \boldsymbol{\theta}_{t-1})' \mathbf{W}_t^{-1} \mathbf{1}_2$$

$$\mathbf{G}_t^* = \begin{pmatrix} 1 & p_{t-1} \\ 0 & p_{t-1} \end{pmatrix}$$

$$\mathbf{W}_t^* = \mathbf{W}_t + \mathbf{1}_2 \mathbf{1}'_2 q_{t-1},$$

Prova: Como as duas distribuições são normais, a conjugação gera outras duas distribuições também normais. Basta se concentrar na expressão da parte exponencial da densi-

dade. Aplicando o log em ambos os lados da equação:

$$\log(f(\boldsymbol{\theta}_t, \boldsymbol{\psi}_{t-1} | \boldsymbol{\theta}_{t-1}, \mathbf{D}_{t-1})) \propto \left\{ -\frac{1}{2}(\boldsymbol{\theta}_t - \mathbf{G}_t \boldsymbol{\theta}_{t-1})' \mathbf{W}_t^{-1} (\boldsymbol{\theta}_t - \mathbf{G}_t \boldsymbol{\theta}_{t-1}) - \frac{1}{2} \frac{\boldsymbol{\theta}_{2,t-1}^2}{q_{t-1}} (\boldsymbol{\psi}_{t-1} - p_{t-1})^2 \right\}.$$

Por conveniência, considere que

$$\begin{aligned} \boldsymbol{\theta}_t - \mathbf{G}_t \boldsymbol{\theta}_{t-1} &= \begin{pmatrix} \boldsymbol{\theta}_{1,t} \\ \boldsymbol{\theta}_{2,t} \end{pmatrix} - \begin{pmatrix} 1 & \boldsymbol{\psi}_{t-1} \\ 0 & \boldsymbol{\psi}_{t-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}_{1,t-1} \\ \boldsymbol{\theta}_{2,t-1} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\theta}_{1,t} - \boldsymbol{\theta}_{1,t-1} \\ \boldsymbol{\theta}_{2,t} \end{pmatrix} - \boldsymbol{\theta}_{2,t-1} \boldsymbol{\psi}_{t-1} \mathbf{1}_2 \\ &= \mathbf{K} - \boldsymbol{\theta}_{2,t-1} \boldsymbol{\psi}_{t-1} \mathbf{1}_2, \end{aligned}$$

onde

$$\mathbf{K}' = (\boldsymbol{\theta}_{1,t} - \boldsymbol{\theta}_{1,t-1}, \boldsymbol{\theta}_{2,t})', \quad (4.3)$$

expandindo o quadrado

$$\begin{aligned} &\propto -\frac{1}{2} \left\{ \mathbf{K}' \mathbf{W}_t^{-1} \mathbf{K} - 2 \mathbf{K}' \mathbf{W}_t^{-1} \mathbf{1}_2 \boldsymbol{\theta}_{2,t-1} \boldsymbol{\psi}_{t-1} + \boldsymbol{\theta}_{2,t-1}^2 \boldsymbol{\psi}_{t-1}^2 \mathbf{1}_2' \mathbf{W}_t^{-1} \mathbf{1}_2 \right\} + \\ &\quad -\frac{1}{2} \left\{ \frac{\boldsymbol{\psi}_{t-1}^2 \boldsymbol{\theta}_{2,t-1}^2}{q_{t-1}} + \frac{p_{t-1}^2 \boldsymbol{\theta}_{2,t-1}^2}{q_{t-1}} - 2 \frac{p_{t-1} \boldsymbol{\theta}_{2,t-1} \boldsymbol{\psi}_{t-1}}{q_{t-1}} \right\}, \end{aligned}$$

a expressão acima pode ser reescrita como

$$\begin{aligned} &\propto -\frac{1}{2} \left\{ (\mathbf{K} - \boldsymbol{\theta}_{2,t-1} \boldsymbol{\psi}_{t-1} \mathbf{1}_2)' \mathbf{W}_t^{-1} (\mathbf{K} - \boldsymbol{\theta}_{2,t-1} \boldsymbol{\psi}_{t-1} \mathbf{1}_2) + \frac{\boldsymbol{\theta}_{2,t-1}^2}{q_{t-1}} (\boldsymbol{\psi}_{t-1} - p_{t-1})^2 \right\} \\ &\propto -\frac{1}{2} \left\{ \mathbf{K}' \mathbf{W}_t^{-1} \mathbf{K} + \left\{ \frac{1}{q_{t-1}} + \mathbf{1}_2' \mathbf{W}_t^{-1} \mathbf{1}_2 \right\} \boldsymbol{\theta}_{2,t-1}^2 \boldsymbol{\psi}_{t-1}^2 \right\} + \\ &\quad -\frac{1}{2} \left\{ -2 \boldsymbol{\psi}_{t-1} \boldsymbol{\theta}_{2,t-1} \left\{ \mathbf{K}' \mathbf{W}_t^{-1} \mathbf{1}_2 + \frac{p_{t-1} \boldsymbol{\theta}_{2,t-1}}{q_{t-1}} \right\} + p_{t-1}^2 \frac{\boldsymbol{\theta}_{2,t-1}^2}{q_{t-1}} \right\} \\ &\propto -\frac{1}{2} \left\{ \mathbf{K}' \mathbf{W}_t^{-1} \mathbf{K} + (\boldsymbol{\theta}_{2,t-1})^2 s_{t-1} \boldsymbol{\psi}_{t-1}^2 - 2 \boldsymbol{\psi}_{t-1} \boldsymbol{\theta}_{2,t-1} s_{t-1} d_{t-1} + p_{t-1}^2 \frac{\boldsymbol{\theta}_{2,t-1}^2}{q_{t-1}} + const \right\}, \end{aligned}$$

onde o termo *const* é um termo que representa constantes com relação a $\boldsymbol{\psi}_{t-1}$ e $\boldsymbol{\theta}_t$

$$\begin{aligned} &\propto -\frac{1}{2} \left\{ \mathbf{K}' \mathbf{W}_t^{-1} \mathbf{K} + (\boldsymbol{\theta}_{2,t-1})^2 s_{t-1} \left\{ \boldsymbol{\psi}_{t-1}^2 - 2 \frac{\boldsymbol{\psi}_{t-1} d_{t-1}}{\boldsymbol{\theta}_{2,t-1}} \right\} + p_{t-1}^2 \frac{\boldsymbol{\theta}_{2,t-1}^2}{q_{t-1}} + const \right\} \\ &\propto -\frac{1}{2} \left\{ \mathbf{K}' \mathbf{W}_t^{-1} \mathbf{K} + (\boldsymbol{\theta}_{2,t-1})^2 s_{t-1} \left\{ \boldsymbol{\psi}_{t-1} - \frac{d_{t-1}}{\boldsymbol{\theta}_{2,t-1}} \right\}^2 - d_{t-1}^2 s_{t-1} + p_{t-1}^2 \frac{\boldsymbol{\theta}_{2,t-1}^2}{q_{t-1}} + const \right\}, \end{aligned}$$

defina

$$D = (\boldsymbol{\theta}_{2,t-1})^2 s_{t-1} \left\{ \psi_{t-1} - \frac{d_{t-1}}{\boldsymbol{\theta}_{2,t-1}} \right\}^2,$$

expandindo o quadrado $(d_{t-1}s_{t-1})^2/s_{t-1}$:

$$\propto -\frac{1}{2} \left\{ D + \mathbf{K}' \left\{ \mathbf{W}_t^{-1} - \frac{\mathbf{W}_t^{-1} \mathbf{1}_2 \mathbf{1}_2' \mathbf{W}_t^{-1}}{s_{t-1}} \right\} \mathbf{K} - 2\mathbf{K}' \mathbf{W}_t^{-1} \mathbf{1}_2 \frac{p_{t-1} \boldsymbol{\theta}_{2,t-1}}{q_{t-1} s_{t-1}} - \frac{\boldsymbol{\theta}_{2,t-1}^2 p_{t-1}^2}{s_{t-1} q_{t-1}^2} + p_{t-1}^2 \frac{\boldsymbol{\theta}_{2,t-1}^2}{q_{t-1}} + const \right\},$$

calculando a inversa da matriz \mathbf{W}_t^* , temos

$$\begin{aligned} \mathbf{W}_t^* &= \left\{ \mathbf{W}_t^{-1} - \frac{\mathbf{W}_t^{-1} \mathbf{1}_2 \mathbf{1}_2' \mathbf{W}_t^{-1}}{s_{t-1}} \right\}^{-1} \\ &= \mathbf{W}_t + \mathbf{1}_2 (s_{t-1} - \mathbf{1}_2' \mathbf{W}_t^{-1} \mathbf{1}_2)^{-1} \mathbf{1}_2' \\ &= \mathbf{W}_t + \mathbf{1}_2 \mathbf{1}_2' q_{t-1}, \end{aligned}$$

de forma que

$$\begin{aligned} &\propto -\frac{1}{2} \left\{ D + \mathbf{K}' \{ \mathbf{W}_t^{*-1} \} \mathbf{K} - 2\mathbf{K}' \mathbf{W}_t^{*-1} (\mathbf{W}_t + \mathbf{1}_2 \mathbf{1}_2' q_{t-1}) \mathbf{W}_t^{-1} \mathbf{1}_2 \frac{p_{t-1} \boldsymbol{\theta}_{2,t-1}}{q_{t-1} s_{t-1}} \right\} + \\ &-\frac{1}{2} \left\{ -\frac{\boldsymbol{\theta}_{2,t-1}^2 p_{t-1}^2}{s_{t-1} q_{t-1}^2} + p_{t-1}^2 \frac{\boldsymbol{\theta}_{2,t-1}^2}{q_{t-1}} + const \right\}, \end{aligned}$$

note as seguintes identidades

$$\begin{aligned} 1 + \mathbf{1}_2' \mathbf{W}_t^{-1} \mathbf{1}_2 q_{t-1} &= 1 + q_{t-1} s_{t-1} - 1 = q_{t-1} s_{t-1} \\ \frac{1}{q_{t-1}} - \frac{1}{s_{t-1} q_{t-1}^2} &= \mathbf{1}_2' \mathbf{W}_t^{*-1} \mathbf{1}_2, \end{aligned}$$

substituindo as identidades acima, temos

$$\propto -\frac{1}{2} \left\{ D + \mathbf{K}' \{ \mathbf{W}_t^{*-1} \} \mathbf{K} - 2\mathbf{K}' \mathbf{W}_t^{*-1} \mathbf{1}_2 p_{t-1} \boldsymbol{\theta}_{2,t-1} + \boldsymbol{\theta}_{2,t-1}^2 p_{t-1}^2 \mathbf{1}_2' \mathbf{W}_t^{*-1} \mathbf{1}_2 + const \right\},$$

a expressão acima é reescrita como

$$(\boldsymbol{\theta}_{2,t-1})^2 s_{t-1} \left\{ \psi_{t-1} - \frac{d_{t-1}}{\boldsymbol{\theta}_{2,t-1}} \right\}^2 + (\boldsymbol{\theta}_t - \mathbf{G}_t^* \boldsymbol{\theta}_{t-1})' \mathbf{W}_t^{*-1} (\boldsymbol{\theta}_t - \mathbf{G}_t^* \boldsymbol{\theta}_{t-1}),$$

e a proposição está provada.

Esta proposição mostra que a marginal $(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1})$ ainda é normal, com média e variância dadas acima. A seguinte proposição demonstra a densidade de suavização de ψ_t .

Proposição 4.2. Dado \mathbf{D}_{t-1} e $\boldsymbol{\theta}_{2,t-1}$, se

$$\begin{aligned} (\boldsymbol{\theta}_{2,t} | \boldsymbol{\theta}_{2,t-1}, \mathbf{D}_{t-1}, \boldsymbol{\psi}_{t-1}) &\sim N[\boldsymbol{\psi}_{t-1} \boldsymbol{\theta}_{2,t-1}, w_{22,t}] \\ (\boldsymbol{\psi}_{t-1} | \boldsymbol{\theta}_{2,t-1}, \mathbf{D}_{t-1}) &\sim N \left[p_{t-1}, \frac{q_{t-1}}{\boldsymbol{\theta}_{2,t-1}^2} \right], \end{aligned}$$

então

$$(\boldsymbol{\psi}_{t-1} | \boldsymbol{\theta}_{2,t}, \boldsymbol{\theta}_{2,t-1}, \mathbf{D}_{t-1}) \sim N \left[r_t + \frac{\boldsymbol{\theta}_{2,t}}{\boldsymbol{\theta}_{2,t-1}} u_t, \frac{k_t}{\boldsymbol{\theta}_{2,t-1}^2} \right],$$

onde

$$\begin{aligned} w_{22,t}^* &= w_{22,t} + q_{t-1} \\ r_t &= p_{t-1}(1 - u_t), \quad u_t = \frac{q_{t-1}}{w_{22,t}^*} \quad \text{e} \quad k_t = q_{t-1}(1 - u_t). \end{aligned}$$

Prova: A prova usa o Lema A.2 de reconstrução da distribuição conjunta, de forma que

$$\left(\begin{array}{c} \boldsymbol{\theta}_{2,t} \\ \boldsymbol{\psi}_{t-1} \end{array} \middle| \boldsymbol{\theta}_{2,t-1}, \mathbf{D}_{t-1} \right) \sim N \left[\left(\begin{array}{c} p_{t-1} \boldsymbol{\theta}_{2,t-1} \\ p_{t-1} \end{array} \right), \left(\begin{array}{cc} w_{22,t} + q_{t-1} & \frac{q_{t-1}}{\boldsymbol{\theta}_{2,t-1}} \\ \frac{q_{t-1}}{\boldsymbol{\theta}_{2,t-1}} & \frac{q_{t-1}}{\boldsymbol{\theta}_{2,t-1}^2} \end{array} \right) \right],$$

logo a distribuição $(\boldsymbol{\psi}_{t-1} | \boldsymbol{\theta}_{2,t}, \boldsymbol{\theta}_{2,t-1}, \mathbf{D}_{t-1})$ tem momentos:

$$\begin{aligned} E[\boldsymbol{\psi}_{t-1} | \boldsymbol{\theta}_{2,t}, \boldsymbol{\theta}_{2,t-1}, \mathbf{D}_{t-1}] &= p_{t-1} + \frac{q_{t-1}}{\boldsymbol{\theta}_{2,t-1} w_{22,t}^*} (\boldsymbol{\theta}_{2,t} - p_{t-1} \boldsymbol{\theta}_{2,t-1}) \\ &= p_{t-1} \left(1 - \frac{q_{t-1}}{w_{22,t}^*} \right) + \frac{\boldsymbol{\theta}_{2,t}}{\boldsymbol{\theta}_{2,t-1}} \frac{q_{t-1}}{w_{22,t}^*} \\ V[\boldsymbol{\psi}_{t-1} | \boldsymbol{\theta}_{2,t}, \boldsymbol{\theta}_{2,t-1}, \mathbf{D}_{t-1}] &= \frac{q_{t-1}}{\boldsymbol{\theta}_{2,t-1}^2} - \left(\frac{q_{t-1}}{\boldsymbol{\theta}_{2,t-1}} \right)^2 \frac{1}{w_{22,t}^*}. \end{aligned}$$

Usando os resultados da proposição acima e a priori $\boldsymbol{\theta}_{2,t-1} | \mathbf{D}_{t-1}$ constrói-se a proposição do suavização do segundo componente do vetor $\boldsymbol{\theta}_t$.

Proposição 4.3. Dado \mathbf{D}_{t-1} , se

$$\begin{aligned} (\boldsymbol{\theta}_{2,t} | \boldsymbol{\theta}_{2,t-1}, \mathbf{D}_{t-1}) &\sim N[p_{t-1} \boldsymbol{\theta}_{2,t-1}, w_{22,t}^*] \\ (\boldsymbol{\theta}_{2,t-1} | \mathbf{D}_{t-1}) &\sim N[m_{2,t-1}, c_{22,t-1}], \end{aligned}$$

então

$$(\theta_{2,t-1} | \theta_{2,t}, \mathbf{D}_{t-1}) \sim N[l_t, h_t],$$

onde

$$\begin{aligned} l_t &= m_{2,t-1} + b_t(\theta_{2,t} - p_{t-1}m_{2,t-1}) \\ h_t &= c_{22,t-1} - b_t^2 x_t \\ b_t &= \frac{p_{t-1}c_{22,t-1}}{x_t} \text{ e } x_t = w_{22,t}^* + p_{t-1}^2 c_{22,t-1}. \end{aligned}$$

*Prova:*Primeiro, com a lei da variância total, temos que

$$\begin{aligned} V[\theta_{2,t} | \mathbf{D}_{t-1}] &= E[V[\theta_{2,t} | \theta_{2,t-1}, \mathbf{D}_{t-1}]] + V[E[\theta_{2,t} | \theta_{2,t-1}, \mathbf{D}_{t-1}]] \\ &= w_{22,t}^* + p_{t-1}^2 c_{22,t-1} \\ Cov[\theta_{2,t}, \theta_{2,t-1} | \mathbf{D}_{t-1}] &= Cov[p_{t-1}\theta_{2,t-1} + w_{22,t}^*, \theta_{2,t-1}] = p_{t-1}c_{22,t-1}. \end{aligned}$$

Usando o Lema A.2 constrói-se a densidade conjunta dos estados em $t-1$ e t , dada por

$$\begin{pmatrix} \theta_{2,t} \\ \theta_{2,t-1} \end{pmatrix} | \mathbf{D}_{t-1} \sim N \left[\begin{pmatrix} p_{t-1}m_{2,t-1} \\ m_{2,t-1} \end{pmatrix}, \begin{pmatrix} w_{22,t} + p_{t-1}^2 c_{22,t-1} & p_{t-1}c_{22,t-1} \\ p_{t-1}c_{22,t-1} & c_{22,t-1} \end{pmatrix} \right],$$

com base na densidade conjunta acima, calcula-se os momentos da densidade condicional $(\theta_{2,t-1} | \theta_{2,t}, \mathbf{D}_{t-1})$:

$$\begin{aligned} E[\theta_{2,t-1} | \theta_{2,t}, \mathbf{D}_{t-1}] &= m_{2,t-1} + p_{t-1}c_{22,t-1}x_t^{-1}(\theta_{2,t} - p_{t-1}m_{2,t-1}) \\ V[\theta_{2,t-1} | \theta_{2,t}, \mathbf{D}_{t-1}] &= c_{22,t-1} - \frac{p_{t-1}^2 c_{22,t-1}^2}{x_t}, \end{aligned}$$

Pode-se reconhecer estes momentos como l_t e h_t , respectivamente, dados acima.

O resultado da Proposição 4.2 junto com a proposta da evolução do amortecedor de tendência ψ são usados para demonstrar a proposição abaixo. Ela apresenta a evolução temporal do amortecedor ψ_t com base no segundo componente do vetor de estados θ_2 nos tempos t e $t-1$. Note que a evolução do amortecedor depende apenas de $\theta_{2,t}$, $\theta_{2,t-1}$ e \mathbf{D}_{t-1} .

Proposição 4.4. Dado $\theta_{2,t-1}$, $\theta_{2,t}$ e \mathbf{D}_{t-1} , e usando os resultados da Proposição 4.2, se

$$\begin{aligned} (\psi_{t-1} | \theta_{2,t}, \theta_{2,t-1}) &\sim N \left[r_t + \frac{\theta_{2,t}}{\theta_{2,t-1}} u_t, \frac{k_t}{\theta_{2,t-1}^2} \right] \\ (\psi_t | \psi_{t-1}, \theta_{2,t}, \theta_{2,t-1}) &\sim N \left[\frac{\theta_{2,t-1}}{\theta_{2,t}} \psi_{t-1}, \frac{1}{\theta_{2,t}^2} \right], \end{aligned}$$

então

$$(\psi_t | \theta_{2,t}, \theta_{2,t-1}) \sim N[p_t, z_t],$$

onde

$$\begin{aligned} r_t &= p_{t-1}(1 - u_t) \\ p_t &= u_t + \frac{\theta_{2,t-1}}{\theta_{2,t}} r_t \\ z_t &= \frac{k_t + 1}{\theta_{2,t}^2}. \end{aligned}$$

Prova: A prova é feita simplesmente aplicando o Lema A.1 nas duas densidades acima e o resultado é imediato.

A última proposição a ser apresentada permite determinar a densidade de ψ_t dado $\theta_{2,t}$ e \mathbf{D}_{t-1} .

Proposição 4.5. *Dado $\theta_{2,t}$ e \mathbf{D}_{t-1} , e usando os resultados das Proposições 4.4 e 4.3, se*

$$\begin{aligned} (\psi_t | \theta_{2,t}, \theta_{2,t-1}, \mathbf{D}_{t-1}) &\sim N \left[u_t + \frac{\theta_{2,t-1}}{\theta_{2,t}} r_t, \frac{k_t + 1}{\theta_{2,t}^2} \right] \\ (\theta_{2,t-1} | \theta_{2,t}, \mathbf{D}_{t-1}) &\sim N[m_{2,t-1} + b_t(\theta_{2,t} - p_{t-1}m_{2,t-1}), c_{22,t-1} - b_t^2 x_t], \end{aligned}$$

então

$$(\psi_t | \theta_{2,t}, \mathbf{D}_{t-1}) \sim N \left[n_t, \frac{q_t^*}{\theta_{2,t}^2} \right],$$

onde

$$\begin{aligned} u_t &= \frac{q_{t-1}}{w_{22,t}^*} \\ x_t &= w_{22,t}^* + p_{t-1}^2 c_{22,t-1} \\ k_t &= q_{t-1}(1 - u_t) \\ r_t &= p_{t-1}(1 - u_t) \\ b_t &= \frac{p_{t-1} c_{22,t-1}}{x_t} \\ n_t &= u_t + \frac{r_t m_{2,t-1}}{\theta_{2,t}} (1 - p_{t-1} b_t) + r_t b_t \\ h_t &= c_{22,t} - b_t^2 x_t \\ q_t^* &= 1 + k_t + r_t^2 h_t. \end{aligned}$$

Prova: Como a relação entre $\boldsymbol{\psi}_t$ e $\boldsymbol{\theta}_{2,t-1}$ é linear na média, basta invocar o Lema A.1 e a prova é imediata.

Como a previsão de Y_t depende apenas de $\boldsymbol{\theta}_t$, existem infinitas formulações para $(Y_t|\boldsymbol{\psi}_{t-1}, \boldsymbol{\theta}_{t-1}, \mathbf{D}_{t-1})$ que geram a densidade da equação de observação $(Y_t|\boldsymbol{\theta}_t) \sim N[\mathbf{F}'_t\boldsymbol{\theta}_t, V_t]$. Várias combinações diferentes de $\boldsymbol{\psi}_{t-1}$ e $\boldsymbol{\theta}_{2,t-1}$ podem gerar o mesmo valor de $\boldsymbol{\theta}_{2,t}$, ou seja, o sistema é não identificável. Além disso, para que as equações de atualização do filtro e de previsão tenham solução analítica, é importante que a densidade $(Y_t|\boldsymbol{\psi}_{t-1}, \boldsymbol{\theta}_{t-1}, \mathbf{D}_{t-1})$ seja tal que a *priori* tenha a forma dada abaixo:

$$(\boldsymbol{\psi}_t|\boldsymbol{\theta}_{2,t}, \mathbf{D}_t) \sim N[p_t, q_t/\boldsymbol{\theta}_{2,t}^2],$$

onde p_t e q_t são constantes em relação a $\boldsymbol{\theta}_t$.

Da Proposição 4.5, temos que

$$(\boldsymbol{\psi}_t|\boldsymbol{\theta}_{2,t}, \mathbf{D}_{t-1}) \sim N\left[n_t, \frac{q_t^*}{\boldsymbol{\theta}_{2,t}^2}\right],$$

Como o sistema é não identificável, a nova *priori* para $\boldsymbol{\psi}_t|\boldsymbol{\theta}_{2,t}, \mathbf{D}_t$ só pode ser dada por sugestões. Várias possibilidades existem, e dado o resultado da Proposição 4.5, uma sugestão plausível é

$$(\boldsymbol{\psi}_t|\boldsymbol{\theta}_{2,t}, \mathbf{D}_t) \sim N\left[p_t, \frac{q_t}{\boldsymbol{\theta}_{2,t}^2}\right],$$

onde

$$\begin{aligned} p_t &= u_t + r_t \frac{m_{2,t-1}}{m_{2,t}} (1 - p_{t-1}b_t) + r_t b_t \\ q_t &= \lambda_t q_t^* = \lambda_t (1 + k_t + r_t^2 h_t), \end{aligned}$$

onde $\lambda_t \in (0, 1)$ é um valor definido pelo usuário e representa o ganho de informação ao se receber a nova observação Y_t , pois espera-se que a variância q_t diminua devido a este ganho. Esta solução permite resolver o problema de forma analítica. Note que a sugestão simplesmente substitui $\boldsymbol{\theta}_{2,t}$ pela sua respectiva média $m_{2,t}$ no resultado da Proposição 4.4. Analisando a expressão para a densidade, nota-se que a média é função do valor da média no tempo anterior, p_{t-1} , e da razão $\boldsymbol{\theta}_{2,t-1}/\boldsymbol{\theta}_{2,t}$, ponderado pelo coeficiente adaptativo u_t (Note que $0 \leq u_t \leq 1$). Pode-se perceber que em cada tempo t , é feita uma nova proposta de *priori* devido ao fato do sistema ser não identificável.

A desvantagem desta proposta é a incapacidade desta evolução lidar com valores de

$\theta_{2,t}$ próximos de zero. Caso isso aconteça, o valor é descartado e o valor válido mais recente é tomado. Define-se valor válido como o valor gerado de ψ que esteja na faixa de valores $(0, 2)$. Este procedimento é válido do ponto de vista bayesiano e evita instabilidades no vetor de estados, como discutido na seção 4.5.

A maneira como o amortecedor evolui no tempo será discutida nas próximas seções. A tabela abaixo resume o modelo proposto e as equações de atualização para a evolução do sistema.

Tabela 4.1: Resumo das Equações do *Damped Holt* Dinâmico

| Modelo <i>Damped Holt</i> com abafamento dinâmico e V_t e W_t conhecidos |
|---|
| Equações de Observação e de Sistema |
| $(Y_t \boldsymbol{\theta}_t) \sim N[\mathbf{F}'_t \boldsymbol{\theta}_t, V_t]$ $(\boldsymbol{\theta}_t \boldsymbol{\theta}_{t-1}, \boldsymbol{\psi}_{t-1}) \sim N[\mathbf{G}_t \boldsymbol{\theta}_{t-1}, \mathbf{W}_t]$ $(\boldsymbol{\theta}_t \boldsymbol{\theta}_{t-1}) \sim N[\mathbf{G}_t^* \boldsymbol{\theta}_{t-1}, \mathbf{W}_t^*]$ |
| onde |
| $\mathbf{1}'_2 = (1, 1), \mathbf{F}_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{G}_t = \begin{pmatrix} 1 & \boldsymbol{\psi}_{t-1} \\ 0 & \boldsymbol{\psi}_{t-1} \end{pmatrix}, \mathbf{G}_t^* = \begin{pmatrix} 1 & p_{t-1} \\ 0 & p_{t-1} \end{pmatrix}, \mathbf{W}_t^* = \mathbf{W}_t + \mathbf{1}_2 \mathbf{1}'_2 q_{t-1},$ |
| Prioris Iniciais |
| $(\boldsymbol{\psi}_0 \boldsymbol{\theta}_0, \mathbf{D}_0) \sim N \left[p_0, \frac{q_0}{\boldsymbol{\theta}_{2,0}^2} \right]$ $(\boldsymbol{\theta}_0 \mathbf{D}_0) \sim N[\mathbf{m}_0, \mathbf{C}_0],$ |
| Equações de atualização |
| $(\boldsymbol{\psi}_t \boldsymbol{\theta}_{2,t}, \mathbf{D}_t) \sim N \left[p_t, \frac{q_t}{\boldsymbol{\theta}_{2,t}^2} \right],$ $(\boldsymbol{\psi}_t \boldsymbol{\psi}_{t-1}, \boldsymbol{\theta}_{2,t}, \boldsymbol{\theta}_{2,t-1}) \sim N \left[\frac{\boldsymbol{\theta}_{2,t-1}}{\boldsymbol{\theta}_{2,t}} \boldsymbol{\psi}_{t-1}, \frac{1}{\boldsymbol{\theta}_{2,t}^2} \right],$ |
| onde |
| $p_t = u_t + r_t \frac{m_{2,t-1}}{m_{2,t}} (1 - p_{t-1} b_t) + r_t b_t$ $q_t = \lambda_t (1 + k_t + r_t^2 h_t)$ $u_t = \frac{q_{t-1}}{w_{22,t}^*}$ $k_t = q_{t-1} (1 - u_t)$ $r_t = p_{t-1} (1 - u_t)$ $x_t = w_{22,t}^* + p_{t-1}^2 c_{22,t-1}$ $b_t = \frac{p_{t-1} c_{22,t-1}}{x_t}$ $h_t = c_{22,t} - b_t^2 x_t$ |

4.3 A Evolução do Amortecedor

A evolução temporal do amortecedor, da maneira como é apresentada pela Proposição 4.2, possui oscilações muito grandes na sua média. Isto é devido sobretudo ao coeficiente adaptativo u_t ser alto e a razão entre os estados se alterar muito com a passagem do tempo. O gráfico abaixo ilustra o problema com uma série da competição M3 (N2533). A série é mostrada no primeiro painel. No segundo estão os valores do segundo componente do vetor de estados ($\theta_{2,t}$). No terceiro painel o valor da média de ψ_t é mostrado. No eixo horizontal está o índice de tempo.

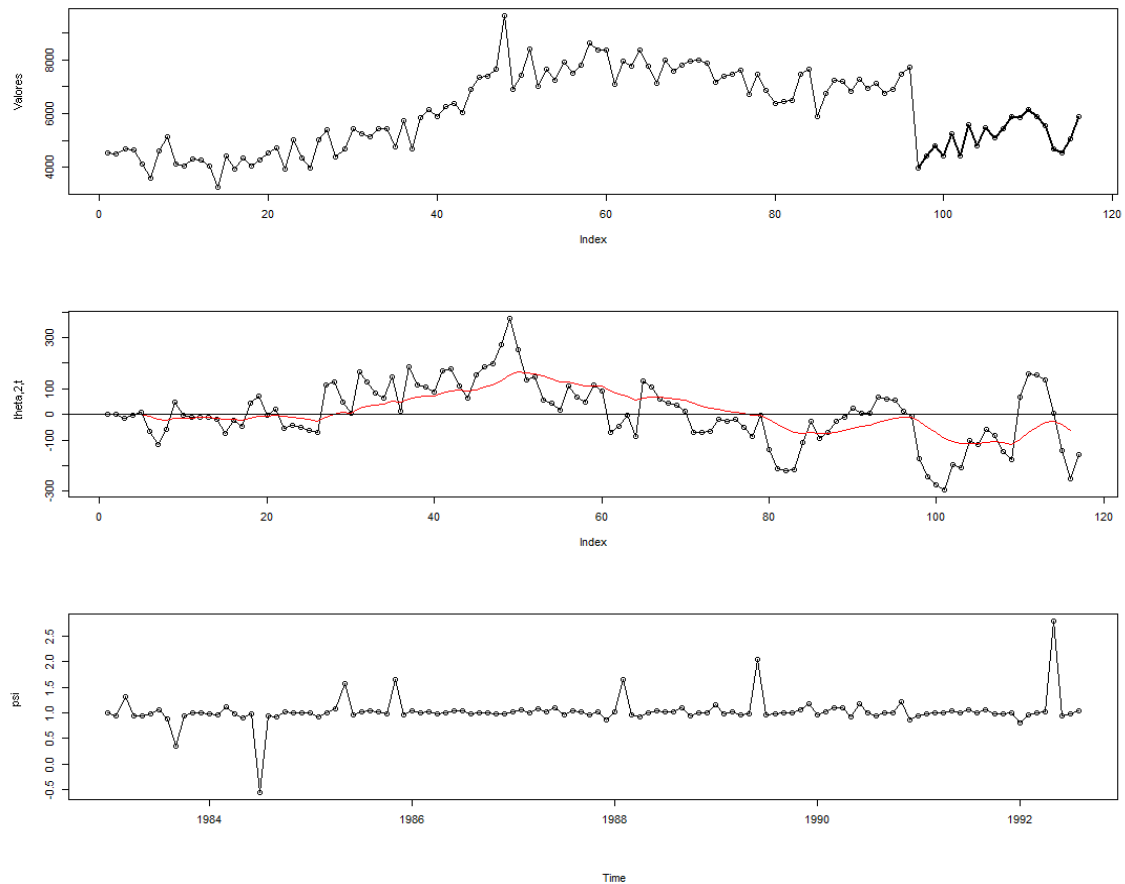


Figura 4.2: Série da competição M3 (N2533) ilustrando o problema da dinâmica do amortecedor.

Uma maneira de solucionar este problema e evitar saltos abruptos no valor da média é diminuir o peso dado a razão entre os estados ($\theta_{2,t-1}/\theta_{2,t}$), de forma que a evolução da média ao longo do tempo seja suavizada. Analisando o coeficiente adaptativo u_t , temos que

$$u_t = \frac{q_{t-1}}{w_{22,t}^*} = \frac{q_{t-1}}{w_{22,t} + q_{t-1}}. \quad (4.4)$$

Note que se a variância $w_{22,t}$ é próxima de zero, então u_t se aproxima de um, logo a evolução é dada principalmente pela razão entre os estados. Agora se $w_{22,t}$ aumenta de magnitude em relação a variância de ψ_t , o valor imediatamente anterior da média de ψ_t , p_{t-1} tem peso maior que a razão entre os estados. Note também que $0 \leq u_t \leq 1$, de forma que u_t serve como uma espécie de peso entre o valor da média de ψ_t e a razão entre os estados. Isto fica mais claro ao analisar a expressão para a média de $(\psi_{t-1} | \theta_{2,t}, \theta_{2,t-1})$, dada por

$$p_{t-1}(1 - u_t) + \frac{\theta_{2,t}}{\theta_{2,t-1}} u_t. \quad (4.5)$$

Uma maneira de controlar a dinâmica de ψ_t é tornar o coeficiente adaptativo u_t constante, de forma que saltos abruptos na razão entre os estados não influenciem muito na média. Temos que u_t é dado por

$$u_t = \frac{q_{t-1}}{w_{22,t}^*} = \frac{\lambda_{t-1} q_{t-1}^*}{\lambda_{t-1} q_{t-1}^* + w_{22,t}}, \quad (4.6)$$

então existem infinitas combinações de $w_{22,t}$ e λ_t que permitem o mesmo valor de u_t . Agora temos que u_t é fixo e não possui mais nenhuma dinâmica temporal ($u_t = u$ para todo t), o seu valor sendo especificado como os outros parâmetros. Isto implica em uma influência direta em W_t^* :

$$\mathbf{W}_t^* = \mathbf{W}_t + \mathbf{1}_2 \mathbf{1}'_2 q_{t-1} = \begin{pmatrix} w_{11,t} & w_{12,t} \\ w_{12,t} & q_{t-1} \frac{1-u_t}{u_t} \end{pmatrix} + \mathbf{1}_2 \mathbf{1}'_2 q_{t-1} \quad (4.7)$$

É interessante que W_t^* seja dada através de fator de descontos, em conformidade com a abordagem bayesiana. A estratégia usual de fatores de descontos produz a seguinte identidade

$$\begin{aligned} \mathbf{W}_t^* &= \mathbf{G}_t^* \mathbf{C}_{t-1} \mathbf{G}_t^{*'} \left(\frac{1}{\delta} - 1 \right) \\ &= \begin{pmatrix} 1 & p_{t-1} \\ 0 & p_{t-1} \end{pmatrix} \begin{pmatrix} c_{11,t-1} & c_{12,t-1} \\ c_{12,t-1} & c_{22,t-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ p_{t-1} & p_{t-1} \end{pmatrix} \left(\frac{1}{\delta} - 1 \right) \\ &= p_{t-1}^2 c_{22,t-1} \mathbf{1}_2 \mathbf{1}'_2 \left(\frac{1}{\delta} - 1 \right) \\ &+ \begin{pmatrix} 2p_{t-1} c_{12,t-1} + c_{11,t-1} & p_{t-1} c_{12,t-1} \\ p_{t-1} c_{12,t-1} & 0 \end{pmatrix} \left(\frac{1}{\delta} - 1 \right) \end{aligned} \quad (4.8)$$

Agora basta verificar quais valores de W_t^* e λ_t tornam as expressões 4.7 e 4.8 iguais com u_t e δ fixados. Agora tem-se que existe apenas um único λ_t para cada u_t e δ fixos,

como descrito na equação abaixo.

$$\lambda_t q_{t-1}^* \frac{1-u_t}{u_t} + \lambda_t q_{t-1}^* = p_{t-1}^2 c_{22,t-1} \left(\frac{1}{\delta} - 1 \right) \quad (4.10)$$

de forma que

$$\lambda_t = \frac{u_t}{q_{t-1}^*} p_{t-1}^2 c_{22,t-1} \left(\frac{1}{\delta} - 1 \right) \quad (4.11)$$

No entanto, o valor de λ_t calculado acima pode ser tal que não exista $w_{11,t} > 0$ que permita a seguinte igualdade

$$w_{11,t} + \lambda_t q_{t-1}^* = (p_{t-1}^2 c_{22,t-1} + 2p_{t-1} c_{12,t-1} + c_{11,t-1}) \left(\frac{1}{\delta} - 1 \right) \quad (4.12)$$

Ainda pode acontecer que $\lambda_t q_{t-1}^*$ seja muito maior que o lado direito da equação acima e caso isso aconteça, tem-se que $w_{11,t}$ teria de ser negativo para tornar a igualdade acima válida. Para solucionar isto, sugere-se proceder da seguinte maneira:

1. Após atualizar o tempo t , fixe δ e u_t com valores pré-definidos;
2. Calcule λ_t , dado por 4.11;
3. Calcule $w_{11,t}$, dado por 4.12. Se $w_{11,t} > 0$, prossiga. Se não for, aumente o fator de desconto δ e volte para o segundo passo;
4. Atualize a variância de $(\psi_t | \theta_{2,t}, \mathbf{D}_t)$, dada por $q_t = \lambda_t q_t^*$;
5. Prossiga com as atualizações restantes;

Ao realizar os testes com esta formulação, notou-se que valores típicos de u_t são na faixa (0.01,0.99), onde valores acima de 0.6 em geral não são significativos, i.e., o efeito na série é o mesmo que se teria se o valor fosse 0.6. Em geral, quando $u_t = 0.5$ a série se comporta como o modelo polinomial de ordem dois tradicional. A figura abaixo ilustra o efeito de diferentes valores de u_t na previsão um passo a frente para uma série da competição M3. Note que conforme u_t diminui de valor, o modelo tende a apresentar maior variabilidade dos estados e a seguir as observações de maneira mais acentuada.

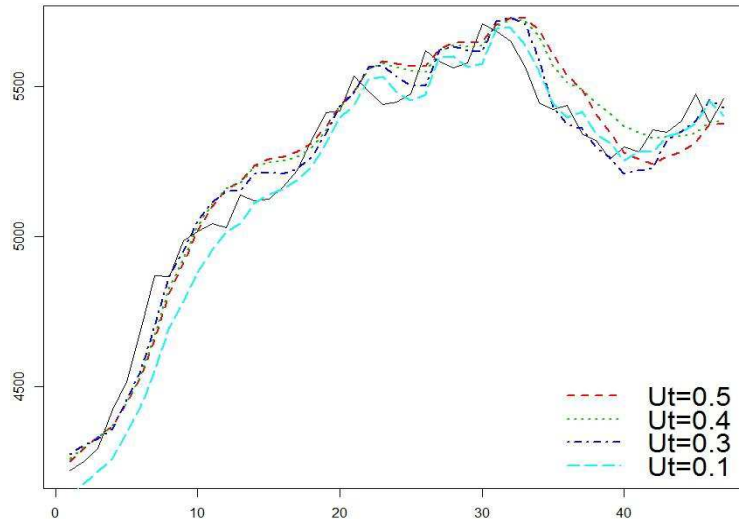


Figura 4.3: Série da competição M3 (N2525) ilustrando a influência de diferentes valores de u_t .

Note como o coeficiente adaptativo u_t influencia na variabilidade do amortecedor ψ_t , como mostra a Figura 4.4 abaixo. Pode-se perceber que conforme u_t diminui a variabilidade do amortecedor aumenta.

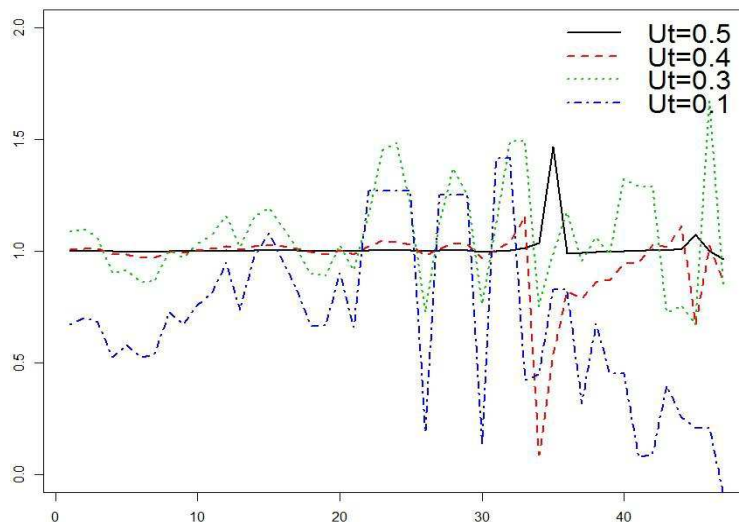


Figura 4.4: Valores da média de ψ_t para a série da competição M3 (N2525) ilustrando a influência de u_t .

A taxa de aumento de variabilidade depende da natureza da série. A influência de um ψ_t dinâmico é realizada pela evolução do segundo componente do vetor de estados $\theta_{2,t}$, como ilustra a Figura 4.5 abaixo. Note como a variabilidade do estado aumenta conforme u_t diminui, devido ao amortecedor dinâmico.

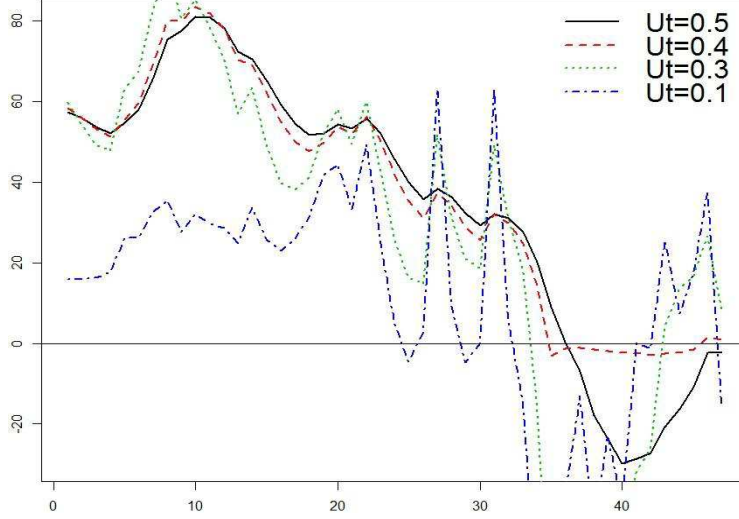


Figura 4.5: Valores da média de $\theta_{2,t}$ para a série da competição M3 (N2525) ilustrando a influência de u_t .

Relembrando a expressão para a média de $(\psi_t | \theta_{2,t}, \mathbf{D}_t)$, temos que

$$p_t = u_t + p_{t-1}(1-u_t)(1-p_{t-1}b_t) \frac{m_{2,t-1}}{m_{2,t}} + p_{t-1}(1-u_t)b_t$$

$$p_t = u_t + (1-u_t) \left(\frac{p_{t-1}}{w_{22,t}^* + p_{t-1}^2 c_{22,t-1}} \left(p_{t-1} c_{22,t-1} + \frac{m_{2,t-1}}{m_{2,t}} w_{22,t}^* \right) \right). \quad (4.13)$$

Note que a média p_t é uma função linear da razão entre os estados com o peso dado pela variância do estado $\theta_{2,t}$, i.e., $w_{22,t}^*$.

4.4 Estimação de Parâmetros

A quádrupla $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$ define completamente o DLM, condicionado a ψ_t . O novo modelo sugerido possui $\mathbf{F}_t = \mathbf{F} = (1, 0)'$ fixo para todo tempo t . A matriz \mathbf{G}_t depende apenas de ψ_{t-1} e portanto é totalmente especificada, mas com dinâmica temporal. Note que é necessário definir um valor para u_t , como discutido na seção anterior. O coeficiente adaptativo u_t tipicamente toma valores no intervalo $(0.01, 0.99)$, onde valores próximos de zero dão maior peso a razão entre os estados $\theta_{2,t-1}/\theta_{2,t}$ e valores próximos de um dão peso maior ao valor imediatamente anterior ψ_{t-1} .

A metodologia sugerida para estimar o coeficiente adaptativo u_t é escolher o valor que minimiza a métrica SMAPE no conjunto de dados de treinamento. Este procedimento é simples e teve resultados encorajadores nos dados referentes à competição internacional

M3, como descrito no próximo capítulo. Outra metodologia sugerida é empregar um conjunto de modelos, cada um com a mesma estrutura, mas com valores de u_t diferentes. Neste caso usa-se multiprocessos classe I para estimar o DLM que possui o u_t mais adequado para a série (para maiores detalhes, veja a seção (3.3 e West & Harrison [39]). Dado o valor para u_t e os dois primeiros momentos para a *priori* ($\psi_0|\theta_0, D_0$), a saber, a média p_0 e a variância q_0 , temos a evolução de ψ_t completamente especificada. Caso não se queira especificar u_t , uma opção é usar a formulação original, i.e., $u_t = q_{t-1}/w_{22,t}^*$. A desvantagem ao se adotar esta opção é a possibilidade de alterações bruscas no valor de ψ_t . Na prática, estes saltos de valor não são frequentes e não alteram muito a previsão da série, mas em algumas situações a alteração é de grande magnitude e pode gerar instabilidade no vetor de estados e portanto invalidar as previsões a partir do momento em que ocorre a alteração brusca de valor, veja a Figura 4.6.

Resta apenas especificar a variância observacional V_t e a variância de evolução \mathbf{W}_t . Ambas podem ser estimadas por métodos frequentistas, como a maximização da função de verossimilhança (Petris *et al.* [29]). Como o enfoque deste trabalho é a metodologia bayesiana, adotou-se a estimação das variâncias por meio de aprendizagem, como descrito nas seções 2.2.8 e 2.2.9. Neste caso, a variância observacional é definida como uma distribuição gama inversa com evolução temporal. Ao invés de especificar a matriz \mathbf{W}_t , usa-se fatores de desconto. Este fator expressa o ganho de variabilidade, ao se passar de $t-1$ para o tempo t , devido ao componente de erro estocástico presente na evolução dos estados (veja seção 2.2.9). Esta interpretação permanece válida no novo modelo proposto. Neste trabalho, a especificação de fatores de desconto é feita pela minimização da medida SMAPE (veja, 3.4) em um conjunto de dados de treinamento. Desta forma encontra-se o fator que melhor se adequa a série em estudo de maneira rápida e simples. Outra opção é utilizar multiprocessos classe I, com uma faixa de valores para o fator de desconto. Tem-se um conjunto de DLMs com a mesma estrutura, e fatores de desconto distintos, um para cada modelo. Durante o estudo das simulações, notou-se que o fator de desconto não possui um grande impacto na capacidade preditiva do modelo em relação aos modelos polinomiais de ordem dois convencionais, definidos na seção 2.2.2. Maiores detalhes são mostrados no próximo capítulo, onde um estudo de simulação com séries artificiais foi feito e a performance preditiva para diferentes fatores de desconto foi analisada.

4.5 Limitações do Modelo

Como foi descrito na seção 4.2, a evolução temporal do amortecedor é a principal inovação do modelo proposto em relação ao modelo polinomial de ordem dois. Os modelos

propostos na literatura assumem que o amortecedor de tendência é invariante no tempo, pois a maior vantagem do amortecedor é melhorar as previsões k passos adiante (Chatfield & Yar [8], Gardner & McKenzie [13]). A estimação é feita com um conjunto de dados de treinamento. Em relação aos modelos existentes, a principal vantagem é a estimação do parâmetro, pois o mesmo possui uma dinâmica temporal que depende apenas dos dados \mathbf{D}_t e dos valores iniciais da distribuição a priori (p_0 e q_0). Caso seja desejável uma dinâmica mais suave, é necessário especificar o coeficiente adaptativo u_t , como descrito na seção anterior. No entanto o modelo, como proposto neste trabalho, possui algumas limitações.

Deve-se notar a impossibilidade do modelo lidar com séries onde o vetor de estados tenha uma dinâmica que permita que o valor do segundo componente do vetor de estados ($\theta_{2,t}$) se aproxime de zero. Na prática isso não aconteça com frequência, pois $\theta_{2,t}$ é contínuo e não deveria se condensar em um único ponto. Note que a dinâmica do amortecedor ψ_t é movida principalmente pelo seu valor anterior no tempo $t - 1$ e pela razão entre os estados $\theta_{2,t-1}/\theta_{2,t}$. Pode-se ter problemas caso a razão atinja valores altos de maneira súbita, o que geralmente é o caso se $\theta_{2,t-1}$ for distante de zero, mas $\theta_{2,t}$ for muito próximo. A Figura 4.6 ilustra o problema. No primeiro painel temos a série original. No segundo, temos os valores de $\theta_{2,t}$ ao longo do tempo e no último painel os respectivos valores de ψ_t . Note que a partir do tempo $t = 115$ o valor de ψ_t apresenta um salto de valor e o sistema fica instável, gerando previsões com erros cada vez mais grosseiros. Tal comportamento, apesar de não ser muito frequente, aconteceu em algumas das séries da competição M3. Este problema ocorre principalmente em séries onde a tendência apresenta uma mudança de nível brusca. Isto implica que a razão $\theta_{2,t-1}/\theta_{2,t}$ apresenta uma mudança de valor alta ao passar do tempo $t - 1$ para t . Existem pelo menos três formas de solucionar isto. A primeira é filtrar o valor absurdo de ψ_t e tomar o último valor válido ($p_t = p_{t-1}$). Esta solução se enquadra dentro do enfoque bayesiano do trabalho e funcionou bem na competição M3, como descrito no Capítulo 5. Outra solução possível é usar valores de u_t altos (em torno de 0.45 já é suficiente na maioria dos casos), de forma que a dinâmica de ψ_t seja governada principalmente pelo seu valor anterior. Lembre que u_t serve como uma espécie de peso entre o valor anterior de ψ_t e a razão entre os estados. Uma terceira solução é apresentada no Capítulo 6, onde uma nova distribuição a priori é proposta para a densidade ($\psi_t | \theta_{2,t}, \mathbf{D}_t$).

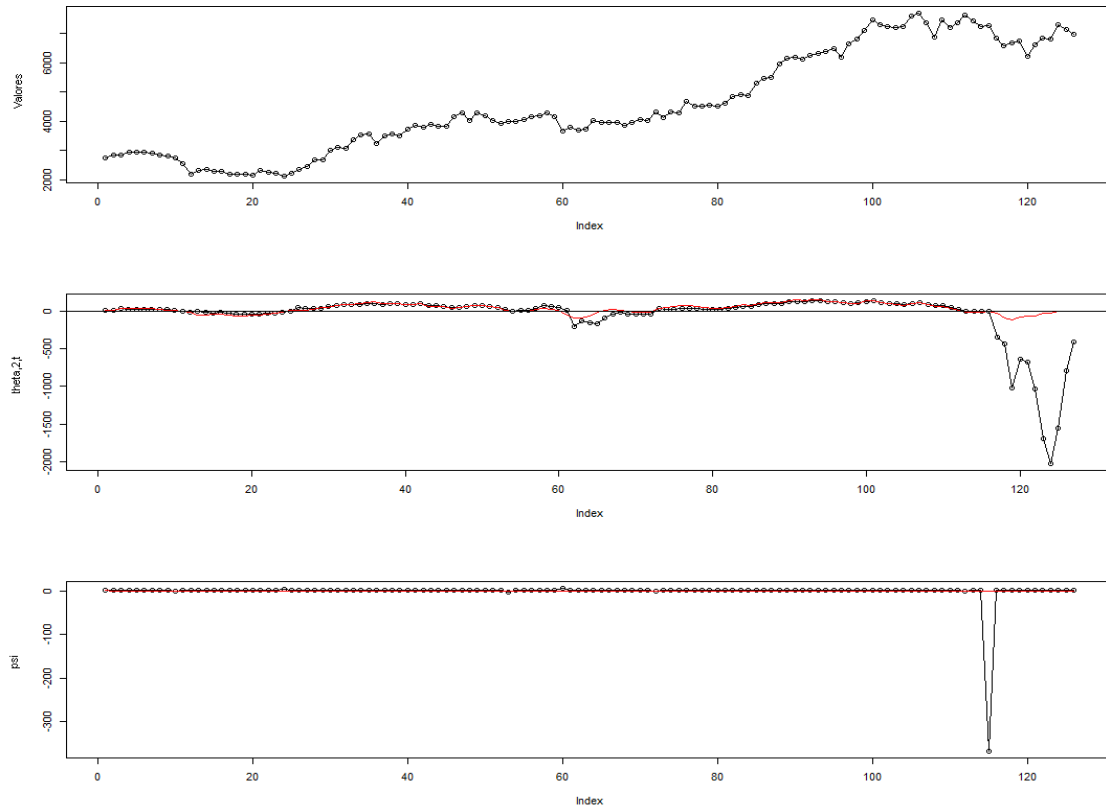


Figura 4.6: Série da competição M3 (N2527) ilustrando o problema da dinâmica do amorcedor com valores próximos de zero.

Ao realizar o estudo com séries simuladas, notou-se também que o modelo não consegue capturar o valor de ψ usado para gerar as séries. Esta deficiência vem da formulação adotada, como discutido na seção 4.3 acima. O Capítulo 6 apresenta uma nova formulação que permite que o ψ_t capture a dinâmica, mas esta formulação não foi explorada em detalhes e fica como sugestão para futuros trabalhos.

Para a previsão k passos adiante, usa-se a média dos últimos dez valores e não o último valor. Adotou-se esta prática para melhorar a capacidade preditiva do modelo na competição, pois o último valor nem sempre correspondia bem ao desenvolvimento posterior da série. Isto se deve sobretudo a séries onde a variabilidade do ψ_t era considerável. Assim, a média se tornava um valor representativo da dinâmica do ψ_t e conseguiu-se ganhos significativos com ela. Caso se adotasse apenas o último valor de ψ_t , os erros de previsão k passos à frente poderiam ser tornar grandes. Por exemplo, se $\psi_t \geq 1$ a previsão k passos à frente se torna exponencial e caso a série não tenha essa característica, os erros podem ser de magnitude considerável. Ao se tomar a média ao invés do último valor apenas, consegue-se um efeito de suavização da dinâmica, de forma que as previsões k passos à frente sejam mais condizentes com o desenvolvimento da série.

4.6 Estudo de uma Série

Para ilustrar o modelo, foi realizado um estudo com uma série em particular, de forma que a principal característica do modelo possa ser mostrada. A série N2528 da competição internacional M3 corresponde a evolução mensal do patrimônio de bancos do Banco Central Fed *Federal Reserve*. A série exibe um crescimento razoavelmente linear com certa variação e sazonalidade mensal. Para fins de comparação, adotou-se o modelo *Damped Holt* como modelo alternativo, com o amortecedor constante e estimado por mínimos quadrados. A figura abaixo ilustra a série, junto com a previsão dezoito passos à frente. O modelo proposto é mais preciso que o modelo tradicional justamente por incorporar a aceleração de crescimento na fase final da série.

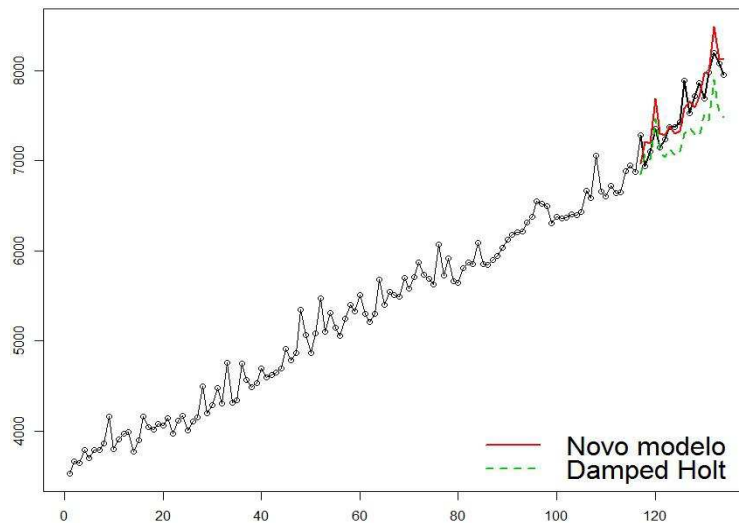


Figura 4.7: Série da competição M3 (N2528) ilustrando a vantagem da dinâmica do amortecedor para previsões.

Ao invés de especificar a matriz \mathbf{W}_t , usou-se fator de desconto 0.8 para todos os componentes do vetor de estados e aprendizagem na variância observacional V_t para ambos os modelos.

Capítulo 5

Análise dos Resultados do novo Modelo

Neste capítulo apresenta-se uma análise da performance preditiva do modelo proposto em um conjunto de séries temporais simuladas e reais. Estudou-se a sua performance em relação a outros modelos através de um estudo de caso na competição internacional M3, realizada em 2000. O novo modelo polinomial de ordem dois com abafamento (*damping*) dinâmico da tendência (*Damped Holt Dinâmico*) apresentou um desempenho satisfatório em relação aos modelos tradicionais e mais usados na prática, como os modelos aditivos *Holt Winters* e *Damped Holt*. As principais descobertas são apresentadas e discutidas neste contexto.

5.1 Simulação de Séries Temporais

A primeira parte da análise consiste em gerar uma sequência de séries com parâmetros pré-definidos numa faixa de valores discreta. Desta forma, consegue-se observar o comportamento do modelo em diferentes condições em algumas regiões do espaço paramétrico. Para a geração das séries é necessário especificar os seguintes parâmetros:

- A variância observacional V_t , com valores nos reais positivos, i.e., $V_t > 0$;
- A matriz de variância de evolução dos estados \mathbf{W}_t , positiva definida, i.e., $|\mathbf{W}_t| > 0$;
- O valor do amortecedor de tendência ψ , com valores na faixa $(0, 2)$;

O estudo consistiu na simulação de séries, cada uma com um conjunto distinto de valores para os parâmetros $\{V_t, W_{11,t}, W_{22,t}, \psi\}$. Como o espaço paramétrico é grande, e

a combinação possível de valores é explosiva, optou-se pela escolha de um conjunto pequeno, mas representativo deste espaço paramétrico. É possível que os resultados possam ser extrapolados para valores fora da faixa escolhida, mas nenhum estudo nesse sentido foi realizado. As seguintes combinações de parâmetros foram usadas:

$$V_t = (100, 10, 1) \quad W_{11,t} = (100, 10, 1) \quad W_{22,t} = (10, 1, 0.1) \quad \psi = (1, 1.01, 1.05).$$

Note que a combinação destes valores gera séries temporais apropriadas para modelos polinomiais de ordem dois. Por exemplo, caso $\psi \leq 1$ para todo o tempo t , a série apresenta mais ruído que sinal de tendência, e portanto o modelo polinomial de ordem dois não seria adequado. Desta forma, garante-se que todas as séries simuladas são consideradas adequadas para o modelo proposto. Tomou-se a variabilidade da tendência $\theta_{2,t}$ como 10% da variabilidade do componente de nível $\theta_{1,t}$. Não se tomou valores de ψ acima de 1.05 pelo fato da série, neste caso, apresentar uma tendência muito forte e a geração criaria séries com comportamento explosivo. Isto ocorre devido ao fato das variâncias permanecerem constantes na simulação da série e o nível aumentar muito em relação a variância observacional e dos estados.

Para avaliar a adequação do modelo às séries geradas, foram adotados quatro medidas de performance preditiva: SMAPE, MedAPE, MASE e *Percent Better*. Para mais detalhes sobre estas métricas, veja seção 3.4. Apenas foi computado o erro de previsão um passo à frente, i.e., calculou-se a média do erro de previsão um passo à frente para toda a série.

Para cada combinação de parâmetros $\{V_t, W_{11,t}, W_{22,t}, \psi\}$, gerou-se 1000 séries simuladas de 100 pontos cada e calculou-se as quatro métricas listadas acima.

Os valores das métricas do novo modelo foram calculados junto com o DLM polinomial de ordem dois tradicional, de forma que o mesmo sirva de base para comparação. Medidas bayesianas de performance, tais como fatores de bayes, não foram adotadas, pois as mesmas não são empregadas nas competições de previsão internacionais e não oferecem uma medida absoluta de desempenho do modelo. As subseções seguintes ilustram os resultados obtidos para cada combinação de parâmetros empregada e para cada métrica adotada. Nas subseções abaixo, cada figura ilustra a métrica usada no eixo vertical e os valores de fator de desconto no eixo horizontal.

5.1.1 Resultados para as Séries com $V_t = 100, W_{11,t} = 1, W_{22,t} = 0.1$

Nesta subseção, é discutido o comportamento do modelo para séries onde o nível de ruído, representado pela variância observacional V_t , é muito maior que o nível de sinal,

dados pela variância do primeiro componente do vetor de estados $W_{11,t}$. Na literatura, para modelos polinomiais de ordem um esta é a razão sinal/ruído (*signal-to-noise ratio*, Petris *et al.* [29]). Lembre que adota-se $W_{22,t}$, a variância da tendência, como sendo 10% da variância do nível. Esta representação é condizente com os valores encontrados na prática.

A Figura 5.1 mostra os valores de SMAPE para as séries simuladas, onde cada ponto representa a média SMAPE de 1000 séries simuladas. Note que o novo modelo performa melhor que o polinomial de ordem dois, sendo que a melhora diminui conforme o fator de desconto decresce. Para o fator de desconto 0.80 os resultados podem ser considerados equivalentes. Veja que para $\psi = 1.05$ o resultado é muito melhor do que para os outros valores. Isto gera evidências de que o modelo performa bem em séries com forte tendência.

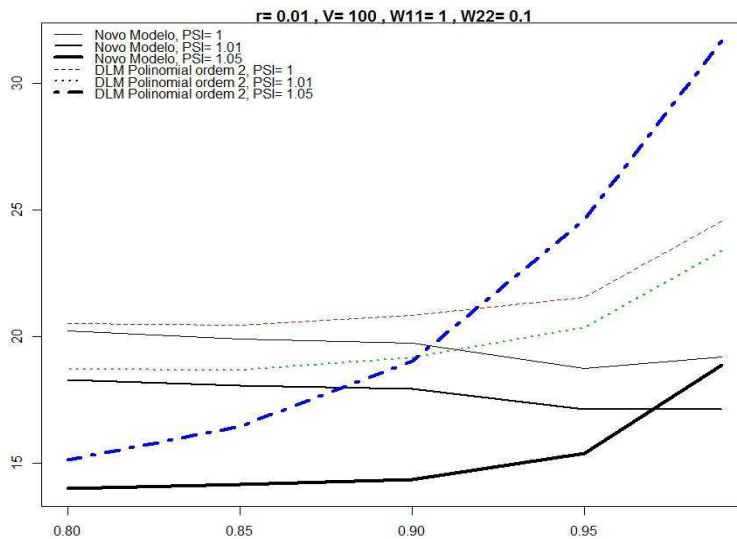


Figura 5.1: SMAPE das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 1, W_{22,t} = 0.1$.

As conclusões acima ainda são válidas para os valores de MedAPE obtidos nas mesmas condições, mas com valores menos significativos, como mostra a Figura 5.2.

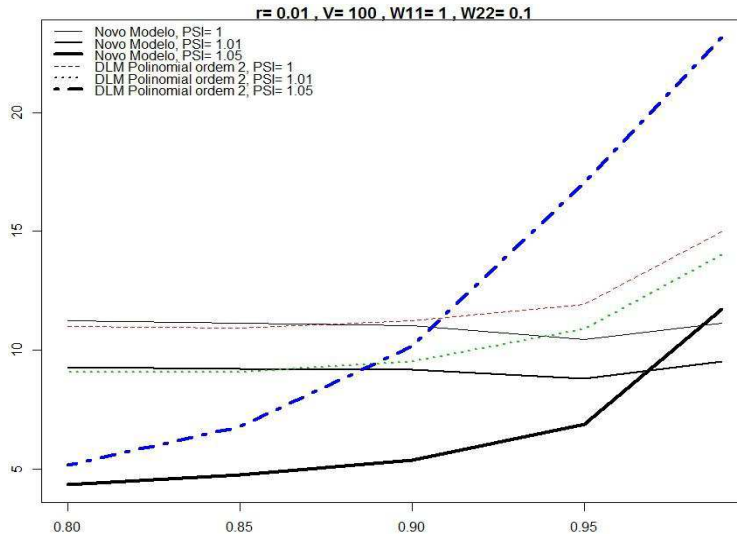


Figura 5.2: MedAPE das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 1, W_{22,t} = 0.1$.

Usando MASE, as diferenças entre os modelos não são tão claras, a não ser para $\psi = 1.05$, onde o novo modelo performa melhor. O modelo polinomial performa pior quando o fator de desconto se encontra acima de 0.95 (veja Figura 5.3).

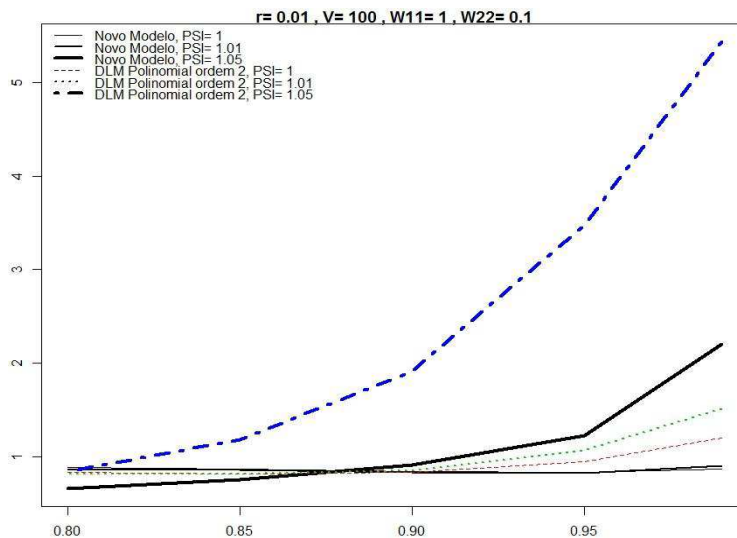


Figura 5.3: MASE das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 1, W_{22,t} = 0.1$.

Por fim, a medida *Percent Better* mostra que o SMAPE do novo modelo é melhor que o SMAPE do modelo polinomial de ordem dois por pelo menos 60% dos casos, sendo que a porcentagem aumenta conforme se incrementa o fator de desconto. Para $\psi = 1.05$ o resultado é ainda melhor, como mostra 5.4.

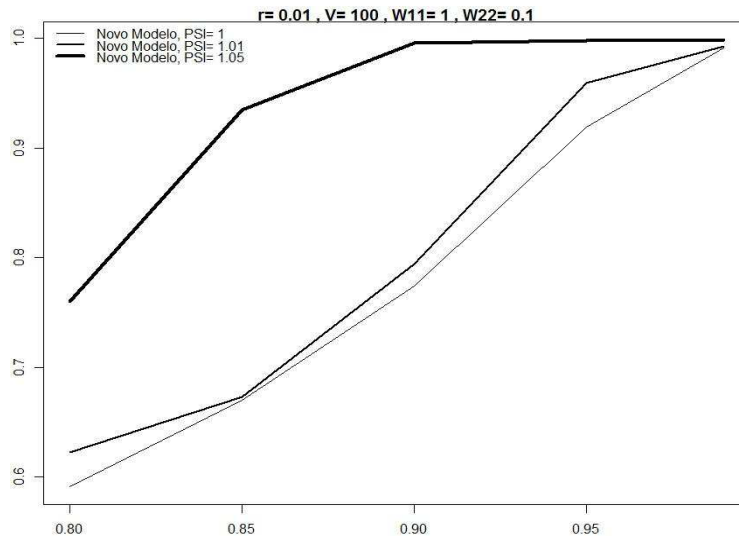


Figura 5.4: *Percent Better* das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 1, W_{22,t} = 0.1$.

A próxima seção mostra os resultados para séries onde o nível de ruído é menor, i.e., onde a razão $W_{11,t}/V_t$ é maior.

5.1.2 Resultados para as Séries com $V_t = 100, W_{11,t} = 10, W_{22,t} = 1$

Aumentando a variância do estado, obtém-se resultados mais claros, onde se pode conjecturar que o novo modelo performa melhor ao longo de todo o espaço paramétrico. A Figura 5.5 mostra os valores de SMAPE para as séries simuladas. Pode-se notar que o novo modelo é visivelmente superior.

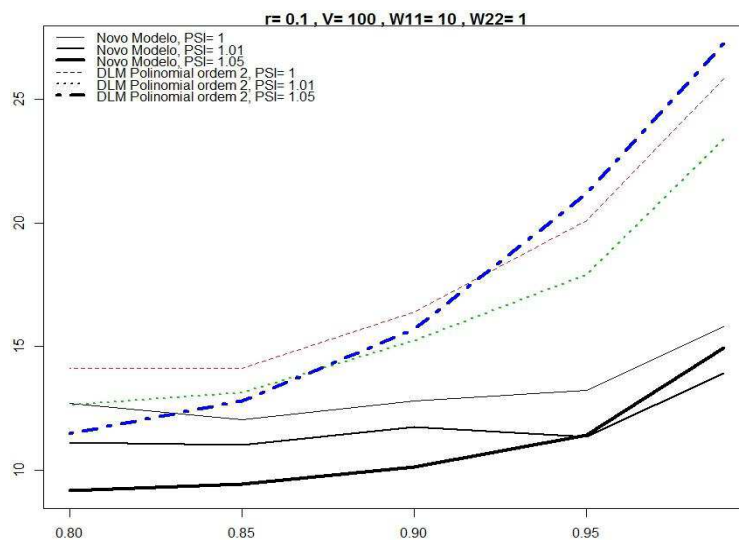


Figura 5.5: SMAPE das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 10, W_{22,t} = 1$.

As conclusões acima ainda são válidas usando a medida MedAPE, como mostra a Figura 5.6.

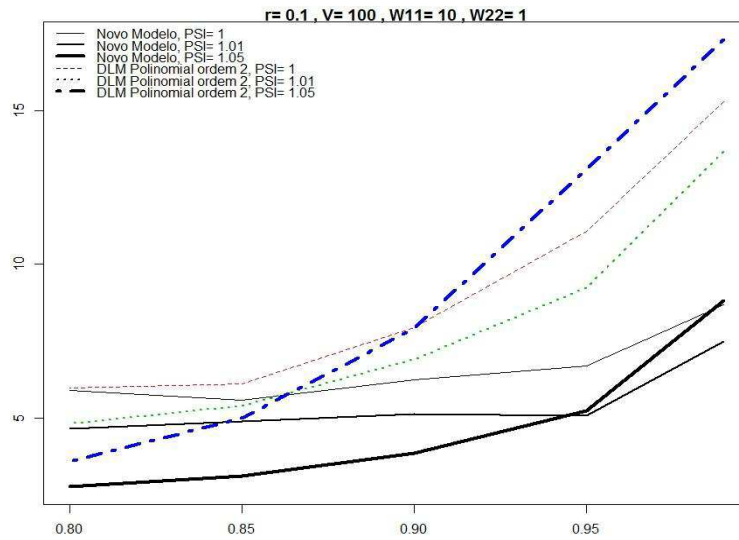


Figura 5.6: MedAPE das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 10, W_{22,t} = 1$.

Usando MASE as medidas não são tão claras, mas consegue-se perceber ganho de performance preditiva a partir de fator de desconto 0.85 e acima (Figura 5.7).

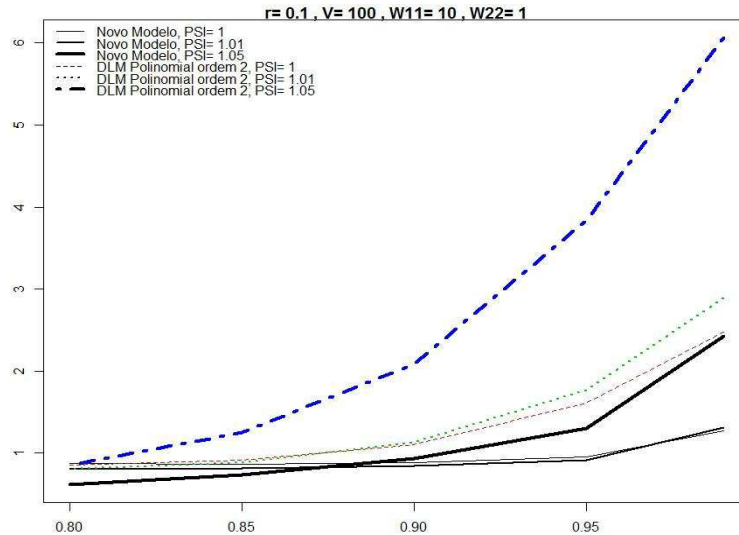


Figura 5.7: MASE das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 10, W_{22,t} = 1$.

Por fim, a medida *Percent Better* mostra a mesma conclusão da subseção anterior, agora com resultados ainda melhores. Note também que a performance melhora conforme se aumenta o valor de ψ usado nas simulações.

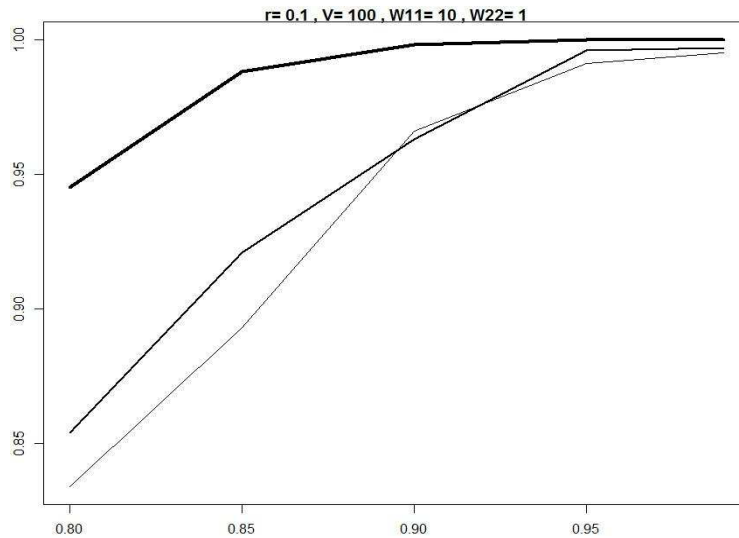


Figura 5.8: *Percent Better* das séries simuladas com parâmetros $V_t = 100, W_{11,t} = 10, W_{22,t} = 1$.

5.1.3 Resultados para as Séries com $V_t = 10, W_{11,t} = 10, W_{22,t} = 1$

A Figura 5.9 mostra os valores de SMAPE para séries onde o nível de sinal em relação ao ruído é ainda maior que o da seção anterior. Os resultados são ainda mais encorajadores para o novo modelo. Veja que independente do valor de ψ usado para gerar a série, o novo modelo performa muito acima do modelo tradicional, sendo que a melhoria aumenta conforme se aumenta o valor de ψ . Conforme o fator de desconto diminui, a performance preditiva aumenta, já que ainda existe um nível de ruído considerável na geração das séries.

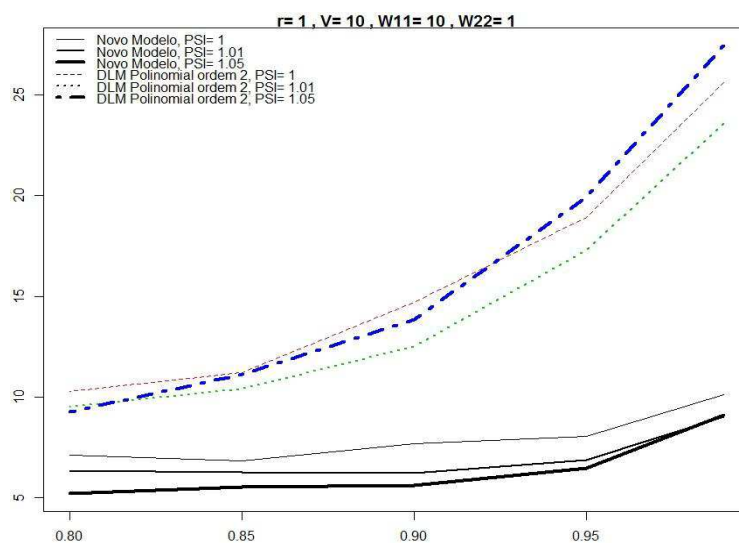


Figura 5.9: SMAPE das séries simuladas com parâmetros $V_t = 10, W_{11,t} = 10, W_{22,t} = 1$.

As conclusões acima ainda são válidas para a medida MedAPE, como mostra a Figura 5.10. Agora os resultados são ligeiramente piores.

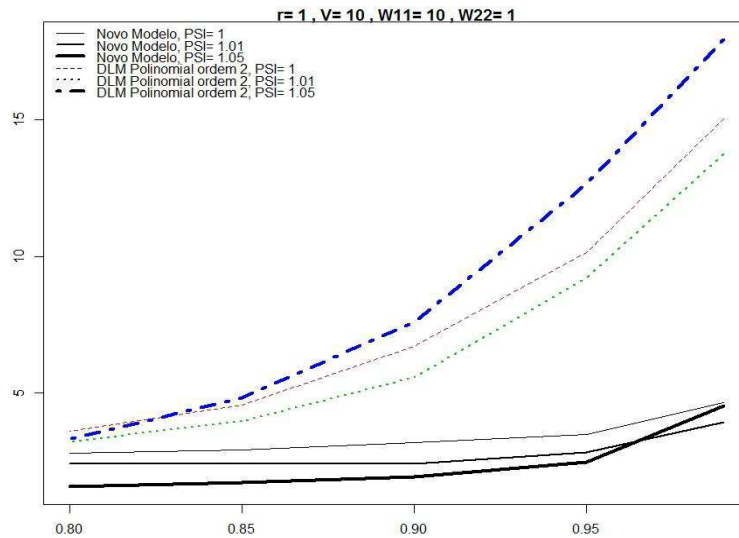


Figura 5.10: MedAPE das séries simuladas com parâmetros $V_t = 10, W_{11,t} = 10, W_{22,t} = 1$.

Usando MASE as conclusões agora são as mesmas obtidas das medidas SMAPE e MedAPE acima. Veja Figura 5.11.

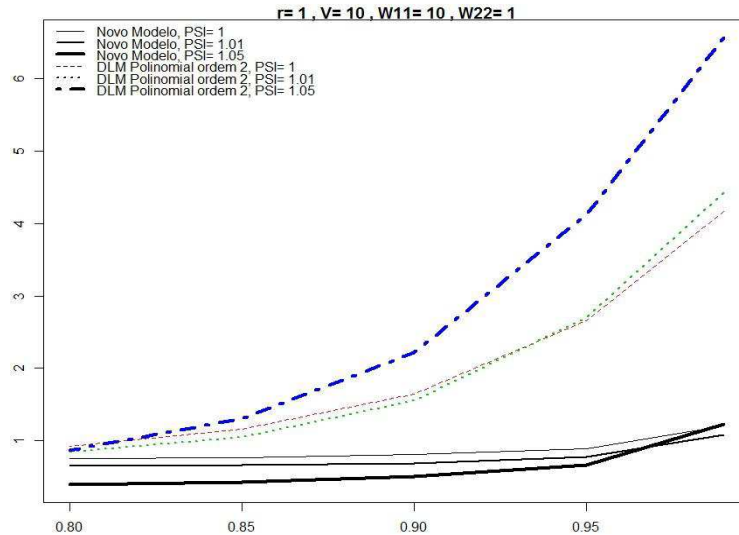


Figura 5.11: MASE das séries simuladas com parâmetros $V_t = 10, W_{11,t} = 10, W_{22,t} = 1$.

Por fim, a medida *Percent Better* para SMAPE é omitida pois os resultados são muito bons, acima de 98% para diferentes valores de ψ .

5.1.4 Resultados para as Séries com $V_t = 1, W_{11,t} = 10, W_{22,t} = 1$

A Figura 5.12, como nas seções anteriores, mostra os valores de SMAPE obtidos com a simulação. Note que os resultados são ainda melhores do que os obtidos na seção anterior. Agora fica mais claro a melhoria na performance preditiva em vários pontos do espaço paramétrico. Neste conjunto de parâmetros as quatro medidas são muito similares em termos de comportamento e as conclusões são as mesmas da seção anterior.

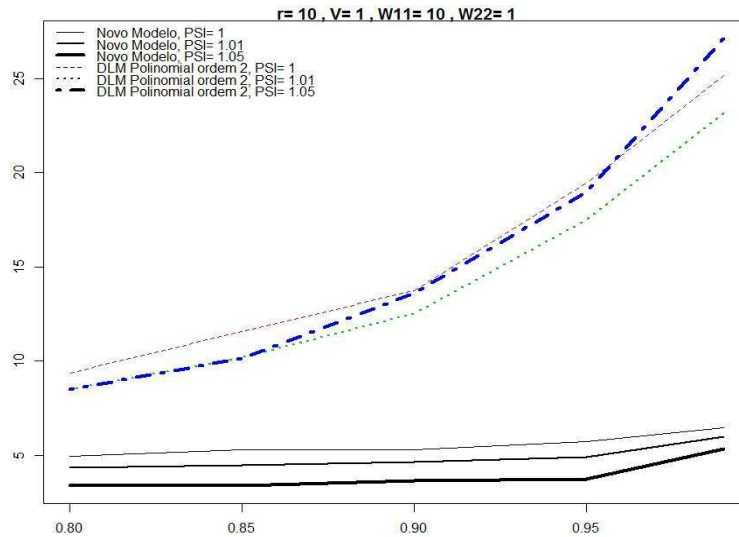


Figura 5.12: SMAPE das séries simuladas com parâmetros $V_t = 1, W_{11,t} = 10, W_{22,t} = 1$.

Para a medida MedAPE, chega-se as mesmas conclusões, como mostra a Figura 5.13.

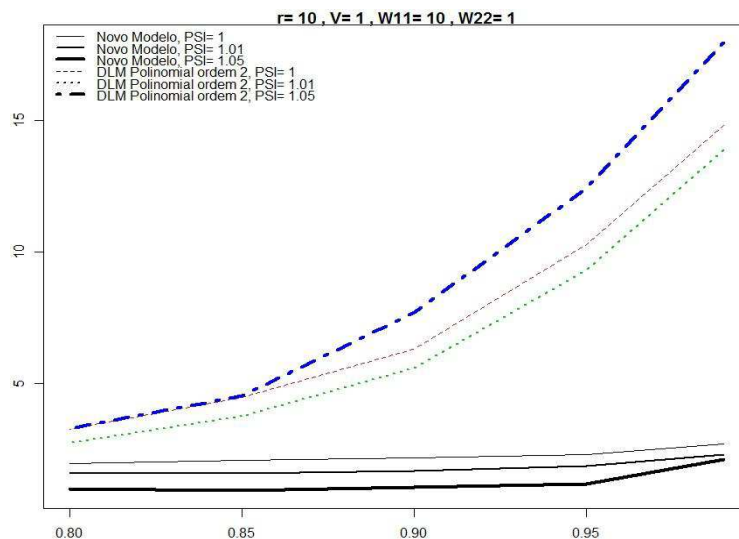


Figura 5.13: MedAPE das séries simuladas com parâmetros $V_t = 1, W_{11,t} = 10, W_{22,t} = 1$.

A medida MASE também é muito similar as duas medidas acima e é mostrada (5.14).

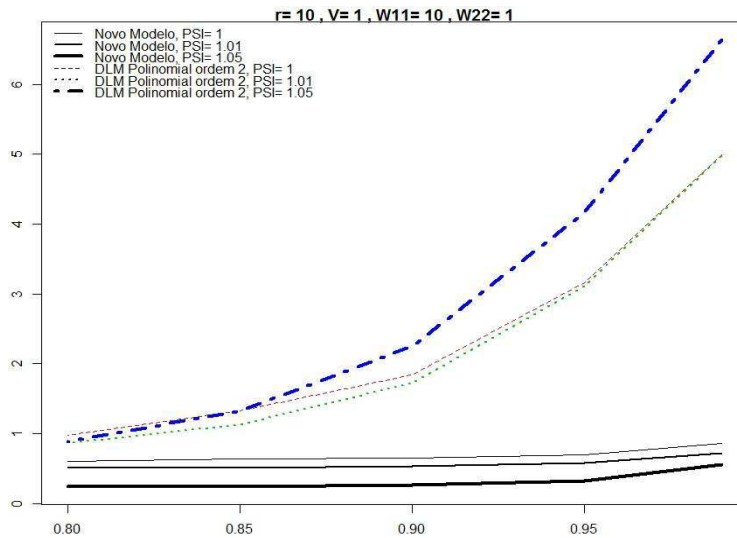


Figura 5.14: MASE das séries simuladas com parâmetros $V_t = 1, W_{11,t} = 10, W_{22,t} = 1$.

A medida *Percent Better* é omitida novamente, pois os valores são acima de 98% em todos os casos.

5.1.5 Resultados para as Séries com $V_t = 1, W_{11,t} = 100, W_{22,t} = 10$

Nesta última seção, os resultados são ainda mais encorajadores, onde agora o nível de sinal é muito alto em relação ao ruído da série. As quatro medidas permitem obter as mesmas conclusões obtidas na seção anterior. Veja 5.15 para a medida SMAPE.

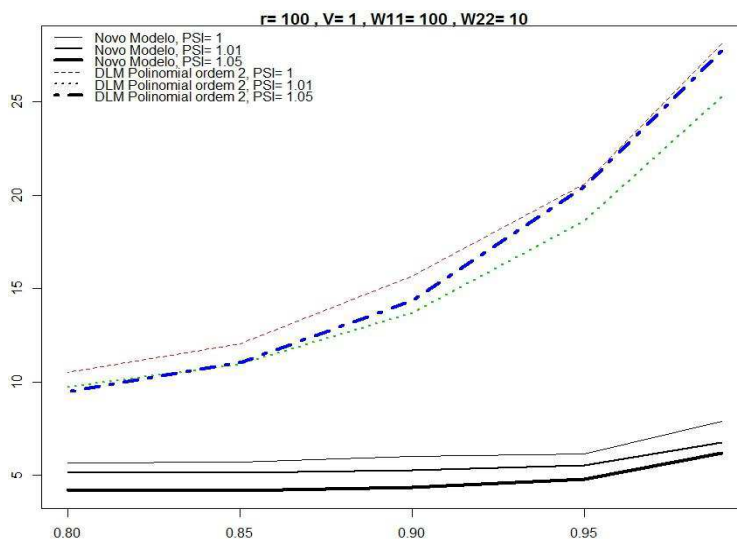


Figura 5.15: SMAPE das séries simuladas com parâmetros $V_t = 1, W_{11,t} = 100, W_{22,t} = 10$.

A Figura 5.16 mostra os resultados para a medida MedAPE.

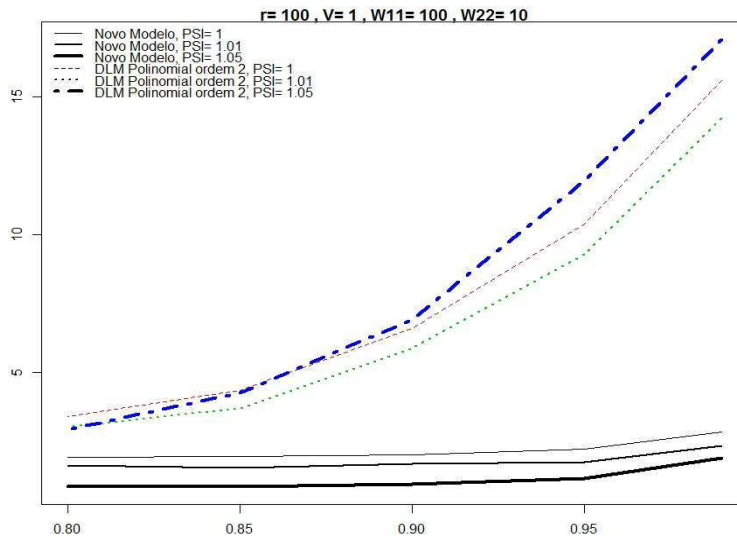


Figura 5.16: MedAPE das séries simuladas com parâmetros $V_t = 1, W_{11,t} = 100, W_{22,t} = 10$.

A medida MASE é ilustrada na Figura 5.17.

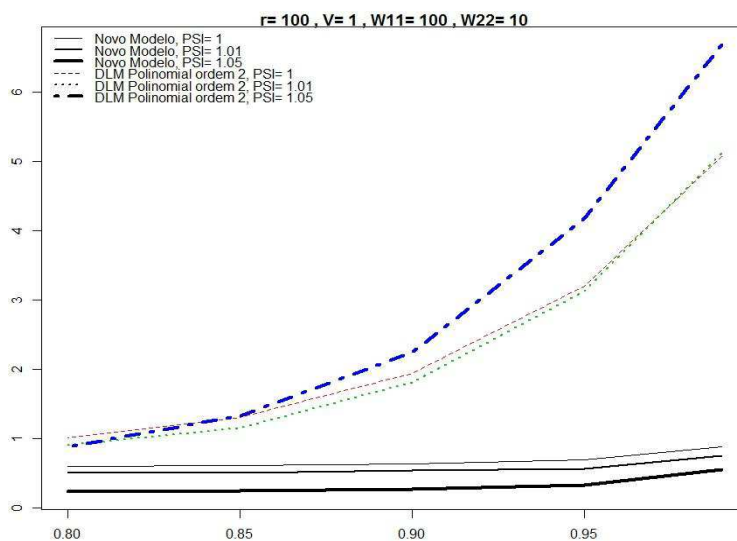


Figura 5.17: MASE das séries simuladas com parâmetros $V_t = 1, W_{11,t} = 100, W_{22,t} = 10$.

Novamente, a medida *Percent Better* de SMAPE é omitida, pois os resultados são todos acima de 98% em todos os valores de V_t, W_t, ψ usados.

Analisando os resultados acima, pode-se perceber que o modelo performa melhor em séries com tendência forte, ou seja, séries onde $\psi > 1$. Conforme a tendência aumenta de magnitude o modelo proposto tende a performar melhor que o modelo polinomial de ordem dois. Além disso, pode-se notar que na medida que a série apresenta mais sinal que ruído, dado pela razão entre as variâncias $W_{11,t}/V_t$ o modelo tende a performar melhor.

Note também que o fator de desconto não apresenta uma influência muito grande na capacidade preditiva do novo modelo em relação ao modelo polinomial de ordem dois.

5.2 Análise da Competição Internacional M3

Além da simulação de séries para análise do comportamento de modelos, estudos empíricos no campo de análise de séries temporais tem comparado a acurácia (*post-sample forecasting accuracy*) de vários métodos, de forma que a sua performance preditiva possa ser determinada de maneira objetiva e mensurável. A competição internacional M3, realizada em 1999/2000, é a mais reconhecida do gênero (Makridakis & Hibon [26]). Outras competições estão em fase de desenvolvimento, como a competição M4 (a mais nova versão da competição M3) e a competição NN3, que é uma replicação da M3 com uma extensão para métodos de redes neurais e inteligência computacional (Cronea *et al.* [9]). Previamente, houveram também as competições M1 e M2, mas foram realizadas em um conjunto de dados menor e menos abrangente. As competições M1, M2 e M3 são estudos empíricos que compararam a performance de um grande número de métodos para análise de séries temporais usando especialistas reconhecidos. Estes provem as previsões, usando o método no qual são especializados. Makridakis & Hibon [26] publicaram os principais resultados da competição M3, resumindo a performance relativa de todos os métodos empregados no estudo.

Os resultados da competição são os seguintes:

- Métodos estatísticos sofisticados ou mais complexos não produzem necessariamente previsões melhores que métodos mais simples;
- O ranking de performance relativa dos vários métodos varia de acordo com a métrica usada;
- A acurácia de vários métodos combinados é, na média, melhor que a acurácia de cada método individual;
- A acurácia dos vários métodos depende do comprimento do horizonte de previsão;

O conjunto de dados do estudo consistiu de 3003 séries com variadas características e diferentes intervalos de tempo entre as observações (anual, quateral, mensal, etc.), além de pertencer a áreas de conhecimento distintas, tais como indústria, demografia, finanças, macroeconomia, microeconomia e meteorologia. Afim de garantir que um número suficiente de dados estava disponível para desenvolver um modelo adequado, foi decidido que cada série teria um número mínimo de observações para cada tipo de dado, de acordo com o intervalo de tempo entre as observações. As séries mensais variam em comprimento de 48 a 126, com mediana 115. Todos os dados da competição são estritamente positivos. Para as previsões k passos à frente ($k \geq 1$), caso alguma delas seja negativa, o valor foi

substituído por zero, de forma que a métrica SMAPE seja sempre positiva. Este procedimento foi empregado para todos os modelos testados e segue o procedimento usado na competição M3.

Para previsões de vários passos adiante, não se adotou uma dinâmica no amortecedor, pois não existem dados que possam orientar a dinâmica. Neste caso, uma solução razoável é adotar uma média dos últimos dez valores do amortecedor como estimativa do coeficiente. A razão para adotar esta estimativa foram os bons resultados obtidos na competição M3. Note que neste caso as previsões serão diferentes do modelo com amortecedor fixo, pois o modelo fixo usa todo o conjunto de treinamento para estimar o seu valor. No caso do modelo com abafamento dinâmico, os últimos valores do amortecedor possivelmente irão capturar o valor considerado adequado, pelo menos para a previsão de curto prazo.

Cinco medidas de acurácia foram empregadas na competição: SMAPE, MedAPE, Percent Better, Average Ranking e RAE mediano. Para uma descrição de cada uma delas, veja (Makridakis & Hibon [26] e a seção 3.4).

Modelos lineares dinâmicos não foram avaliados na competição. A tese de Zhai [41] continuou o estudo da competição M3 e avaliou a performance de três paradigmas, a saber: Modelos Lineares Dinâmicos, Redes Neurais e modelos ARIMA (*Autoregressive Integrated Moving Average models*). O objetivo foi descobrir como os diferentes tipos de séries temporais afetam a capacidade preditiva destes três paradigmas. Além disso, foi exposto uma metodologia de seleção de modelos de previsão de acordo com as características dos dados. Para comparar os resultados foi empregada a métrica SMAPE (veja seção 3.4), a mesma usada na competição internacional M3. A tese de Karlon [23] focou apenas na performance de DLMS otimizados para a competição M3, o que mostrou que a performance deste paradigma é dependente da classe de séries temporais. Neste caso, Karlon conseguiu atingir um performance preditiva superior a de Zhai através da técnica de seleção de modelos, i. e., foi determinado a arquitetura ótima para cada série temporal e então foi determinado se o DLM escolhido supera os paradigmas da competição M3 e da pesquisa feita por Zhai.

Neste estudo, foram consideradas apenas as 1428 séries mensais da competição, seguindo a mesma metodologia de (Taylor [35]). Neste caso, procedimentos automáticos de previsão raramente são usados para séries com frequência maior que a mensal. Note também que o novo modelo, por não precisar realizar simulações, pode também ser usado para séries de alta frequência, tais como séries com intervalos de um minuto ou de apenas segundos, como é o caso de muitas séries financeiras de preços de ativos listados em bolsa.

Os dados não foram dessazonalizados, como usualmente se procede antes de aplicar o modelo para previsão e estimação de parâmetros. Ao invés de dessazonalizar os dados, aplicou-se um componente de sazonalidade de forma livre (veja seção 2.2.3), já que a sazonalidade é mensal. Para determinar se a sazonalidade era necessária para o *Damped Holt* Dinâmico, criou-se dois modelos, sendo um deles com um componente extra correspondendo a sazonalidade e o outro com apenas o componente de tendência linear. Calculava-se a métrica SMAPE para os dados de treinamento e comparava-se o valor de cada modelo, sendo que o modelo com o menor valor era o escolhido para realizar a previsão k passos à frente, seguindo a metodologia de Karlon. Para o modelo *Damped Holt* tradicional, usou-se mínimos quadrados para estimar se a sazonalidade era necessária, já que este é o procedimento mais empregado na literatura.

Uma vez detalhado como as séries são tratadas, parte-se agora para definir os dois principais objetivos da análise: descobrir se existe valor em atribuir uma dinâmica temporal no amortecedor de tendência e quantificar como o novo método performa em relação ao modelo tradicional *Damped Holt*, com ψ fixo e $0 \leq \psi \leq 1.1$ e aos DLM polinomiais de ordem dois, o caso particular com $\psi_t = 1$ para todo tempo t e o DLM polinomial de ordem um.

Foram produzidas previsões para os quatro métodos considerados adequados para séries com tendência:

- *Damped Holt* com erro e sazonalidade aditivos, com parâmetros estimados por mínimos quadrados e $0 \leq \psi \leq 1.1$;
- O novo modelo *Damped Holt* Dinâmico, com parâmetros (u_t e fator de desconto δ) estimados pela minimização da métrica SMAPE, aprendizagem na variância e $0 \leq \psi \leq 1.1$;
- DLM polinomial de ordem dois com aprendizagem na variância e fator de desconto estimado pela minimização da métrica SMAPE, com $\psi = 1$;
- DLM polinomial de ordem um com aprendizagem na variância e fator de desconto estimado pela minimização da métrica SMAPE;

Uma nota é conveniente neste momento. Adotou-se a estimação de parâmetros por mínimos quadrados para o modelo *Damped Holt* por este ser o procedimento mais empregado na prática e também o mais adotado pela literatura. Como o enfoque deste trabalho é bayesiano, nas seções seguintes são apresentados mais resultados com o mesmo modelo, mas com a estimação dos parâmetros por multiprocessos classe I. As previsões foram feitas na série original, sem nenhuma transformação das variáveis.

Os valores iniciais da *priori* $\theta_0|\mathbf{D}_0$ foram escolhidos como $\mathbf{m}_0 = (y_1, 1)'$ e $\mathbf{C}_0 = \text{diag}(1e7, 1e7)$, onde y_1 é o primeiro valor da série.

Usando cada um dos quatro métodos acima foram produzidos previsões nos horizontes de 1 a 18 para cada uma das 1428 séries, de maneira similar ao procedimento adotado na competição M3 e nas outras competições internacionais M1, M2 e NN3.

5.2.1 Resultados para todas as Séries

As tabelas abaixo resumam os resultados para cada um dos quatro métodos aplicados a todas as 1428 séries mensais da competição M3. Várias métricas foram propostas na literatura e decidiu-se adotar as mais comumente usadas nas competições internacionais, SMAPE (*Symmetric Mean Absolute Percentage Error*), MedAPE (*Median Average Percentage Error*) e MASE (*Mean Absolute Scaled Error*). A medida MASE foi proposta na literatura em 2006, veja (Hyndman & Koehler [18]), e a mesma não foi adotada na competição M3 por ter sido sugerida anos mais tarde, e portanto não pode ser usada para fins de comparação com os métodos empregados na M3. No entanto, como a métrica foi usada na competição NN3 e é usada com certa frequência na literatura, foi decidido adotá-la também. Para mais detalhes, veja seção 3.4. Por simplicidade, reporta-se a média de cada métrica para diferentes horizontes de previsão: 1-6, para curto prazo, 7-12, para médio prazo, 13-18 para longo prazo e, para complementar, o horizonte 1-12 e o horizonte completo 1-18. Alguns horizontes individuais de curto prazo também são reportados. Os melhores resultados para cada horizonte de previsão são escritos em negrito.

A Tabela 5.1 resume o erro SMAPE, MedAPE e MASE de cada um dos quatro métodos no curto prazo (horizontes 1-6). Note que o novo método é ligeiramente melhor para horizontes de curto prazo de acordo com a SMAPE e MASE nos horizontes de 3 a 6, e em primeiro ou segundo lugar de acordo com o MedAPE. Para previsões de um passo a frente, o método ficaria em sexto lugar na competição internacional. Os DLM polinomiais tradicionais não performaram bem em relação ao modelos com amortecedores, sendo que o modelo polinomial de ordem um obteve a pior performance nos horizontes 1-6.

Tabela 5.1: SMAPE, MedAPE e MASE de horizontes 1-6 para as 1428 séries mensais da competição M3.

| Métrica | Método | Horizonte de Previsão | | | | | |
|---------|------------------------|-----------------------|--------------|--------------|--------------|--------------|--------------|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| SMAPE | <i>Damped Holt</i> | 11.6 | 12 | 12.5 | 14 | 15.15 | 13.33 |
| | Novo Método | 11.73 | 11.45 | 12.19 | 13.32 | 13.44 | 12.99 |
| | DLM Polinomial ordem 2 | 11.62 | 11.9 | 13.13 | 14.44 | 14.12 | 13.96 |
| | DLM Polinomial ordem 1 | 13.96 | 13.08 | 14.91 | 16.16 | 14.8 | 14.73 |
| MedAPE | <i>Damped Holt</i> | 4.17 | 4.21 | 4.46 | 5.94 | 6.05 | 6.29 |
| | Novo Método | 4.22 | 4.52 | 4.45 | 6.14 | 5.84 | 5.78 |
| | DLM Polinomial ordem 2 | 4.44 | 4.61 | 4.75 | 6.33 | 6.11 | 6 |
| | DLM Polinomial ordem 1 | 6.6 | 6.36 | 7.22 | 8.66 | 7.57 | 7.75 |
| MASE | <i>Damped Holt</i> | 0.93 | 1.05 | 1.21 | 1.55 | 1.69 | 1.8 |
| | Novo Método | 1.02 | 1.08 | 1.2 | 1.49 | 1.6 | 1.68 |
| | DLM Polinomial ordem 2 | 1.02 | 1.1 | 1.25 | 1.55 | 1.66 | 1.78 |
| | DLM Polinomial ordem 1 | 1.68 | 1.76 | 1.99 | 2.26 | 2.32 | 2.46 |

Por simplicidade e em conformidade com a maneira como os dados foram reportados na competição M3, foram calculadas os valores médios das métricas no curto (1-4 e 1-6), médio (7-12 e 1-12) e longo prazo (13-18), assim como a performance geral, dada pelo horizonte completo 1-18. A Tabela 5.2 mostra os resultados para todas as métricas usadas. O modelo *Damped Holt* Dinâmico performa um pouco melhor que todos os outros modelos no curto e médio prazo. No longo prazo, o modelo proposto é um pouco pior que o *Damped Holt*, mas melhor que o modelo polinomial de ordem dois, de acordo com a métrica MedAPE. Para os horizontes de 1 a 6 e 7 a 12 o modelo tem valores menores em todas as métricas adotadas, o que sugere que o modelo é potencialmente melhor que o tradicional *Damped Holt* no curto e médio prazo. Para o horizonte 13-18, o novo modelo é ligeiramente pior, ficando em segundo lugar em todas as métricas, exceto o SMAPE, onde fica em primeiro lugar. Isto é esperado, já que o modelo *Damped Holt* é ótimo para séries com tendência moderada, o que é o caso em muitas séries da competição. No horizonte completo de 1-18, contudo, o *Damped Holt* dinâmico é um pouco melhor de acordo com o SMAPE e o MASE, e fica em segundo lugar de acordo com o MedAPE. Note que os modelos polinomiais tradicionais, consistentemente performam pior de acordo com todas as métricas. Como este modelo não performa bem no longo prazo nas outras métricas, é provável que previsões muito ruins em algumas séries tenham prejudicado o desempenho do novo modelo e do *Damped Holt* tradicional. Uma confirmação disso é o fato do modelo *Damped Holt* ter performado bem no longo prazo na métrica MedAPE, a mediana dos erros absolutos. O fato do novo método ter performado ligeiramente melhor no curto e médio prazo em relação ao modelo *Damped Holt* tradicional é muito encorajador, já que

é um dos modelos mais empregados na prática a várias décadas e tem sido usado como *benchmark* em competições internacionais.

Tabela 5.2: SMAPE, MedAPE e MASE de horizontes curto, médio e longo prazo para as 1428 séries mensais da competição M3.

| Métrica | Método | Horizonte de Previsão | | | | | |
|---------|------------------------|-----------------------|--------------|--------------|--------------|--------------|--------------|
| | | 1-4 | 1-6 | 7-12 | 13-18 | 1-12 | 1-18 |
| SMAPE | <i>Damped Holt</i> | 12.52 | 13.1 | 14.73 | 19.55 | 13.91 | 15.79 |
| | Novo Método | 12.17 | 12.52 | 14.13 | 19.47 | 13.32 | 15.37 |
| | DLM Polinomial ordem 2 | 12.77 | 13.2 | 15.49 | 21.77 | 14.34 | 16.82 |
| | DLM Polinomial ordem 1 | 14.53 | 14.61 | 15.71 | 19.82 | 15.16 | 16.71 |
| MedAPE | <i>Damped Holt</i> | 4.7 | 5.19 | 6.94 | 9.37 | 6.07 | 7.17 |
| | Novo Método | 4.83 | 5.16 | 6.67 | 9.71 | 5.91 | 7.18 |
| | DLM Polinomial ordem 2 | 5.03 | 5.37 | 7.06 | 10.1 | 6.22 | 7.51 |
| | DLM Polinomial ordem 1 | 7.21 | 7.36 | 8.89 | 11.98 | 8.12 | 9.41 |
| MASE | <i>Damped Holt</i> | 1.18 | 1.37 | 2.15 | 3.07 | 1.76 | 2.2 |
| | Novo Método | 1.2 | 1.34 | 2.02 | 3.11 | 1.68 | 2.16 |
| | DLM Polinomial ordem 2 | 1.23 | 1.39 | 2.1 | 3.23 | 1.75 | 2.24 |
| | DLM Polinomial ordem 1 | 1.92 | 2.08 | 2.88 | 3.87 | 2.48 | 2.94 |

Para confirmar se as métricas usadas tiveram em geral coerência para as séries da competição, calculou-se a percentagem de vezes que as métricas tiveram resultados divergentes, i.e., o número de vezes em que a métrica SMAPE do novo modelo era melhor que a mesma métrica para o *Damped Holt* e ao mesmo tempo a métrica MedAPE do novo modelo era pior que a mesma métrica do modelo *Damped Holt*. Isto aconteceu em 166 séries de 1428 (11.6%), o que indica que nestas séries existem *outliers* que podem estar prejudicando a performance do modelo. A divergência entre o SMAPE e o MASE foi menor, de apenas 1.3%, o que está dentro de limites aceitáveis.

O novo modelo assume que os resíduos possuem distribuição normal. Como as séries da competição vem de diferentes categorias e possuem tamanhos diferentes, foi realizado um teste de normalidade dos resíduos para assegurar que esta hipótese do modelo tenha sido atendida pela maioria das séries. Das 1428 séries mensais, 1427 passam no teste de *shapiro* de normalidade com 5% de significância. Deste modo, podemos ter certa segurança que a hipótese de normalidade praticamente não influenciou a performance do modelo na competição.

5.2.2 Resultados para todas as Séries usando Multiprocessos

A abordagem bayesiana permite que a estimação dos parâmetros seja feita por uma mistura de densidades, como descrito na seção 3.3. Neste caso, a previsão é composta de um conjunto de previsões de cada modelo individual, ponderado pela sua probabilidade a *posteriori*, dadas as observações. Para fins de comparação com os resultados acima, foram também feitas análises dos quatro principais modelos já existentes com os dados da competição, agora usando multiprocessos classe I como método de estimação de parâmetros e da densidade preditiva. Como o enfoque deste trabalho é bayesiano, considera-se importante apresentar os resultados também com uma estimação bayesiana. A acurácia de previsão pode ser melhorada através da combinação de múltiplas previsões individuais. Não existe uma única técnica ou modelo que consegue performar bem em todas as séries e horizontes de previsão. Logo, ao combinar as previsões de modelos individuais, pode-se reduzir o viés de especificação dos modelos individuais e aumentar a capacidade preditiva do supermodelo resultante da combinação.

A Tabela 5.3 ilustra os resultados obtidos com esta abordagem no curto prazo, com horizontes 1-6. Os resultados para o novo modelo foram calculados usando multiprocessos para uma malha de valores de u_t na faixa (0.1,0.8) com intervalo de 0.05. No horizonte de curto prazo 1-3 o modelo novo ainda performa melhor nas métricas MedAPE e MASE. Para os horizontes 3-6 ele ainda fica em primeiro ou segundo lugar em relação aos outros modelos, ficando em primeiro lugar na métrica MASE. A pior performance ficou com o modelo polinomial de ordem um e o modelo polinomial de ordem dois foi o segundo pior em quase todos os horizontes.

Seguindo a metodologia da seção anterior, são reportados também os resultados para horizontes de curto, médio e longo prazo, como mostra a Tabela 5.4. Para os horizontes de 1 a 4 e 1 a 6, o novo modelo fica em primeiro lugar em todas as métricas. Novamente, todas as medidas confirmam que a performance do novo modelo é boa no curto prazo. Para o horizonte de médio prazo, 7 a 12, o modelo proposto fica em primeiro ou segundo lugar. No longo prazo, o novo modelo fica em segundo lugar nas métricas MedAPE e MASE. O fato do novo modelo ficar em segundo lugar na medida SMAPE no prazo completo 1-18 indica a presença de *outliers* que podem estar prejudicando a performance do modelo. Isto motivou o estudo de condições onde o modelo pode ter boa acurácia. Os resultados estão na seção seguinte. Para horizontes de longo prazo, não houve consenso entre as métricas sobre qual modelo performa bem. Para o horizonte completo 1-18, o novo modelo ficou com a melhor performance de acordo com o MedAPE e o MASE, e

Tabela 5.3: SMAPE, MedAPE e MASE de horizontes 1-6 para as 1428 séries mensais da competição M3, com multiprocessos classe I.

| Métrica | Método | Horizonte de Previsão | | | | | |
|---------|------------------------|-----------------------|--------------|--------------|--------------|--------------|--------------|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| SMAPE | <i>Damped Holt</i> | 11.74 | 11.44 | 12.52 | 14.04 | 12.88 | 12.55 |
| | Novo Método | 11.59 | 11.52 | 12.24 | 13.39 | 13.38 | 12.85 |
| | DLM Polinomial ordem 2 | 11.75 | 11.74 | 12.57 | 13.71 | 13.58 | 13.15 |
| | DLM Polinomial ordem 1 | 12.76 | 12.31 | 13.6 | 15.07 | 13.12 | 12.6 |
| MedAPE | <i>Damped Holt</i> | 4.42 | 4.67 | 4.96 | 6.51 | 5.66 | 5.84 |
| | Novo Método | 4.31 | 4.61 | 4.47 | 6.11 | 5.76 | 5.9 |
| | DLM Polinomial ordem 2 | 4.65 | 5.05 | 4.95 | 6.44 | 5.9 | 6.16 |
| | DLM Polinomial ordem 1 | 5.83 | 5.55 | 5.95 | 7.74 | 6.77 | 6.75 |
| MASE | <i>Damped Holt</i> | 1.03 | 1.1 | 1.23 | 1.55 | 1.6 | 1.71 |
| | Novo Método | 1.01 | 1.08 | 1.19 | 1.5 | 1.59 | 1.68 |
| | DLM Polinomial ordem 2 | 1.04 | 1.12 | 1.22 | 1.53 | 1.62 | 1.72 |
| | DLM Polinomial ordem 1 | 1.46 | 1.56 | 1.78 | 2.08 | 2.09 | 2.2 |

em terceiro lugar de acordo com o SMAPE.

Tabela 5.4: SMAPE, MedAPE e MASE de horizontes curto, médio e longo prazo para as 1428 séries mensais da competição M3, com multiprocessos classe I.

| Métrica | Método | Horizonte de Previsão | | | | | |
|---------|------------------------|-----------------------|--------------|--------------|--------------|--------------|--------------|
| | | 1-4 | 1-6 | 7-12 | 13-18 | 1-12 | 1-18 |
| SMAPE | <i>Damped Holt</i> | 12.44 | 12.55 | 13.76 | 18.63 | 13.15 | 14.98 |
| | Novo Método | 12.18 | 12.49 | 13.85 | 19.29 | 13.17 | 15.21 |
| | DLM Polinomial ordem 2 | 12.44 | 12.75 | 14.32 | 19.9 | 13.54 | 15.66 |
| | DLM Polinomial ordem 1 | 13.44 | 13.24 | 14.05 | 18.12 | 13.65 | 15.14 |
| MedAPE | <i>Damped Holt</i> | 5.13 | 5.34 | 6.6 | 9.54 | 5.97 | 7.16 |
| | Novo Método | 4.88 | 5.19 | 6.66 | 9.57 | 5.93 | 7.14 |
| | DLM Polinomial ordem 2 | 5.27 | 5.52 | 6.92 | 9.93 | 6.22 | 7.46 |
| | DLM Polinomial ordem 1 | 6.27 | 6.43 | 7.95 | 10.7 | 7.19 | 8.36 |
| MASE | <i>Damped Holt</i> | 1.23 | 1.37 | 2.08 | 3.25 | 1.73 | 2.24 |
| | Novo Método | 1.2 | 1.34 | 2 | 3.07 | 1.67 | 2.14 |
| | DLM Polinomial ordem 2 | 1.23 | 1.37 | 2.04 | 3.1 | 1.71 | 2.17 |
| | DLM Polinomial ordem 1 | 1.72 | 1.86 | 2.7 | 3.62 | 2.28 | 2.73 |

5.2.3 Resultados para todas as Séries onde a Tendência é forte

Alguns métodos podem ser muito úteis quando aplicados seletivamente a um conjunto apropriado de séries temporais. Já existe uma ampla literatura para protocolos de seleção

de modelos (Billah *et al.* [5], Tashman & Kruk [34], Taylor [36], Billah *et al.* [6]). Na competição M3, o novo modelo produziu melhores previsões pela métrica SMAPE em 707 das 1428 séries. Isto nos fez investigar as condições sob as quais o modelo teria uma performance boa em relação aos modelos tradicionais polinomiais e o *Damped Holt*.

Os modelos com amortecedores são conhecidos por performar bem quando a série possui tendência forte (Taylor [35], Gardner [12]). A fim de investigar mais a fundo a performance do modelo em situações onde o mesmo é considerado adequado, foram refeitos os cálculos das métricas para séries onde o amortecedor $\psi > 1$. O objetivo é quantificar os ganhos de performance em relação aos métodos tradicionais existentes e validar o potencial do modelo e o valor de se ter um amortecedor com dinâmica temporal.

Das 1428 séries mensais da competição, 153 tiveram o valor do amortecedor maior que um, segundo o método de estimação dos mínimos quadrados para o modelo *Damped Holt* tradicional. Curiosamente, isto contradiz o achado por Taylor [35], usando o mesmo método. Um valor maior do que um para ψ sugere que o *Damped Holt* não poderá fazer as previsões de modo satisfatório e que outros métodos podem ser melhores. Foi investigado se o novo modelo poderia performar melhor nestas condições, dado que o amortecedor agora é dinâmico e portanto teria mais chances de capturar o valor de ψ mais adequado. As Tabelas 5.5 e 5.6 sumarizam as medidas SMAPE, MedAPE e MASE para os quatros métodos em estudo aplicados ao subconjunto de 153 séries onde $\psi > 1$.

Tabela 5.5: SMAPE, MedAPE e MASE de horizontes 1-6 para as 153 séries mensais da competição M3 onde $\psi > 1$.

| Métrica | Método | Horizonte de Previsão | | | | | |
|---------|------------------------|-----------------------|--------------|--------------|--------------|--------------|--------------|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| SMAPE | <i>Damped Holt</i> | 13.56 | 14.79 | 14.53 | 17.51 | 16.99 | 12.62 |
| | Novo Método | 13.16 | 10.79 | 11.97 | 13.05 | 13.27 | 11.76 |
| | DLM Polinomial ordem 2 | 12.8 | 12.58 | 14.68 | 16.56 | 14.67 | 12.93 |
| | DLM Polinomial ordem 1 | 17.13 | 13.29 | 16.88 | 17.18 | 18.72 | 16.89 |
| MedAPE | <i>Damped Holt</i> | 5.17 | 5.34 | 3.92 | 5.56 | 5.89 | 6.05 |
| | Novo Método | 5.76 | 4.06 | 3.51 | 5.22 | 5.56 | 5.46 |
| | DLM Polinomial ordem 2 | 5.06 | 5.4 | 4.45 | 5.85 | 5.64 | 4.86 |
| | DLM Polinomial ordem 1 | 9.41 | 8.88 | 9.76 | 10.64 | 11.02 | 10.14 |
| MASE | <i>Damped Holt</i> | 0.86 | 0.84 | 0.8 | 1.16 | 1.15 | 1.22 |
| | Novo Método | 0.91 | 0.77 | 0.78 | 0.98 | 0.98 | 1.02 |
| | DLM Polinomial ordem 2 | 0.97 | 0.92 | 0.91 | 1.13 | 1.16 | 1.12 |
| | DLM Polinomial ordem 1 | 2.38 | 2.29 | 2.59 | 2.75 | 2.82 | 2.83 |

Para os horizontes 1 a 6, o novo método ficou em primeiro ou segundo lugar, ficando em primeiro nos horizontes 2-5 em todas as métricas. A prior performance ficou com o modelo polinomial de ordem um, como era de se esperar, pois este modelo não se adequa

bem a séries com forte tendência. O tradicional *Damped Holt* também performou bem nas previsões mais próximas, de horizontes 1-3.

Em todos os horizontes agregados de curto, médio e longo prazo, o novo modelo performou muito melhor que os outros modelos em todas as métricas usadas. O modelo polinomial de ordem dois ficou em segundo lugar em todas as métricas e em todos os horizontes considerados. A pior performance ficou com o modelo polinomial de ordem um, como era de se esperar para séries com tendência forte.

Tabela 5.6: SMAPE, MedAPE e MASE de horizontes curto, médio e longo prazo para as 153 séries mensais da competição M3 onde $\psi > 1$.

| Métrica | Método | Horizonte de Previsão | | | | | |
|---------|------------------------|-----------------------|--------------|--------------|-------------|--------------|--------------|
| | | 1-4 | 1-6 | 7-12 | 13-18 | 1-12 | 1-18 |
| SMAPE | <i>Damped Holt</i> | 15.1 | 15 | 14.61 | 20.35 | 14.8 | 16.65 |
| | Novo Método | 12.24 | 12.33 | 13.37 | 16.4 | 12.85 | 14.04 |
| | DLM Polinomial ordem 2 | 14.15 | 14.04 | 14.43 | 19.85 | 14.23 | 16.1 |
| | DLM Polinomial ordem 1 | 16.12 | 16.68 | 17.24 | 21.06 | 16.96 | 18.33 |
| MedAPE | <i>Damped Holt</i> | 5 | 5.32 | 6.66 | 8.97 | 5.99 | 6.99 |
| | Novo Método | 4.64 | 4.93 | 6.41 | 8.13 | 5.67 | 6.49 |
| | DLM Polinomial ordem 2 | 5.19 | 5.21 | 6.61 | 8.74 | 5.91 | 6.85 |
| | DLM Polinomial ordem 1 | 9.67 | 9.98 | 11.79 | 14.47 | 10.88 | 12.08 |
| MASE | <i>Damped Holt</i> | 0.91 | 1 | 1.6 | 2.41 | 1.3 | 1.67 |
| | Novo Método | 0.86 | 0.91 | 1.38 | 1.84 | 1.15 | 1.38 |
| | DLM Polinomial ordem 2 | 0.98 | 1.04 | 1.43 | 2.05 | 1.23 | 1.5 |
| | DLM Polinomial ordem 1 | 2.5 | 2.61 | 3.22 | 4.26 | 2.92 | 3.36 |

5.2.4 Resultados para todas as Séries onde AIC seleciona Modelos Ordem Dois

Seguindo a idéia da seção anterior, buscou-se as séries onde o *Akaike Information Criterion* (AIC) indicasse que os modelos polinomiais de ordem dois fossem adequados em relação aos modelos de ordem um. O critério AIC foi desenvolvido usando princípios de entropia da informação e possui ampla aceitação na literatura (Hyndman & Khandakar [20], Koehler & Murphree [24]). O AIC é uma medida de qualidade do ajuste de um modelo estatístico para um certo conjunto de dados. Como tal, pode ser usado para seleção de modelos. Esta medida lida com uma ponderação entre qualidade de ajuste e complexidade do modelo. O AIC oferece uma estimativa relativa da informação perdida quando um certo modelo é usado para representar o processo que gera os dados obser-

vados. Note que o AIC não consegue determinar se o modelo se ajusta bem aos dados de maneira absoluta. Existe uma versão para amostras finitas, denominada AIC corrigido (AICc) e esta é a versão que usou-se para determinar, dentre as 1428 séries mensais da competição, as séries onde o AICc julga que o modelo polinomial de ordem dois é mais adequado para descrever a série que o modelo polinomial de ordem um. Neste caso 644 séries foram selecionadas. Para permitir que os diferentes modelos se adaptassem aos dados, descartou-se os valores das cinco primeiras previsões. Calculou-se o valor das métricas nestas séries e o resultado é mostrado na Tabela 5.7 para os horizontes de curto prazo, 1 a 6.

Tabela 5.7: SMAPE, MedAPE e MASE de horizontes 1-6 para as 644 séries mensais da competição M3 onde o AICc seleciona o modelo de ordem dois.

| Métrica | Método | Horizonte de Previsão | | | | | |
|---------|------------------------|-----------------------|-------------|-------------|-------------|-------------|-------------|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| SMAPE | <i>Damped Holt</i> | 3.72 | 4.31 | 4.54 | 6.19 | 7.29 | 7.43 |
| | Novo Método | 3.34 | 3.85 | 4.22 | 5.92 | 6.37 | 6.78 |
| | DLM Polinomial ordem 2 | 3.79 | 4.27 | 4.83 | 6.5 | 6.8 | 7.19 |
| | DLM Polinomial ordem 1 | 5.79 | 6.46 | 7.71 | 8.47 | 9.46 | 9.76 |
| MedAPE | <i>Damped Holt</i> | 1.01 | 1.52 | 1.82 | 2.47 | 2.92 | 3.18 |
| | Novo Método | 0.89 | 1.35 | 1.73 | 2.47 | 2.78 | 3.17 |
| | DLM Polinomial ordem 2 | 1.21 | 1.66 | 1.74 | 2.49 | 2.74 | 2.99 |
| | DLM Polinomial ordem 1 | 2.21 | 2.75 | 3.33 | 4.29 | 4.32 | 5.05 |
| MASE | <i>Damped Holt</i> | 0.92 | 1.27 | 1.56 | 2.22 | 2.53 | 2.81 |
| | Novo Método | 0.82 | 1.12 | 1.41 | 2.02 | 2.3 | 2.59 |
| | DLM Polinomial ordem 2 | 1.07 | 1.31 | 1.55 | 2.1 | 2.36 | 2.62 |
| | DLM Polinomial ordem 1 | 2.29 | 2.65 | 3.04 | 3.52 | 3.82 | 4.13 |

Segundo todas as métricas, o novo modelo é o melhor nos horizontes de 1 a 4. De acordo com o SMAPE e MASE, ele é melhor nos horizontes 5 a 6 e fica em segundo no MedAPE, o que é um bom indicativo de sua performance no curto prazo.

Valores médios para os horizontes de curto, médio e longo prazo são sumarizados na Tabela 5.8. Agora pode-se notar que no curto prazo, horizontes de 1 a 4 e 1 a 6, para todas as métricas, o novo modelo performa melhor. Para o médio prazo ele ainda é o melhor de acordo com o SMAPE e fica em segundo de acordo com o MedAPE e MASE. No longo prazo, horizontes 13-18, ele fica em segundo lugar no SMAPE e em terceiro no MASE e MedAPE. No geral, nos horizontes 1-12 o novo modelo fica em primeiro ou segundo lugar e no horizonte completo, 1-18, performa melhor de acordo com o SMAPE e em segundo de acordo com o MedAPE e o MASE. A pior performance fica, como era de se esperar, com o modelo polinomial de ordem um. O modelo polinomial de ordem dois performou bem no médio e longo prazo de acordo com o MedAPE e o MASE.

Tabela 5.8: SMAPE, MedAPE e MASE de horizontes curto, médio e longo prazo para as 644 séries mensais da competição M3 onde o AICc seleciona o modelo de ordem dois.

| Métrica | Método | Horizonte de Previsão | | | | | |
|---------|------------------------|-----------------------|-------------|-------------|--------------|-------------|-------------|
| | | 1-4 | 1-6 | 7-12 | 13-18 | 1-12 | 1-18 |
| SMAPE | <i>Damped Holt</i> | 4.69 | 5.58 | 8.49 | 12.26 | 7.04 | 8.78 |
| | Novo Método | 4.33 | 5.08 | 8.46 | 12.57 | 6.77 | 8.7 |
| | DLM Polinomial ordem 2 | 4.85 | 5.56 | 8.51 | 13.35 | 7.04 | 9.14 |
| | DLM Polinomial ordem 1 | 7.11 | 7.94 | 10.64 | 14 | 9.29 | 10.86 |
| MedAPE | <i>Damped Holt</i> | 1.71 | 2.15 | 3.93 | 6.01 | 3.04 | 4.03 |
| | Novo Método | 1.61 | 2.06 | 3.76 | 6.11 | 2.91 | 3.98 |
| | DLM Polinomial ordem 2 | 1.78 | 2.14 | 3.42 | 5.6 | 2.78 | 3.72 |
| | DLM Polinomial ordem 1 | 3.14 | 3.66 | 5.88 | 8.74 | 4.77 | 6.09 |
| MASE | <i>Damped Holt</i> | 1.49 | 1.89 | 3.44 | 5.2 | 2.66 | 3.51 |
| | Novo Método | 1.34 | 1.71 | 3.32 | 5.24 | 2.51 | 3.42 |
| | DLM Polinomial ordem 2 | 1.51 | 1.84 | 3.18 | 5.19 | 2.51 | 3.4 |
| | DLM Polinomial ordem 1 | 2.87 | 3.24 | 4.96 | 6.82 | 4.1 | 5.01 |

Capítulo 6

Conclusão e Trabalhos Futuros

6.1 Conclusão

Neste trabalho, apresentamos uma nova abordagem para modelos lineares dinâmicos polinomiais de ordem dois com abafamento da tendência, adotando o enfoque bayesiano.

A inovação deste está na atribuição de uma estrutura dinâmica na evolução do amortecedor de tendência. Em toda literatura pesquisada, o parâmetro de abafamento é fixado, sendo estimado de forma a minimizar alguma medida de eficiência. Desta forma, esta proposta é inédita. As principais vantagens deste método são:

1. O modelo possui solução analítica e não existe a necessidade de realizar simulações para obter os momentos da densidade *posteriori* dos estados ou do amortecedor via técnicas computacionalmente exigentes, tais como *Markov Chain Monte Carlo* ou o amostrador de *Gibbs*. Isto permite que o modelo seja empregado para prever o desenvolvimento de várias séries simultaneamente e de maneira automatizada e em larga escala, como é o caso de preços de ativos financeiros listados em bolsa ou itens de monitoramento *online*.
2. Como o modelo assume uma distribuição normal, usando o teorema da superposição (veja West & Harrison [39], Capítulo 6), ele pode ser facilmente incorporado a outros componentes do DLM para a criação de um modelo adequado para a série em estudo. Desta forma, o modelo proposto pode ser mais uma opção a ser usada na definição de DLMs para previsão de séries temporais.
3. Em geral, o modelo é mais preciso para previsões de curto prazo que o modelo polinomial de ordem dois existente e pode ser considerado como uma alternativa

em casos onde deseja-se adicionar um componente para modelar a tendência. Note também que o componente pode lidar com vários tipos de tendência, sejam elas linear ou exponencial. Os resultados das simulações e o estudo de caso da competição internacional M3 ajudam a confirmar esta hipótese.

Nos estudos de simulação, o método proposto nesse trabalho mostra desempenho similar ou melhor ao DLM polinomial de ordem dois já existente. Ao simular séries temporais com diferentes valores no espaço paramétrico, foi possível observar o comportamento do novo modelo em relação ao modelo polinomial de ordem dois tradicional. Neste caso os resultados sugerem que o novo modelo pode apresentar capacidade preditiva superior em parte do espaço paramétrico. Observou-se também que a performance preditiva do novo modelo melhora muito conforme o valor de ψ aumenta ou conforme a razão entre as variâncias observacional e dos estados, denotada por $W_{11,t}/V_t$, aumenta. Isto permite identificar as condições onde o modelo tem maior chance de performar muito acima da média dos modelos lineares já existentes. Os resultados obtidos permitem recomendar o modelo para modelar séries com tendência forte ou onde o critério AIC corrigido sugere o uso de modelos de ordem dois.

De um modo geral, considerando os resultados obtidos na competição M3, o novo método teve um desempenho satisfatório e conseguiu ter performance melhor que o modelo *Damped Holt* tradicional em quase metade das séries da competição internacional M3. Para horizontes de curto e médio prazo o modelo foi ligeiramente melhor que os modelos polinomiais tradicionais e o modelo *Damped Holt*. Este resultado, confirmado por todas as métricas usadas, pode indicar que o novo modelo é uma alternativa a se considerar ao modelar séries com tendência.

É importante considerar, no entanto, que o novo método foi aplicado de forma indiscriminada em todas as séries da competição. Em muitas séries, é razoável supor que o modelo polinomial de ordem dois não é adequado para descrever o comportamento da série e portanto não se espera que as previsões do modelo sejam boas. Seria interessante desenvolver ou aplicar metodologias para seleção e identificação de modelos para séries temporais, de forma que os modelos sejam aplicados somente quando considerados adequados. Acredita-se que os resultados podem melhorar consideravelmente.

Outra limitação é o fato do modelo ser univariado e não poder ser empregado para previsão conjunta de séries correlacionadas. Não foi possível desenvolver um modelo, com solução analítica, para séries multivariadas, devido às limitações de tempo. Pode-se argumentar que mesmo que as séries tenham correlação alta, é bem provável que a dinâmica do amortecedor ψ_t seja similar para cada uma delas e provavelmente a evolução dos

amortecedores também tenha correlação. Já existem na literatura modelos multivariados para séries com amortecedores constantes no tempo, veja (Hyndman *et al.* [19]).

6.2 Trabalhos Futuros

Nesta seção são apresentadas algumas idéias para futuras pesquisas relacionadas com o que foi desenvolvido neste trabalho.

- Fazer novos estudos com diferentes formulações na evolução temporal do amortecedor $\psi_t | \psi_{t-1}$ e estudar o impacto destas na capacidade preditiva do modelo. Um aspecto importante é poder capturar o valor adequado do amortecedor. Por exemplo, se a série é um passeio aleatório, então a dinâmica do amortecedor deve ser tal que o valor de ψ_t decaia para zero, degenerando para o modelo polinomial de ordem um. Uma proposta para a nova *priori* que se adequa a este caso é dada abaixo:

$$\begin{aligned} p_t &= \frac{m_{2,t}}{m_{2,t} + 1}(u_t + r_t b_t) + \frac{m_{2,t-1}}{m_{2,t} + 1} r_t (1 - p_{t-1} b_t) \\ q_t &= 1 + s_t + r_t^2 h_t. \end{aligned}$$

Caso o estado atinja valores perto de -1 de forma súbita, o valor de p_t calculado é descartado e adota-se o último valor válido. Mais estudos são necessários para validar esta nova proposta de hiperparâmetros da *priori* para a evolução do amortecedor.

- Refazer os testes com o modelo multiplicativo, *Damped Pegels* dinâmico, no lugar do modelo aditivo e quantificar os possíveis ganhos de performance com esta formulação.
- Avaliar a relação entre o coeficiente adaptativo u_t e a acurácia de previsão fora da amostra de treinamento, de modo que haja uma metodologia para a estimação de u_t que gera previsões mais precisas, principalmente no curto prazo. Outra opção é estudar o emprego de multiprocessos classe I ou II no novo modelo e realizar as previsões usando a mistura de densidades resultante, ao invés de usar apenas um único modelo com parâmetros otimizados.
- Estender a formulação atual para o caso multivariado, de forma que séries correlacionadas possam influenciar na evolução temporal do amortecedor e trazer ganhos preditivos ao modelo. É muito provável que neste caso a formulação não tenha mais solução analítica e métodos de simulação sejam necessários para obter uma aproximação da distribuição a *posteriori* dos parâmetros.

- Comparar a formulação com dados mais recentes vindos da nova competição M4, ainda em fase de desenvolvimento e avaliar se a performance preditiva se mantém.

Apêndice A

Lemas empregados

A.1 Teoria de Distribuição Normal

As duas proposições abaixo são usadas na prova das equações de atualização dos estados, também chamado de filtro de *Kalman*. Ambas usam conceitos da teoria de distribuição normal multivariada.

Lema A.1. *Sejam G uma matrix $d \times s$. Se $X|Y \sim N_d(\mathbf{G}Y, \mathbf{W})$ e $Y \sim N_s(\mathbf{m}, \mathbf{C})$ então*

$$X \sim N_d(\mathbf{a}, \mathbf{R}), \quad (\text{A.1})$$

onde $\mathbf{a} = \mathbf{G}\mathbf{m}$ e $\mathbf{R} = \mathbf{G}\mathbf{C}\mathbf{G}' + \mathbf{W}$.

O resultado usa conceitos padrões da teoria de distribuição normal, veja Johnson & Wichern [22] ou Anderson [1].

O lema abaixo permite a reconstrução das condicionais dada a densidade conjunta.

Lema A.2. *Sejam $X|Y \sim N_{d_1}(\boldsymbol{\mu}_1 + \mathbf{G}(Y - \boldsymbol{\mu}_2), \mathbf{W})$ e $Y \sim N_{d_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$, então*

$$\mathbf{Z} = \begin{pmatrix} X \\ \dots \\ Y \end{pmatrix} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (\text{A.2})$$

com

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \dots \\ \boldsymbol{\mu}_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \vdots & \boldsymbol{\Sigma}_{12} \\ \dots & \dots & \dots \\ \boldsymbol{\Sigma}_{21} & \vdots & \boldsymbol{\Sigma}_{22} \end{pmatrix}, |\boldsymbol{\Sigma}_{22}| > 0, d = d_1 + d_2, \boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}'_{12}, \quad (\text{A.3})$$

e

$$\Sigma_{11} = \mathbf{W} + \mathbf{G}\Sigma_{22}\mathbf{G}', \quad \Sigma_{12} = \mathbf{G}\Sigma_{22}. \quad (\text{A.4})$$

Alternativamente, se $\mathbf{Z} \sim N_d(\boldsymbol{\mu}, \Sigma)$ então $\mathbf{Y}|\mathbf{X} \sim N_{d_2}(\boldsymbol{\mu}_{y|x}, \Sigma_{y|x})$, onde

$$\boldsymbol{\mu}_{y|x} = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{X} - \boldsymbol{\mu}_1), \quad (\text{A.5})$$

$$\Sigma_{y|x} = \Sigma_{22} - \Sigma'_{12}\Sigma_{11}^{-1}\Sigma_{12}. \quad (\text{A.6})$$

O resultado usa conceitos padrões da teoria de distribuição normal, veja Johnson & Wichern [22] ou Anderson [1].

Referências Bibliográficas

- [1] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. Second Edition. John Wiley, New York.
- [2] Armstrong, J. (2007). Significance tests harm progress in forecasting. *Internacional Journal of Forecasting*, (23), 321–327.
- [3] Armstrong, J. & Collopy, F. (1992). Error Measures For Generalizing About Forecasting Methods: Empirical Comparisons. *Internacional Journal of Forecasting*, (8), 69–80.
- [4] Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* Second Edition. Springer–Verlag, New York.
- [5] Billah, B., King, M., Snyder, R. & Koehler, A. (2005). Exponential Smoothing Model Selection for Forecasting. *Working Paper 6/05 Monash University*.
- [6] Billah, M., Hyndman, R. & Koehler, A. (2003). Empirical Information Criteria for Time Series Forecasting Model Selection. *Working Paper 02/03 Monash University*.
- [7] Carmona, R. A. (2004). *Statistical analysis of financial data in S-plus*. Springer–Verlag, New York.
- [8] Chatfield, C. & Yar, M. (1988). Holt-Winters Forecasting: Some Practical Issues. *Journal of the Royal Statistical Society*, **37**(2), 129–140.
- [9] Cronea, S., Hibon, M. & Nikolopoulos, K. (2011). Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *International Journal of Forecasting*, (27), 635–660.
- [10] Durbin, J. & Koopman, S. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford.

- [11] Flores, B. (1986). A Pragmatic View of Accuracy Measurement in Forecasting. *Internacional Journal of Management Science*, **14**(2), 93–98.
- [12] Gardner, E. (2006). Exponential smoothing: The state of the art Part 2. *Internacional Journal of Forecasting*, (22), 637–666.
- [13] Gardner, E. & McKenzie, E. (2009). Why the damped trend works. *Technical Report University of Houston*.
- [14] Goodwin, P. & Lawton, R. (1999). On the asymmetry of the symmetric MAPE. *Internacional Journal of Forecasting*, (15), 405–408.
- [15] Gooijer, J. & Hyndman, R. (2005). 25 Years of IIF Time Series Forecasting: A Selective Review. *Working Paper 12/05 Monash University*.
- [16] Hang, D. (2013). High-Frequency Bayesian Modeling and Analysis of Stochastic Volatility in Finance. *Duke University*. <http://www.stat.duke.edu/research/BEST/BEST2010-11/BESThonmention2010-11.Derrick.pdf>.
- [17] Harvey, A. (1989). *Forecasting, Structural Time Series Models and the Kalman filter*. Cambridge University Press, Cambridge.
- [18] Hyndman, R. & Koehler, A. (2006). Another look at measures of forecast accuracy. *Internacional Journal of Forecasting*, (22), 679–688.
- [19] Hyndman, R., Koehler, A., Ord, J. & Snyder, R. (2008). *Forecasting with Exponential smoothing*. Springer, Berlin.
- [20] Hyndman, R. J. & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, **27**(3), 1–22.
- [21] James, B. R. (1981). *Probabilidade: Um Curso em Nível Intermediário*. IMPA–Projeto Euclides, Rio de Janeiro, RJ.
- [22] Johnson, R. A. & Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*. Third Edition. Prentice Hall, New Jersey.
- [23] Karlon, K. (2006). *Determining Optimal architecture for DLMs in Time Series Applications*. Master Thesis - University of North Carolina Wilmington, Wilmington.
- [24] Koehler, A. & Murphree, E. (1988). A Comparison of the Akaike and Schwarz Criteria for Selecting Model Order. *Journal of the Royal Statistical Society*, **37**(2), 187–195.

- [25] Koninga, A., Fransesa, P., Hibon, M., Stekle, H., Goodwin, P. & Lawton, R. (2005). The M3 competition: Statistical tests of the results. *Internacional Journal of Forecasting*, (21), 397–409.
- [26] Makridakis, S. & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *Internacional Journal of Forecasting*, (16), 451–476.
- [27] McKenzie, E. & Gardner, E. (2010). Damped trend exponential smoothing: A modelling viewpoint. *Internacional Journal of Forecasting*, (26), 661–665.
- [28] Migon, H., Gamerman, D., Lopez, H. & Ferreira, M. (2005). Bayesian dynamic models. In D. Day & C. Rao, editors, *Handbook of Statistics*, volume 25, chapter 19, pages 553–588. Elsevier B.V., New York.
- [29] Petris, G., Petrone, S. & Campagnoli, P. (2009). *Dynamic Linear Models with R*. Springer, New York.
- [30] Pole, A., West, M. & Harrison, J. (1994). *Applied Bayesian forecasting and time series analysis*. Chapman & Hall, New York.
- [31] Robert, C. & Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York.
- [32] Solhjell, I. (2009). *Bayesian Forecasting and Dynamic Models applied to strain data from the GÅŪTA River Bridge*. Master Thesis - University of Oslo, Oslo.
- [33] Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Third Edition. Springer, New York.
- [34] Tashman, L. & Kruk, J. (1996). The use of protocols to select exponential smoothing procedures: a reconsideration of forecasting competitions. *International Journal of Forecasting*, (12), 235–253.
- [35] Taylor, J. (2003). Exponential Smoothing with a damped multiplicative trend. *Internacional Journal of Forecasting*, (19), 715–725.
- [36] Taylor, J. (2008). Exponentially weighted information criteria for selecting among forecasting models. *International Journal of Forecasting*, (24), 513–524.
- [37] Venables, W. & Ripley, B. (2002). *Modern Applied Statistics with S* Second Edition. Springer, New York.
- [38] Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer, USA.

- [39] West, M. & Harrison, J. (1997). *Dynamic Models and Bayesian Forecasting*. Springer, New York.
- [40] Yelland, P. & Lee, E. (2003). Forecasting Product Sales with Dynamic Linear Mixture Models. *Technical Report SUN Microsystems SMLI TR-2003-122*.
- [41] Zhai, Y. (2005). *Time Series Forecasting Competition among three sophisticated Paradigms*. Master Thesis - University of North Carolina Wilmington, Wilmington.