

UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
DEPARTAMENTO DE ESTATÍSTICA  
CURSO DE ESTATÍSTICA

Isabela Abreu Cutrim

**Uma Revisão de Modelos Lineares Dinâmicos**

Juiz de Fora

2014

Isabela Abreu Cutrim

## Uma Revisão de Modelos Lineares Dinâmicos

Monografia apresentada ao curso de Estatística da Universidade Federal de Juiz de Fora, como requisito para a obtenção do grau de Bacharel em Estatística.

**Orientador: Joaquim Henriques Vianna Neto**  
**Doutor em Estatística – UFRJ**

Juiz de Fora

2014

Isabela Abreu Cutrim

Uma Revisão de Modelos Lineares Dinâmicos

Monografia apresentada ao curso de Estatística da Universidade Federal de Juiz de Fora, como requisito para a obtenção do grau de Bacharel em Estatística.

Aprovada em 14 de julho de 2014

BANCA EXAMINADORA

---

Joaquim Henriques Vianna Neto (Orientador)

Doutor em Estatística – UFRJ

---

Augusto Carvalho de Souza

Mestre em Estudos Populacionais e Pesquisas Sociais - ENCE

---

Alfredo Chaoubah

Doutor em Engenharia Elétrica – PUC-Rio

## **Agradecimentos**

Agradeço à minha mãe Eliana, que sem ela nada disto seria possível, que ininterruptamente me deu base e apoio. Agradeço à minha irmã Isadora, que sempre dividiu, não só, o ventre e o colo da mãe, mas como tudo nessa vida, alegrias, medos e conquistas. Agradeço à Pedrina e à Nilza pela preocupação e carinho.

Agradeço ao Joaquim pela orientação e paciência. Agradeço aos amigos Camila, Carolina, Vítor, Anna e Franciele, por terem começado essa jornada ao meu lado, pelas dúvidas divididas e por poder aprender e ensinar juntos. Agradeço à minha nova turma, com a qual cursei meus dois últimos anos, pelo ânimo renovado.

E por fim, agradeço ao meu marido e companheiro Pedro, por me animar e nunca me deixar desistir e por dividir a vida comigo.

## **Resumo**

A modelagem dinâmica é uma proposta de análise e previsão de séries temporais do ponto de vista Bayesiano. Este trabalho de conclusão de curso tem como objetivo apresentar os Modelos Lineares Dinâmicos (MLD), seu uso e contexto de aplicação. São mostradas e demonstradas propriedades dos MLDs, como o Filtro de Kalman, a suavização, a previsão um passo à frente e, também,  $k$  passos à frente. São revisados métodos de Monte Carlo via Cadeia de Markov (MCMC) em algoritmos recursivos usados para estimar parâmetros, são eles: amostrador de Gibbs, Metropolis-Hastings e FFBS.

**Palavras-chave:** Modelos Lineares Dinâmicos, MLD, Séries Temporais, Filtro de Kalman, Suavização, Previsão, Inferência Bayesiana, Amostrador de Gibbs, Metropolis-Hastings, FFBS.

## **Abstract**

The state space models are a proposal for time series analysis and forecasting in viewpoint Bayesian. This course conclusion work aims to present the Dynamics Linear Models (DLM), its use and application context. DLM's properties are shown and demonstrated, like Kalman filter, the smoothing, the forecasting one-step-ahead and so the forecasting  $k$  steps ahead. Markov chain Monte Carlo methods (MCMC) are reviewed in recursive algorithms used to estimate parameters, they are: Gibbs sampler, Metropolis-Hastings e FFBS.

**Keywords:** Dynamics Linear Models, DLM, Time Series, Kalman Filter, Smoothing, Forecasting, Bayesian Inference, Gibbs Sampler, Metropolis-Hastings, FFBS.

## Sumário

1. Introdução .....	8
2. ....	11
2.1. Modelos Dinâmicos .....	11
2.2. Modelos Lineares Dinâmicos .....	12
2.3. Exemplos de Modelos Lineares Dinâmicos .....	13
2.3.1. Passeio aleatório mais ruído .....	13
2.3.2. Modelo de crescimento linear .....	14
2.4. Modelo de Regressão Linear .....	15
2.5. Um MLD que generaliza um Modelo de Regressão Linear .....	17
2.6. Filtro de Kalman .....	18
2.6.1. Filtro de Kalman para MDG .....	19
2.6.2. Filtro de Kalman para MLD .....	21
2.7. Suavização .....	24
2.7.1. Suavização Recursiva para MDG .....	25
2.7.2. Suavizador de Kalman .....	26
2.8. Previsão $k$ Passos à Frente .....	27
2.8.1. Recursão Preditiva para MDG .....	28
2.8.2. Recursão Preditiva para MLD .....	29
2.9. Checagem do Modelo .....	30
3. Modelos com componentes desconhecidas .....	33
3.1. Fator de Desconto .....	34
3.2. Métodos de Monte Carlo via Cadeias de Markov .....	35
3.2.1. Amostrador de Gibbs .....	35
3.2.2. Metropolis-Hastings .....	36
3.2.3. Amostrador de Gibbs para MLD .....	38
3.2.4. FFBS .....	39
4. Conclusão .....	41
Referências bibliográficas .....	42

## Capítulo 1

### Introdução

No princípio da década de 60, na engenharia, que se iniciou a Modelagem Dinâmica, também conhecida como modelos de espaço de estados, primeiramente com Kalman (1960) através de um novo ponto de vista e posteriormente pela solução encontrada por Kalman e Bucy (1963), nomeada Filtro de Kalman. Porém, estas descobertas só foram utilizadas por estatísticos cerca de dez anos depois (PETRIS et al. 2009) e, a partir de então, o interesse e o uso de Modelos Dinâmicos na análise de séries temporais vem crescendo consideravelmente.

Modelos Dinâmicos proporcionam a modelagem de uma série temporal usando a combinação de diversas componentes, como tendência, sazonalidade e componente regressiva. Ao mesmo tempo, têm uma elegante e poderosa estrutura probabilística que evolui ao longo do tempo, garantindo aos modelos flexibilidade e aplicabilidade em uma ampla variedade de problemas nas mais diversas áreas de conhecimento.

Os modelos dinâmicos consideram uma série afetada pelo tempo, através de deformações dinâmicas e aleatórias. Eles seguem a interpretação natural das séries temporais combinando componentes sazonais ou regressivos. E ao mesmo tempo, com uma estrutura probabilística robusta, oferecendo um sistema flexível com inúmeras aplicações. Os problemas de estimação e inferência são resolvidos por algoritmos recursivos, que seguem naturalmente o enfoque Bayesiano.

Os modelos de espaço de estados podem ser usados em séries temporais univariadas ou multivariadas, também na presença de mudanças estruturais não estacionárias e padrões irregulares. Ao se deparar com uma série bastante previsível, onde seu comportamento repete ao longo do tempo, segue uma tendência e possui um componente sazonal bem regular, como a série exibida na Figura 1.1, um modelo simples considerando a tendência e a sazonalidade, seria satisfatório. De fato, uma análise de série temporal básica depende da possibilidade de encontrar uma regularidade razoável no comportamento do fenômeno estudado, prevendo assim o comportamento futuro. No entanto, em séries temporais onde claramente há uma mudança na sazonalidade e/ou do nível, como as séries exibidas nas Figuras 1.2 e 1.3, respectivamente, sentimos a necessidade de modelos mais flexíveis, que não assumam um padrão regular nem um sistema estável, mas que se ajustem bem aos dados, contemplando possíveis instabilidades no nível ou na variância, pontos de mudança, desvios estruturais ou

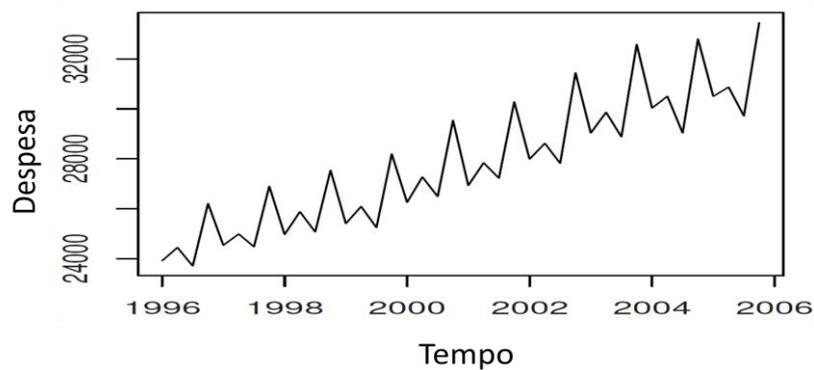


Figura 1.1<sup>1</sup> – Despesa familiar em comida, do 1º trimestre de 1996 ao 4º de 2005.

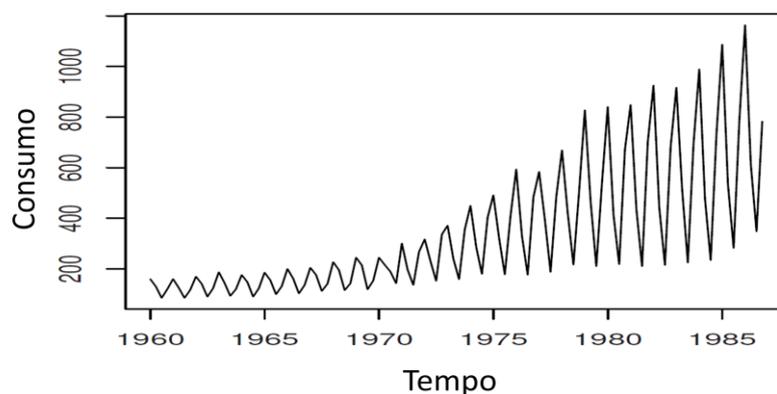


Figura 1.2<sup>2</sup> - Consumo de gás no Reino Unido, do 1º trimestre de 1960 ao 4º de 1986.

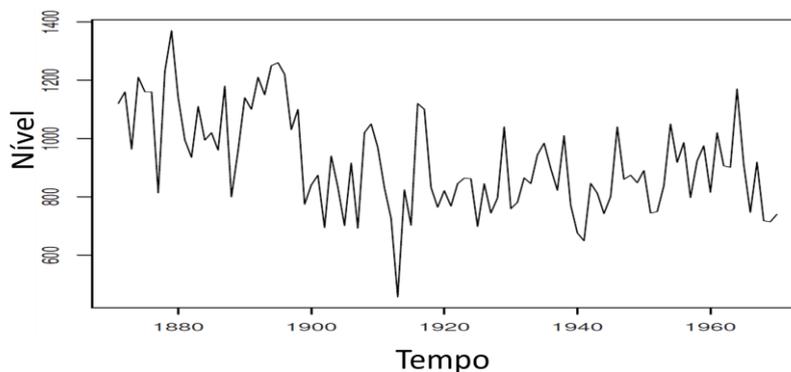


Figura 1.3<sup>3</sup> – Medidas anuais do nível do rio Nilo em Ashwam, de 1871 a 1970.

saltos repentinos. Os modelos de espaço de estados possuem esta flexibilidade e podem ser aplicados, por exemplo, a séries não estacionárias, como a exibida na Figura 1.4, sem a necessidade de qualquer transformação preliminar dos dados.

<sup>1</sup> PETRIS et al, 2009

<sup>2</sup> PETRIS et al, 2009

<sup>3</sup> PETRIS et al, 2009

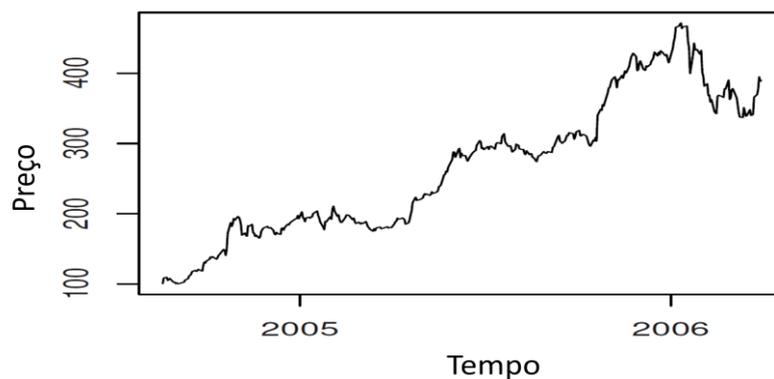


Figura 1.4<sup>4</sup> - Preço diário do Google Inc.

Os problemas de estimação e previsão na modelagem dinâmica são resolvidos por algoritmos recursivos, calculando as distribuições condicionais das quantidades de interesse dadas as informações observadas. Neste sentido, estes modelos são naturalmente tratados seguindo o paradigma Bayesiano.

O objetivo deste trabalho é fazer uma revisão de uma classe específica de modelos dinâmicos, chamada de Modelos Lineares Dinâmicos (MLD), que assumem uma suposição de normalidade da variável dependente e das suas perturbações. Trataremos, no capítulo 2, de apresentar e demonstrar a modelagem matematicamente e como ela nos possibilita passear entre os dados e incorporar os eventos aleatórios associados à passagem do tempo. Sem nos aprofundarmos, faremos também uma breve apresentação dos Modelos Dinâmicos Generalizados (MDG), que são uma extensão dos MLD, porém sem a suposição de normalidade. No capítulo 3, apresentaremos técnicas e algoritmos recursivos para o caso dos MLD com parâmetros desconhecidos, com os quais, geralmente, nos deparamos. Por fim, no capítulo 4, apresentaremos as conclusões.

---

<sup>4</sup> PETRIS et al, 2009

## Capítulo 2

### 2.1. Modelos Dinâmicos

Formalmente, um modelo dinâmico é formado por uma série temporal em  $(\theta_t)_{t \geq 0}$  em  $\mathbb{R}^p$  e uma série temporal em  $(Y_t)_{t \geq 1}$  em  $\mathbb{R}^n$ , que satisfaz as seguintes suposições:

- $(\theta_t)_{t \geq 0}$  é uma cadeia de Markov, ou seja,

$$f(\theta_t | \theta_1, \dots, \theta_{t-1}) = f(\theta_t | \theta_{t-1}) \text{ para } t > 1. \quad (2.1)$$

- Condicionado a  $(\theta_t)_{t \geq 0}$ , os  $Y_t$  são independentes e  $Y_t$  depende somente de  $\theta_t$ , ou seja,

$$f(y_1, \dots, y_t | \theta_0, \theta_1, \dots, \theta_t) = \prod_{i=1}^t f(y_i | \theta_0, \theta_1, \dots, \theta_t) = \prod_{i=1}^t f(y_i | \theta_i), \text{ para } t > 1. \quad (2.2)$$

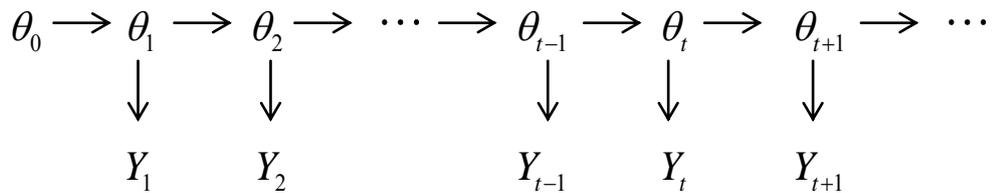
Como consequência de (2.1), usando o teorema da multiplicação, temos que

$$f(\theta_1, \dots, \theta_t) = \left( \prod_{i=1}^t f(\theta_i | \theta_{i-1}) \right) f(\theta_0), \text{ para } t > 1. \quad (2.3)$$

Agora, como consequência de (2.2) e (2.3), o modelo espaço de estado é completamente especificado pela distribuição  $f(\theta_0)$  e pelas distribuições condicionais  $f(\theta_t | \theta_{t-1})$  e  $f(y_t | \theta_t)$  para  $t \geq 1$ . De fato, pelo teorema da multiplicação, para  $t \geq 1$

$$\begin{aligned} f(\theta_0, \dots, \theta_t, y_1, \dots, y_t) &= f(y_1, \dots, y_t | \theta_0, \dots, \theta_t) f(\theta_0, \dots, \theta_t) \\ &= (\text{pela equação 2.2}) = \\ &= \left( \prod_{i=1}^t f(y_i | \theta_i) \right) f(\theta_0, \dots, \theta_t) \\ &= (\text{pela equação 2.3}) = \\ &= \left( \prod_{i=1}^t f(y_i | \theta_i) \right) \left( \prod_{i=1}^t f(\theta_i | \theta_{i-1}) \right) f(\theta_0) \end{aligned}$$

O fluxo de informação assumido pelo modelo espaço de estado é representado na figura abaixo.



## 2.2. Modelos Lineares Dinâmicos

Uma classe específica de modelos dinâmicos é a de modelos de espaço de estados Gaussianos e lineares, mais conhecidos como Modelos Lineares Dinâmicos (MLD). Um MLD é especificado por uma distribuição a priori normal para o vetor de estados no instante zero,

$$\theta_0 \sim N(m_0; C_0),$$

Juntamente com um par de equações para cada tempo  $t \geq 1$ :

- Equação das observações:

$$Y_t = F_t \theta_t + v_t, \text{ onde } v_t \sim N(0; V_t), \quad (2.4)$$

- Equação do sistema:

$$\theta_t = G_t \theta_{t-1} + w_t, \text{ onde } w_t \sim N(0; W_t). \quad (2.5)$$

Sendo, para  $t \geq 1$ ,

- $\theta_t$  é um vetor  $p$ -dimensional denominado parâmetro de estado do modelo dinâmico;
- $F_t$  é uma matriz de regressão ( $n \times p$ ), cujos elementos são valores conhecidos;
- $G_t$  é uma matriz ( $p \times p$ ) conhecida que descreve a evolução temporal dos parâmetros de estado;
- $V_t$  é uma matriz de covariância ( $n \times n$ ) conhecida associada ao erro observacional  $v_t$ ;
- $W_t$  é uma matriz de covariância ( $p \times p$ ) conhecida associada ao erro de evolução dos parâmetros de estado  $w_t$ .

Considerando  $t = 1, 2, \dots$ , as equações (2.4) e (2.5) podem ser reescritas da seguinte forma, respectivamente:

- Equação das observações:

$$Y_t | \theta_t \sim N(F_t \theta_t; V_t) \text{ e}$$

- Equação do sistema:

$$\theta_t | \theta_{t-1} \sim N(G_t \theta_{t-1}; W_t).$$

A suposição de normalidade é sensível em muitas aplicações e pode ser justificada por argumentos que utilizem o Teorema do Limite Central. No entanto, existem extensões importantes, como erros de caudas pesadas para modelar outliers ou os modelos dinâmicos generalizados, usados dentre outras aplicações para tratar séries temporais discretas. O preço a

ser pago ao remover a suposição de normalidade é aumentar a dificuldade computacional (PETRIS et al. 2009).

### 2.3. Exemplos de Modelos Lineares Dinâmicos

Vamos apresentar agora alguns exemplos de MLD. Nosso objetivo é ilustrar como a estrutura descrita em (2.4) e (2.5) pode caracterizar uma grande variedade de modelos.

#### 2.3.1. Passeio aleatório mais ruído

O modelo mais simples para uma série temporal univariada é o passeio aleatório mais ruído, definido por

$$Y_t = \mu_t + v_t, \text{ onde } v_t \sim N(0; V) \text{ e}$$
$$\mu_t = \mu_{t-1} + w_t, \text{ onde } w_t \sim N(0; W).$$

Repare que este é um MLD com:

- $n = p = 1$ ,
- $\theta_t = \mu_t$  para  $t \geq 1$ ,
- $F_t = G_t = 1$  para  $t \geq 1$ ,
- $V_t = V$  para  $t \geq 1$  e
- $W_t = W$  para  $t \geq 1$ .

Intuitivamente, este modelo é apropriado para séries temporais que não possuem variação de tendência ou sazonalidade. O modelo assume que as observações provêm de um ruído mais um nível  $\mu_t$  que, por sua vez, está sujeito a perturbações aleatórias ao longo do tempo. Este modelo é também chamado de modelo de nível local. Repare ainda que se  $W$  for aproximadamente zero, temos aproximadamente um modelo normal de média e variância constantes.

Com o intuito de exemplificar, apresentaremos nas Figuras 2.1 e 2.2 o comportamento deste modelo. Os dados deste exemplo correspondem às vendas mensais de uma bala (SALES), de janeiro de 1976 a novembro de 1981, do arquivo CANDY.DAT do pacote BATS, no programa R, realizado por Migon e Gamerman (2004).

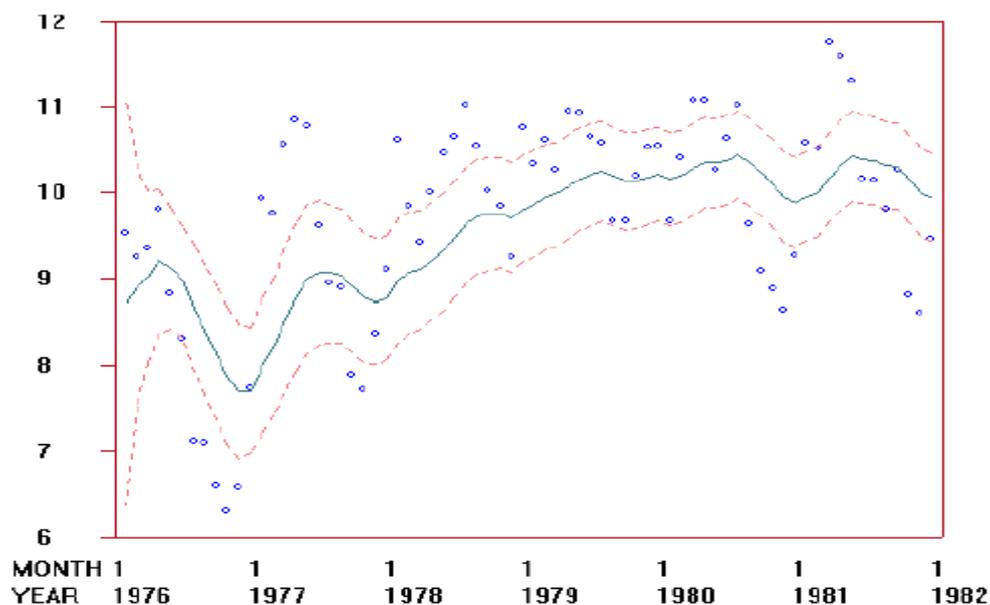


Figura 2.1<sup>5</sup> - Estimativa do nível utilizando o modelo passeio aleatório mais ruído.

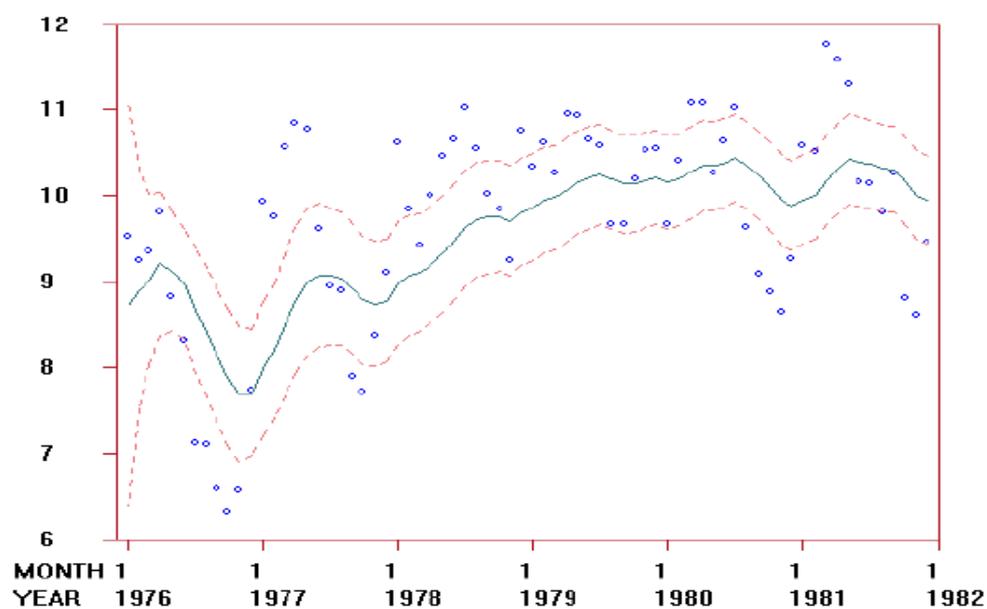


Figura 2.2<sup>6</sup> - Estimativa das vendas utilizando o modelo passeio aleatório mais ruído.

### 2.3.2. Modelo de crescimento linear

Um modelo um pouco mais elaborado é o modelo de crescimento linear, ou modelo de tendência linear local, que tem a mesma equação de observação do modelo de nível local, mas

<sup>5</sup> MIGON & GAMERMAN, 2004

<sup>6</sup> MIGON & GAMERMAN, 2004

considera, na estrutura dinâmica de  $\mu_t$ , um incremento (inclinação do nível) que varia ao longo do tempo:

$$\begin{aligned} Y_t &= \mu_t + v_t, & v_t &\sim N(0, V), \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + w_{t,1}, & w_{t,1} &\sim N(0, \sigma_\mu^2), \\ \beta_t &= \beta_{t-1} + w_{t,2}, & w_{t,2} &\sim N(0, \sigma_\beta^2). \end{aligned}$$

Repare que este é um MLD com:

- $n = 1$ ,
- $p = 2$ ,
- $F_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$  para  $t \geq 1$ ,
- $\theta_t = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix}$  para  $t \geq 1$ ,
- $V_t = V$  para  $t \geq 1$  e
- $G_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$  para  $t \geq 1$ ,
- $w_t = \begin{bmatrix} w_{t,1} \\ w_{t,2} \end{bmatrix}$
- $W_t = \begin{bmatrix} \sigma_\mu^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix}$  para  $t \geq 1$ .

Vale observar que para  $\sigma_\beta^2$  próximo de zero, temos aproximadamente um modelo com incremento constante ao longo do tempo.

Note que neste exemplo, as matrizes  $F_t$ ,  $G_t$  e as matrizes de covariâncias  $V_t$  e  $W_t$  são constantes ao longo do tempo. Nestes casos, o modelo é dito invariante no tempo.

Observe nas Figuras 2.3 e 2.2, como este modelo se adéqua melhor às observações.

## 2.4. Modelo de Regressão Linear

Como foi introduzido anteriormente, os modelos lineares dinâmicos pressupõem normalidade, o que nos possibilita fazer uma comparação inicial com o modelos de regressão linear simples ou múltipla.

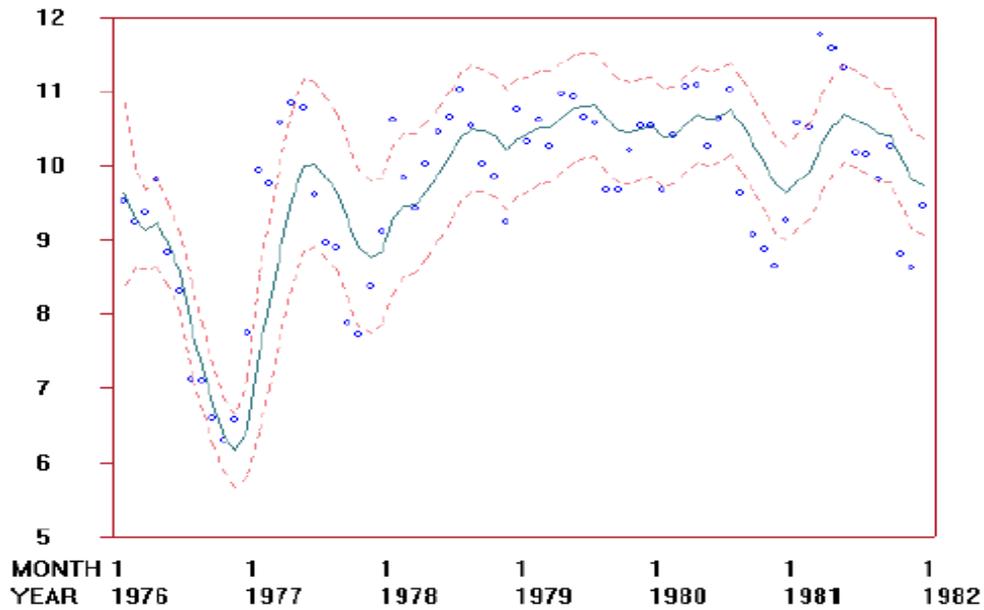


Figura 2.3<sup>7</sup> - Estimativa do nível utilizando o modelo de crescimento linear.

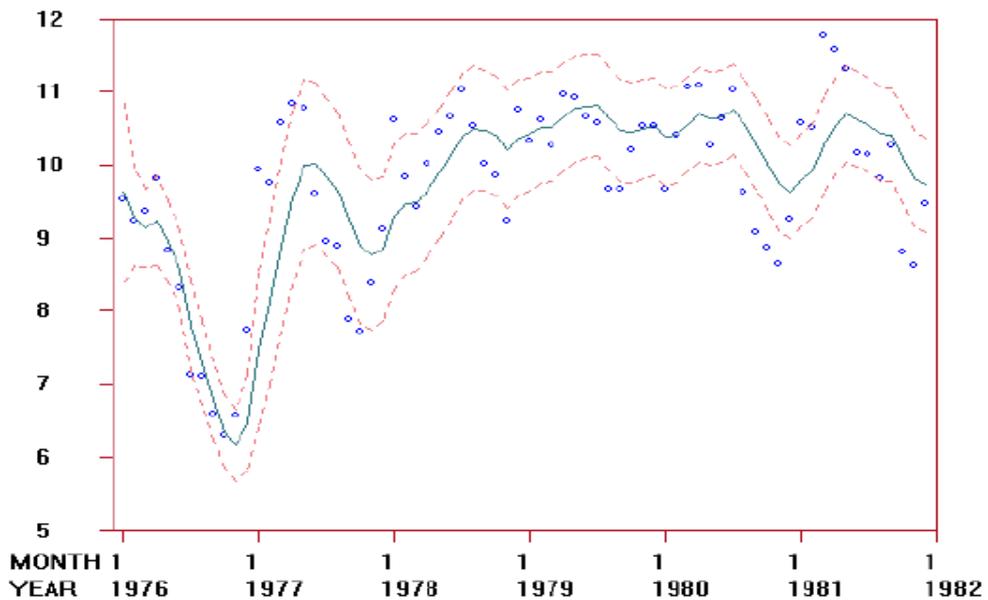


Figura 2.4<sup>8</sup> - Estimativa das vendas utilizando o modelo de crescimento linear.

Em um Modelo de Regressão Linear Simples (MRLS) temos um processo estocástico  $(Y_t)_{t \geq 1}$ , onde  $Y_t$ , geralmente denominada variável resposta, é descrito por

$$Y_t = \theta_0 + \theta_1 x_t + v_t,$$

<sup>7</sup> MIGON & GAMERMAN, 2004

<sup>8</sup> MIGON & GAMERMAN, 2004

onde  $v_t$  é um erro aleatório não observado com distribuição  $N(0, \sigma^2)$ . Em outras palavras,  $Y_t$  é descrito pela soma de uma componente determinística  $(\theta_0 + \theta_1 x_t)$  e uma componente aleatória  $v_t$ , sendo a determinística função linear de uma quantidade regressora  $x_t$ . Já a componente aleatória representa os inúmeros fatores que, conjuntamente podem interferir em  $Y_t$ .

A diferença para o Modelo de Regressão Linear Múltipla (MRLM) é que podemos expressar o valor esperado de  $Y_t$  como função de várias variáveis regressoras (CHARNET et al., 2008). O modelo de regressão múltipla pode ser dado da forma:

$$Y_t = \theta_0 + \theta_1 x_{t,1} + \dots + \theta_q x_{t,q} + v_t,$$

onde  $v_t$  é o erro e tem distribuição  $N(0, \sigma^2)$ , o modelo também pode ser descrito de forma compacta e matricial:

$$Y_t = F_t \theta + v_t,$$

onde  $F_t = \begin{bmatrix} 1 & x_{t,1} & \dots & x_{t,q} \end{bmatrix}$  e  $\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_q \end{bmatrix}$ .

O coeficiente  $\theta_i$  informa, para  $i \in \{1, \dots, q\}$ , a influência da quantidade explicativa  $x_{t,i}$  sobre a resposta  $Y_t$ . Na prática,  $\theta$  é desconhecido e pode ser estimado através de um conjunto de observações pareadas  $\{(y_t; x_t)\}_{t \geq 1}$ , sendo  $y_t$  uma realização de  $Y_t$  (MIGON & GAMERMAN, 2004).

## 2.5. Um MLD que generaliza um Modelo de Regressão Linear

Uma série temporal é um conjunto de observações obtidas sequencialmente ao longo do tempo. A característica mais importante desse tipo de dados é que as observações vizinhas são dependentes. Enquanto nos modelos de regressão a ordem das observações é irrelevante, em séries temporais a ordem dos dados é crucial (EHLERS, 2009). Para analisar uma série temporal  $(Y_t)_{t \geq 1}$  podemos pensar em um MRLM

$$Y_t = F_t \theta + v_t, \quad (2.6)$$

onde  $v_t \sim N(0, \sigma^2)$ ,  $F_t = [1 \quad x_{t,1} \quad \cdots \quad x_{t,q}]$  e  $\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_q \end{bmatrix}$ .

No entanto, no modelo (2.6) a estrutura de evolução temporal é estática. Já em um MLD  $\theta$  admite uma evolução ao longo do tempo e, assim, a forma do modelo tem a possibilidade de mudar. A partir de (2.6), o modelo dinâmico pode ser apresentado, como uma extensão do modelo de regressão:

$$Y_t = F_t \theta_t + v_t, \text{ para } t > 1, \quad (2.7)$$

A diferença entre as equações (2.6) e (2.7) é a indexação do parâmetro  $\theta$ , que passa a evoluir temporalmente, como as variáveis e o ruído,  $\theta$  deixa de ser fixo e inalterado como na equação (2.6), onde  $\theta_{t-1} = \theta_t = \theta_{t+1} = \theta$ . Além disso, como já vimos, em modelos lineares dinâmicos assume-se uma forma geral para relacionar  $\theta_t$  ao seu antecessor  $\theta_{t-1}$ , conhecida como equação do sistema:

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

Repare que se a matriz de evolução  $G_t$  for a matriz identidade e  $W_t$  for zero, o modelo se reduz ao caso estático da regressão linear.

## 2.6. Filtro de Kalman

Como vimos, um MLD possui parâmetros sucessivos relacionados entre si. Para estimar estes parâmetros podemos usar um algoritmo de filtragem chamado filtro de Kalman, no qual as estimativas para o instante  $t+1$  são obtidas através da estimativa no instante anterior  $t$ , mais uma perturbação devida às mudanças do sistema (MIGON & GAMERMAN, 2004).

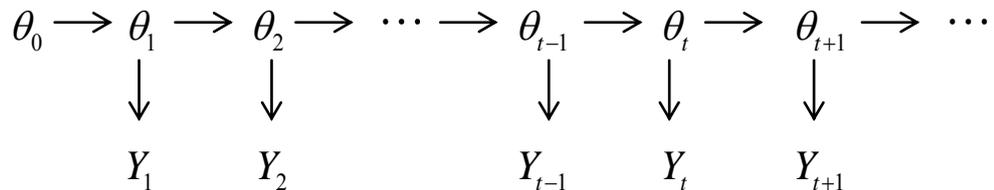
Há centenas de anos, a filtragem é utilizada cotidianamente, por retirar impurezas, geralmente da água, pela passagem dela por uma camada capaz de reter sua sujidade. Nos sistemas de comunicação sempre há ruídos que confundem a mensagem e em termodinâmica a filtragem é utilizada a fim de minimizar as imprecisões causadas por estes barulhos. Em diversos sistemas de controle, o controle é fundamentado em medições anteriores do processo. Estas medições são carregadas de vieses aleatórios devidos ao processo, que são e devem ser eliminados através da filtragem (ANDERSON E MOORE, 1979).

Primeiramente, descreveremos os passos recursivos do Filtro de Kalman para os MLDs Generalizados, mesmo não aprofundando neste tópico, mas para servir de base para depois aplicá-los para os MLDs Gaussianos.

### 2.6.1. Filtro de Kalman para MDG

Nos Modelos Dinâmicos Generalizados, pode não haver a suposição de normalidade da distribuição. Como já discutimos, esta generalização é requerida para modelar séries temporais discretas. Se a variável resposta  $Y_t$  é, por exemplo, a ausência ou presença de certa característica no decorrer do tempo, usaríamos a distribuição de Bernoulli, ou então, se  $Y_t$  é uma contagem em determinado período de tempo, modelamos pela Poisson, etc.

Através da estrutura Markoviana de um modelo dinâmico, é possível chegar à *priori* de hoje a partir da *posteriori* de ontem, isto é, com as últimas informações obtidas de um passado próximo pode-se conhecer o presente. E o método Bayesiano nos possibilita, com a informação da *priori* de hoje, obtida pela equação do sistema, chegar à *posteriori* de hoje. Para amanhã e dias futuros, o ciclo se repete. Resumindo, nos modelos de espaço de estados, informações futuras são relacionadas à informações passadas, formando uma estrutura dependente, como representada abaixo.



Quando  $\{F, G, V, W\}_t$  são conhecidos, a inferência sobre os parâmetros de estado  $\theta_t$  pode ser realizada, nesta subclasse de modelos, através de algoritmos recursivos orientados pela estrutura acima, como o Filtro de Kalman.

**Resultado (filtro recursivo):** A partir das suposições (2.1) e (2.2) e considerando

$$D_t = \{\text{informação total obtida até o instante } t\} = \{D_0, y_1, \dots, y_t\},$$

o modelo dinâmico generalizado estabelece que:

- i. A distribuição preditiva um passo à frente para os parâmetros de estado podem ser computadas a partir da distribuição filtrada  $f(\theta_{t-1} | D_{t-1})$ , conforme

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

- ii. A distribuição preditiva um passo à frente para as observações podem ser computadas pela distribuição preditiva dos estados, segundo

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

- iii. A distribuição filtrada pode ser computada a partir das distribuições dadas acima, por

$$f(\theta_t | D_t) = \frac{f(y_t | \theta_t) f(\theta_t | D_{t-1})}{f(y_t | D_{t-1})}.$$

A propósito estas recursões também mostram que  $f(\theta_t | D_t)$  resume a informação contida em observações passadas  $D_t$ , que são suficientes para prever  $Y_{t+k}$ , para qualquer  $k > 0$ , isto é, a partir da previsão um passo à frente, temos condições de prever  $k$  passos à frente.

Prova: Esta demonstração está relacionada com as propriedades de independência condicionada. Para provar (i), note que  $\theta_t$  é condicionalmente independente de  $D_{t-1}$ , dado  $\theta_{t-1}$ . Então,

$$\begin{aligned} f(\theta_t | D_{t-1}) &= \int f(\theta_{t-1}, \theta_t | D_{t-1}) d\theta_{t-1} \\ &= \int f(\theta_t | \theta_{t-1}, D_{t-1}) f(\theta_{t-1} | D_{t-1}) d\theta_{t-1} \\ &= \int f(\theta_t | \theta_{t-1}) f(\theta_{t-1} | D_{t-1}) d\theta_{t-1}. \end{aligned}$$

Para provar (ii), note que  $Y_t$  é condicionalmente independente de  $D_{t-1}$ , dado  $\theta_t$ . Logo,

$$\begin{aligned} f(y_t | D_{t-1}) &= \int f(y_t, \theta_t | D_{t-1}) d\theta_{t-1} \\ &= \int f(y_t | \theta_t, D_{t-1}) f(\theta_t | D_{t-1}) d\theta_{t-1} \\ &= \int f(y_t | \theta_t) f(\theta_t | D_{t-1}) d\theta_{t-1}. \end{aligned}$$

A parte (iii) é provada pelo Teorema de Bayes e por  $Y_t$  ser condicionalmente independente de  $D_{t-1}$  dado  $\theta_t$ . Então:

$$\begin{aligned} f(\theta_t | D_t) &= \frac{f(\theta_t | D_{t-1}) f(D_t | \theta_t, D_{t-1})}{f(D_t | D_{t-1})} \\ &= \frac{f(\theta_t | D_{t-1}) f(y_t | \theta_t)}{f(y_t | D_{t-1})}. \end{aligned}$$

□

## 2.6.2. Filtro de Kalman para MLD

Os resultados anteriores resolvem inicialmente os problemas de filtragem e previsão, porém, em geral a computação das distribuições condicionais relevantes não é algo trivial. MLDs são um caso especial onde as recursões simplificam consideravelmente. A prova é feita usando resultados da distribuição Normal multivariada, que o vetor aleatório  $(\theta_0, \theta_1, \dots, \theta_t, Y_1, \dots, Y_t)$  é normalmente distribuído para qualquer  $t \geq 1$ . Sendo assim, as distribuições marginais e condicionais também são Gaussianas, elas são completamente determinadas pelas suas médias e variâncias. As soluções dos problemas de filtragem são resolvidas pelo renomado Filtro de Kalman (PETRIS et al., 2009).

Então, as previsões são feitas a partir da combinação de informações obtidas *a priori*,  $(\theta_t | D_{t-1})$ , com a equação das observações,  $y_t = F_t \theta_t + v_t$ , chegando então à chamada distribuição preditiva,  $(y_t | D_{t-1})$ , na qual as previsões serão baseadas.

O Filtro de Kalman refere-se a um conjunto de procedimentos recursivos para estimação em modelos dinâmicos. Considerando o MLD definido em (2.4) e (2.5) com  $\{F, G, V, W\}_t$  conhecidos para  $t=1,2,\dots$ . O Filtro de Kalman é dividido em três etapas dependentes: evolução, previsão e atualização.

Considerando  $D_t = \{\text{informação total obtida até o instante } t\}$ , o filtro de Kalman pode ser assim ilustrado:

$$\begin{array}{ccccc} \dots & (\theta_{t-1} | D_{t-1}) & \xrightarrow{\text{evolução}} & (\theta_t | D_{t-1}) & \xrightarrow{\text{previsão}} & (\theta_t | D_t) & \dots \\ & & & \downarrow \text{atualização} & & & \\ & & & (Y_t | D_{t-1}) & & & \end{array}$$

A evolução é a passagem de  $(\theta_{t-1} | D_{t-1})$  para  $(\theta_t | D_{t-1})$ , previsão é a obtenção de  $(Y_t | D_{t-1})$  e atualização é a passagem de  $(\theta_t | D_{t-1})$  para  $(\theta_t | D_t)$ .

**Resultado (filtro de Kalman):** Considerando o MLD especificado por (2.4) e (2.5), a partir da distribuição

$$\theta_{t-1} | D_{t-1} \sim N(m_{t-1}, C_{t-1}),$$

temos que:

i. A distribuição preditiva um passo à frente de  $\theta_t$  dado  $D_{t-1}$  é Normal, com parâmetros:

$$\begin{aligned} a_t &= E(\theta_t | D_{t-1}) = G_t m_{t-1}, \\ R_t &= \text{Var}(\theta_t | D_{t-1}) = G_t C_{t-1} G_t' + W_t. \end{aligned}$$

ii. A distribuição preditiva um passo à frente de  $Y_t$  dado  $D_{t-1}$  é Normal, com parâmetros:

$$\begin{aligned} f_t &= E(Y_t | D_{t-1}) = F_t a_t, \\ Q_t &= \text{Var}(Y_t | D_{t-1}) = F_t R_t F_t' + V_t. \end{aligned}$$

iii. A distribuição de filtragem de  $\theta_t$  dado  $D_t$  é Normal, com parâmetros:

$$\begin{aligned} m_t &= E(\theta_t | D_t) = a_t + R_t F_t' Q_t^{-1} e_t, \\ C_t &= \text{Var}(\theta_t | D_t) = R_t - R_t F_t' Q_t^{-1} F_t R_t, \end{aligned}$$

onde  $e_t = Y_t - f_t$  é o erro de previsão.

Prova: O vetor aleatório  $(\theta_0, \theta_1, \dots, \theta_t, Y_1, \dots, Y_t)$  tem distribuição conjunta, para qualquer  $t > 0$ , dada por

$$f(D_t, \theta_0, \dots, \theta_t) = \left( \prod_{j=1}^t f(y_j | \theta_j) f(\theta_j | \theta_{j-1}) \right) f(\theta_0)$$

onde as distribuições marginais e condicionais são normais, logo a distribuição conjunta dada acima também é normalmente distribuída, para  $t \geq 1$ . Consequentemente a distribuição de qualquer subvetor também é Gaussiana. Então a distribuição preditiva e a distribuição de filtragem são todas normais, logo podemos calcular suas médias e variâncias.

Para provar (i), considerando  $\theta_t | D_{t-1} \sim N(a_t, R_t)$ , usando (2.5),  $a_t$  e  $R_t$  podem ser assim obtidas:

$$\begin{aligned} a_t &= E(\theta_t | D_{t-1}) \\ &= E(E(\theta_t | \theta_{t-1}, D_{t-1}) | D_{t-1}) \\ &= E(G_t \theta_{t-1} | D_{t-1}) \\ &= G_t m_{t-1} \end{aligned}$$

e

$$\begin{aligned} R_t &= \text{Var}(\theta_t | D_{t-1}) \\ &= E(\text{Var}(\theta_t | \theta_{t-1}, D_{t-1}) | D_{t-1}) + \text{Var}(E(\theta_t | \theta_{t-1}, D_{t-1}) | D_{t-1}) \\ &= W_t + G_t C_{t-1} G_t'. \end{aligned}$$

Para provar (ii), considerando  $Y_t | D_{t-1} \sim N(f_t, Q_t)$ , usando (2.4),  $f_t$  e  $Q_t$  podem ser assim obtidas:

$$\begin{aligned}
f_t &= E(Y_t | D_{t-1}) \\
&= E(E(Y_t | \theta_t, D_{t-1}) | D_{t-1}) \\
&= E(F_t \theta_t | D_{t-1}) \\
&= F_t a_t
\end{aligned}$$

e

$$\begin{aligned}
Q_t &= \text{Var}(Y_t | D_{t-1}) \\
&= E(\text{Var}(Y_t | \theta_t, D_{t-1}) | D_{t-1}) + \text{Var}(E(Y_t | \theta_t, D_{t-1}) | D_{t-1}) \\
&= V_t + F_t R_t F_t'.
\end{aligned}$$

Provaremos, por fim (iii). A fim de calcular a distribuição de filtragem no tempo  $t$ , temos que aplicar a fórmula de Bayes para combinar a *priori*  $f(\theta_t | D_{t-1})$  e a probabilidade  $f(y_t | \theta_t)$ . No caso MLD, todas as distribuições são normais e o problema é o mesmo da inferência Bayesiana para o modelo linear

$$Y_t = F_t \theta_t + v_t, \quad v_t \sim N(0, V_t),$$

com o parâmetro regressivo  $\theta_t$ , seguindo a *priori* Gaussiana  $N(a_t, R_t)$ . Considerando  $V_t$  conhecido, temos que

$$\theta_t | D_t \sim N(m_t, C_t),$$

onde, pelo teorema de Bayes,  $m_t = m_0 + C_0 X' (X C_0 X' + V)^{-1} (y - X m_0)$  e, conseqüentemente,

$$m_t = a_t + R_t F_t' Q_t^{-1} (Y_t - F_t a_t).$$

Ainda pelo teorema de Bayes,  $C_t = C_0 - C_0 X' (X C_0 X' + V)^{-1} X C_0$  e, portanto

$$C_t = R_t - R_t F_t' Q_t^{-1} F_t R_t.$$

□

O filtro de Kalman nos permite calcular as distribuições preditivas e filtradas recursivamente, inicializando por  $\theta_0 \sim N(m_0, C_0)$ , e então calculando  $f(\theta_1, y_1)$ , e procedendo recursivamente assim que novas observações se tornam disponíveis.

Nas Figuras 2.5 e 2.6, note o desempenho dos modelos, apresentados nas seções 2.3.1 e 2.3.2, respectivamente, na previsão um passo à frente. Onde foi utilizado o filtro de Kalman e a mesma base de dados usada para gerar as Figuras 2.1, 2.2, 2.3 e 2.4.

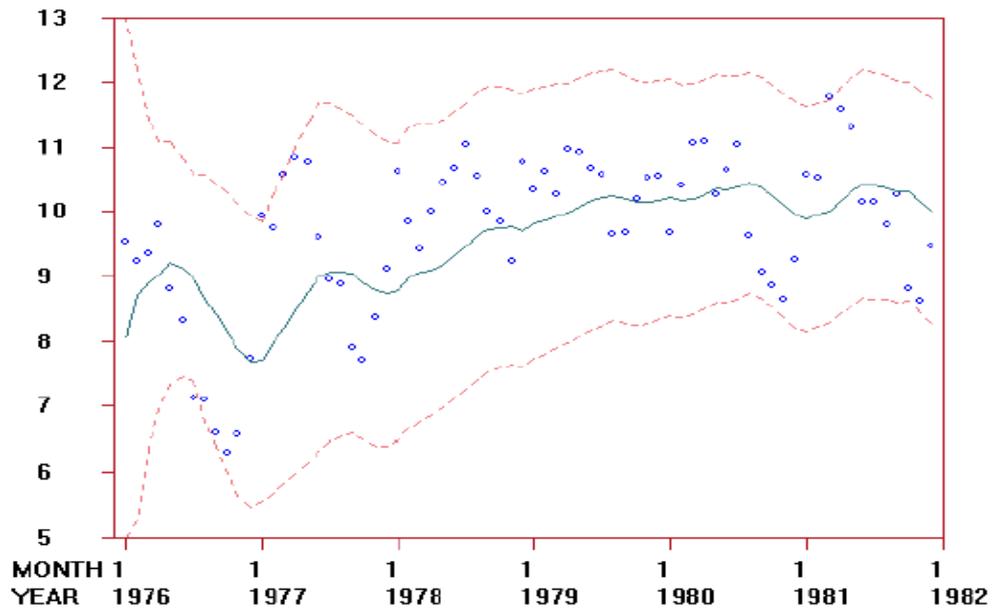


Figura 2.5<sup>9</sup> - Previsão um passo à frente utilizando o modelo passeio aleatório mais ruído.

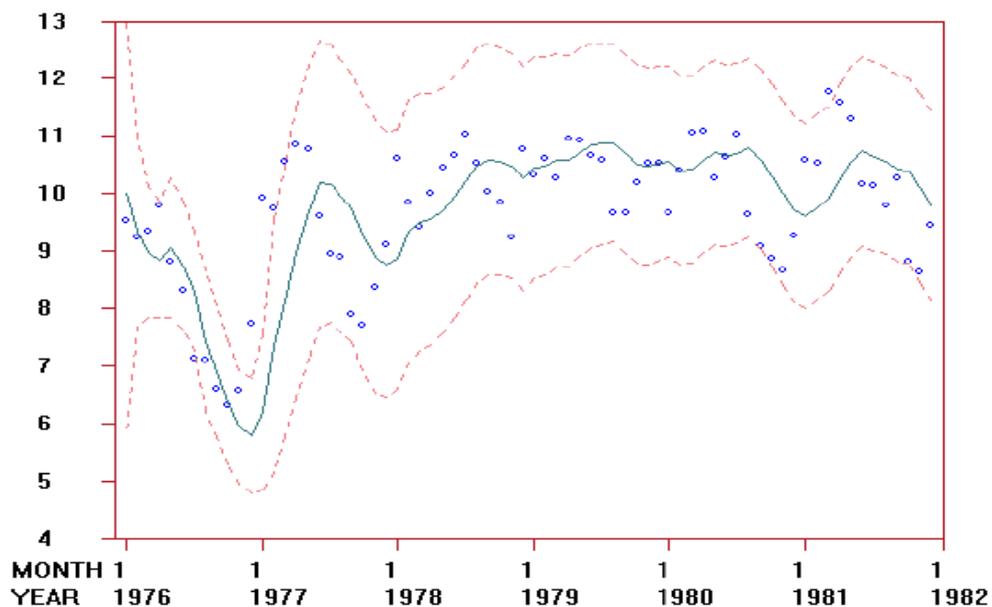


Figura 2.6<sup>10</sup> - Previsão um passo à frente utilizando o modelo de crescimento linear.

## 2.7. Suavização

A suavização ou análise retrospectiva é uma operação de passagem de informação para trás que usa toda a série observada para reavaliar a inferência feita durante o procedimento sequencial. Essa reavaliação é devida à utilização de observações colhidas após

<sup>9</sup> MIGON & GAMERMAN, 2004

<sup>10</sup> MIGON & GAMERMAN, 2004

o período de interesse, com mais informação, sabemos mais e dispomos de mais instrumentos para entender o que se passou (MIGON & GAMERMAN, 2004).

A possibilidade de fazer estimação e previsão sequencialmente, assim que novos dados se tornam disponíveis, é um dos prós do MLD. Na análise de séries temporais, em  $Y_t$  há um número limitado de observações, onde  $t=1, \dots, T$  e o desejo é de reconstruir retrospectivamente o comportamento e a estrutura desta série. Neste caso, usa-se um algoritmo recursivo inverso para calcular as distribuições condicionadas de  $\theta_t$  dado  $D_T$ , para qualquer  $t < T$ , tendo como inicialização a distribuição de filtragem  $f(\theta_t | D_t)$  e estimando retrospectivamente o histórico dos parâmetros de estado (PETRIS et al., 2009).

### 2.7.1. Suavização recursiva para MDG

Em modelos dinâmicos generalizados, definidos por (2.1) e (2.2), as seguintes afirmações sustentam:

- i. Condicionado a  $D_T$ , a sequência de parâmetros de estado  $(\theta_0, \dots, \theta_T)$ , tem probabilidades de transição para trás, dadas por

$$f(\theta_t | \theta_{t+1}, D_T) = \frac{f(\theta_{t+1} | \theta_t) f(\theta_t | D_t)}{f(\theta_{t+1} | D_t)}.$$

- ii. As distribuições suavizadas de  $\theta_t$  dado  $D_T$ , podem ser calculadas de acordo com a seguinte recursão inversa em  $t$ , iniciando em  $f(\theta_T | D_T)$ ,

$$f(\theta_t | D_T) = f(\theta_t | D_t) \int \frac{f(\theta_{t+1} | \theta_t)}{f(\theta_{t+1} | D_t)} f(\theta_{t+1} | D_T) d\theta_{t+1}.$$

Prova: Para provar (i), note que  $\theta_t$  e  $Y_{t+1:T}$  são condicionalmente independentes dado  $\theta_{t+1}$ ; além de  $\theta_{t+1}$  e  $D_t$  serem condicionalmente independentes dado  $\theta_t$ . Usando o Teorema de Bayes, temos que

$$\begin{aligned} f(\theta_t | \theta_{t+1}, D_T) &= f(\theta_t | \theta_{t+1}, D_t) \\ &= \frac{f(\theta_t | D_t) f(\theta_{t+1} | \theta_t, D_t)}{f(\theta_{t+1} | D_t)} \\ &= \frac{f(\theta_t | D_t) f(\theta_{t+1} | \theta_t)}{f(\theta_{t+1} | D_t)}. \end{aligned}$$

Para provar (ii), basta encontrar a marginal de  $f(\theta_t, \theta_{t+1} | D_T)$  em relação à  $\theta_{t+1}$ :

$$\begin{aligned}
f(\theta_t | D_T) &= \int f(\theta_t, \theta_{t+1} | D_T) d\theta_{t+1} \\
&= \int f(\theta_{t+1} | D_T) f(\theta_t | \theta_{t+1}, D_T) d\theta_{t+1} \\
&= \int f(\theta_{t+1} | D_T) \frac{f(\theta_{t+1} | \theta_t) f(\theta_t | D_t)}{f(\theta_{t+1} | D_t)} d\theta_{t+1} \\
&= f(\theta_t | D_t) \int f(\theta_{t+1} | \theta_t) \frac{f(\theta_{t+1} | D_T)}{f(\theta_{t+1} | D_t)} d\theta_{t+1}.
\end{aligned}$$

□

### 2.7.2. Suavizador de Kalman

Para um modelo linear dinâmico gaussiano, a suavização recursiva pode ser estabelecida mais explicitamente pelas médias e variâncias das distribuições de suavização.

Para o MLD, se  $\theta_{t+1} | D_T \sim N(s_{t+1}, S_{t+1})$ , então  $\theta_t | D_T \sim N(s_t, S_t)$ , onde

$$\begin{aligned}
s_t &= m_t + C_t G_{t+1}' R_{t+1}^{-1} (s_{t+1} - a_{t+1}) \\
S_t &= C_t - C_t G_{t+1}' R_{t+1}^{-1} (R_{t+1} - S_{t+1}) R_{t+1}^{-1} G_{t+1} C_t.
\end{aligned}$$

O suavizador de Kalman nos permite calcular as distribuições de  $\theta_t | D_T$ , começando por  $t = T - 1$ , que neste caso  $\theta_T | D_T \sim N(s_T = m_T, S_T = C_T)$ , e então prosseguindo inversamente para calcular as distribuições de  $\theta_t | D_T$  para  $t = T - 2, t = T - 3, \dots$ , isto é, utiliza-se todos os dados disponíveis para estimar estas distribuições. Note que a suavização recursiva depende somente dos dados obtidos pela filtragem e previsão um passo à frente, através do Filtro de Kalman.

Prova: É baseada nas propriedades da distribuição Normal multivariada que a distribuição condicional de  $\theta_t$  dado  $D_T$  é Normal, isto basta para calcular sua média e variância. Então,

$$\begin{aligned}
s_t &= E(\theta_t | D_T) \\
&= E(E(\theta_t | \theta_{t+1}, D_T) | D_T)
\end{aligned}$$

e

$$\begin{aligned}
S_t &= \text{Var}(\theta_t | D_T) \\
&= \text{Var}(E(\theta_t | \theta_{t+1}, D_T) | D_T) + E(\text{Var}(\theta_t | \theta_{t+1}, D_T) | D_T).
\end{aligned}$$

Como mostrado na prova da seção 2.7.1,  $\theta_t$  e  $Y_{t+1:T}$  são condicionalmente independentes dado  $\theta_{t+1}$ , de modo que  $f(\theta_t | \theta_{t+1}, D_T) = f(\theta_t | \theta_{t+1}, D_t)$ . Pode-se usar o Teorema de Bayes

para encontrar esta distribuição. Note que  $f(\theta_{t+1} | \theta_t, D_t) = f(\theta_{t+1} | \theta_t)$  pode ser expressa pela equação do sistema (2.5), logo

$$\theta_{t+1} | \theta_t \sim N(G_{t+1}\theta_t, W_{t+1}).$$

A priori é  $f(\theta_t | D_t)$ , que é normalmente distribuída com parâmetros  $(m_t, C_t)$ . Por  $m_n = m_0 + C_0 X' (XC_0 X' + V)^{-1} (y - X m_0)$  e  $C_n = C_0 - C_0 X' (XC_0 X' + V)^{-1} X C_0$ , temos que

$$\begin{aligned} E(\theta_t | \theta_{t+1}, D_t) &= m_t + C_t G_{t+1}' (G_{t+1} C_t G_{t+1}' + W_{t+1})^{-1} (\theta_{t+1} - G_{t+1} m_t) \\ &= m_t + C_t G_{t+1}' R_{t+1}^{-1} (\theta_{t+1} - a_{t+1}) \end{aligned}$$

e

$$Var(\theta_t | \theta_{t+1}, D_t) = C_t - C_t G_{t+1}' R_{t+1}^{-1} G_{t+1} C_t,$$

do qual resulta, tendo assumido que  $E(\theta_{t+1} | D_T) = s_{t+1}$  e  $Var(\theta_{t+1} | D_T) = S_{t+1}$ ,

$$s_t = E(E(\theta_t | \theta_{t+1}, D_t) | D_T) = m_t + C_t G_{t+1}' R_{t+1}^{-1} (s_{t+1} - a_{t+1})$$

e

$$\begin{aligned} S_t &= Var(E(\theta_t | \theta_{t+1}, D_t) | D_T) + E(Var(\theta_t | \theta_{t+1}, D_t) | D_T) \\ &= C_t - C_t G_{t+1}' R_{t+1}^{-1} G_{t+1} C_t + C_t G_{t+1}' R_{t+1}^{-1} S_{t+1} R_{t+1}^{-1} G_{t+1} C_t \\ &= C_t - C_t G_{t+1}' R_{t+1}^{-1} (R_{t+1} - S_{t+1}) R_{t+1}^{-1} G_{t+1} C_t. \end{aligned}$$

□

## 2.8. Previsão $k$ Passos à Frente

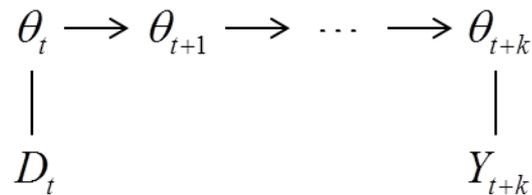
Há muitas aplicações para previsões maiores, para um futuro mais longínquo que apenas para o instante seguinte. Previsões socioeconômicas, por exemplo, como estará a situação financeira do brasileiro daqui a 10 anos? Previsões com a intenção de fazer um plano governamental não podem ser basear apenas para o dia seguinte, mês, ou ano, é necessário que haja para 10, 20, 50 anos, porém, quanto maior for esse espaço de tempo, mais informações são perdidas e incertezas adicionadas ao modelo.

Para séries temporais, onde novas observações são disponibilizadas sequencialmente, a cada dia ou a cada hora, a previsão um passo à frente é natural e pode ser muito bem aplicada. À medida que novos dados são incluídos no modelo, as previsões e atualizações são feitas, também, sequencialmente. Porém, quando se tem  $D_t$  e a intenção é estimar valores futuros,  $Y_{t+k}$ , ou parâmetros de estado,  $\theta_{t+k}$ , a previsão um passo à frente pode ser uma ferramenta

para entender o comportamento do modelo. As distribuições preditivas um passo à frente, dos estados e observações, no MLD, podem ser obtidas através do Filtro de Kalman, apresentado na seção 2.6.

Dado que temos informações até o instante  $t$ , para estimar as médias e variâncias das distribuições condicionadas dos estados e das observações para um instante futuro  $t+k$ , apresentaremos uma fórmula recursiva. Pela natureza Markoviana do modelo, a distribuição de filtragem, no instante  $t$ , age como uma distribuição inicial de evolução futura do modelo. Isto é, a distribuição conjunta dos parâmetros de estado presentes e futuros  $(\theta_{t+k})_{k \geq 0}$  e as observações futuras,  $(Y_{t+k})_{k \geq 1}$ , são aquelas que provêm do modelo dinâmico com distribuições condicionais  $f(\theta_{t+k} | \theta_{t+k-1})$  e  $f(y_{t+k} | \theta_{t+k})$ , e distribuição inicial  $f(\theta_t | D_t)$ , a informação sobre o futuro fornecido pelos dados está toda contida nesta distribuição.

Para os MLD, os dados são usados apenas para obter a média de  $f(\theta_t | D_t)$ ,  $m_t$ , que fornece um resumo dos dados que é suficiente para fins preditivos. O ciclo abaixo representa uma estrutura de dependência entre as variáveis, mostrando o trajeto entre  $Y_{1t}$  e  $Y_{t+k}$ , onde se vê que  $D_t$  fornece informações sobre  $\theta_t$ , que por sua vez provê informações sobre o parâmetro de estado de evolução futura até  $\theta_{t+k}$ , e conseqüentemente sobre  $Y_{t+k}$  e quanto maior for  $k$ , as previsões serão cada vez mais imprecisas.



### 2.8.1. Recursão Preditiva para MDG

Para modelos dinâmicos generalizados, definidos por (2.1) e (2.2), as seguintes afirmações asseguram que, para qualquer  $k > 0$ :

- i. A distribuição preditiva  $k$  passos à frente dos parâmetros de estado é

$$f(\theta_{t+k} | D_t) = \int f(\theta_{t+k} | \theta_{t+k-1}) f(\theta_{t+k-1} | D_t) d\theta_{t+k-1};$$

- ii. A distribuição preditiva  $k$  passos à frente das observações é

$$f(y_{t+k} | D_t) = \int f(y_{t+k} | \theta_{t+k}) f(\theta_{t+k} | D_t) d\theta_{t+k}.$$

Prova: Para prova (i), usamos as propriedades de independência condicional do modelo, então

$$\begin{aligned} f(\theta_{t+k} | D_t) &= \int f(\theta_{t+k}, \theta_{t+k-1} | D_t) d\theta_{t+k-1} \\ &= \int f(\theta_{t+k} | \theta_{t+k-1}, D_t) f(\theta_{t+k-1} | D_t) d\theta_{t+k-1} \\ &= \int f(\theta_{t+k} | \theta_{t+k-1}) f(\theta_{t+k-1} | D_t) d\theta_{t+k-1}, \end{aligned}$$

A prova de (ii) é também baseada nas propriedades de independência condicional do modelo, temos que

$$\begin{aligned} f(y_{t+k} | D_t) &= \int f(y_{t+k}, \theta_{t+k} | D_t) d\theta_{t+k} \\ &= \int f(y_{t+k} | \theta_{t+k}, D_t) f(\theta_{t+k} | D_t) d\theta_{t+k} \\ &= \int f(y_{t+k} | \theta_{t+k}) f(\theta_{t+k} | D_t) d\theta_{t+k}. \end{aligned}$$

□

## 2.8.2. Recursão Preditiva para MLD

Para modelos lineares dinâmicos definidos em (2.4) e (2.5), considere  $a_t(0) = m_t$  e  $R_t(0) = C_t$ , isto é, a média e variância da distribuição preditiva um passo à frente de  $\theta_t$  dado  $D_{t-1}$ , no instante zero, é igual a média e variância da distribuição de filtragem de  $\theta_t$  dado  $D_t$ , respectivamente. Então, para  $k \geq 1$ , eles seguem as seguintes afirmações:

- i. A distribuição de  $\theta_{t+k}$  dado  $D_t$  é Normal, com

$$\begin{aligned} a_t(k) &= G_{t+k} a_{t,k-1}, \\ R_t(k) &= G_{t+k} R_{t,k-1} G_{t+k}' + W_{t+k}; \end{aligned}$$

- ii. A distribuição de  $Y_{t+k}$  dado  $D_t$  é Normal, com

$$\begin{aligned} f_t(k) &= F_{t+k} a_t(k), \\ Q_t(k) &= F_{t+k} R_t(k) F_{t+k}' + V_{t+k}. \end{aligned}$$

Prova: Sabe-se já que todas as distribuições condicionadas são Gaussianas. Então para  $k > 1$

$$\begin{aligned} a_t(k) &= E(\theta_{t+k} | D_t) \\ &= E(E(\theta_{t+k} | D_t, \theta_{t+k-1}) | D_t) \\ &= E(G_{t+k} \theta_{t+k-1} | D_t) \\ &= G_{t+k} a_{t,k-1}, \end{aligned}$$

$$\begin{aligned}
R_t(k) &= \text{Var}(\theta_{t+k} | D_t) \\
&= \text{Var}(E(\theta_{t+k} | D_t, \theta_{t+k-1}) | D_t) + E(\text{Var}(\theta_{t+k} | D_t, \theta_{t+k-1}) | D_t) \\
&= G_{t+k} R_{t,k-1} G_{t+k}' + W_{t+k},
\end{aligned}$$

$$\begin{aligned}
f_t(k) &= E(Y_{t+k} | D_t) \\
&= E(E(Y_{t+k} | D_t, \theta_{t+k}) | D_t) \\
&= E(F_{t+k} \theta_{t+k} | D_t) \\
&= F_{t+k} a_t(k),
\end{aligned}$$

$$\begin{aligned}
Q_t(k) &= \text{Var}(Y_{t+k} | D_t) \\
&= \text{Var}(E(Y_{t+k} | D_t, \theta_{t+k}) | D_t) + E(\text{Var}(Y_{t+k} | D_t, \theta_{t+k}) | D_t) \\
&= F_{t+k} R_t(k) F_{t+k}' + V_{t+k}.
\end{aligned}$$

□

## 2.9. Checagem do Modelo

Como foi visto, com a modelagem dinâmica podemos calcular previsões um passo à frente  $f_t = E(Y_t | D_{t-1})$ , e o erro de previsão é assim definido:

$$e_t = Y_t - E(Y_t | D_{t-1}) = Y_t - f_t,$$

isto é, o erro é, por definição, a observação verdadeira menos sua previsão, para um dado instante.

Os erros de previsão podem ser escritos, de maneira alternativa, em termos da estimação um passo à frente, como apresentado abaixo

$$\begin{aligned}
e_t &= Y_t - F_t a_t \\
&= F_t \theta_t + v_t - F_t a_t \\
&= F_t (\theta_t - a_t) + v_t.
\end{aligned}$$

Seja  $(e_t)$  a sequência de erros de previsão de um MLD, para  $t \geq 1$ , que seguem as seguintes propriedades:

- i. O valor esperado de  $e_t$  é zero;
- ii. O vetor aleatório  $(e_t)$  não tem correlação com qualquer função de  $Y_1, \dots, Y_{t-1}$ ;
- iii. Para qualquer  $s < t$ ,  $e_t$  e  $Y_s$  são descorrelacionados;
- iv. Para qualquer  $s < t$ ,  $e_t$  e  $e_s$  são descorrelacionados;
- v.  $e_t$  é uma função linear de  $Y_1, \dots, Y_t$ ;
- vi.  $(e_t)_{t \geq 1}$  é um processo Gaussiano.

Prova: Prova-se (i) tomando valores esperados iterados,

$$E(e_t) = E(E(Y_t - f_t | D_{t-1})) = 0.$$

Para provar (ii), considere  $Z = g(Y_1, \dots, Y_{t-1})$ . Então:

$$\begin{aligned} \text{Cov}(e_t, Z) &= E(e_t Z) \\ &= E(E(e_t Z | D_{t-1})) \\ &= E(E(e_t | D_{t-1}) Z) \\ &= 0. \end{aligned}$$

Prova-se (iii), se as observações são univariadas, segue-se de (ii), considerando  $Z = Y_s$ .

Caso contrário, use (ii) para cada componente de  $Y_s$ .

Para provar (iv), também segue-se de (ii), tomando  $Z = e_s$ , se as observações são univariadas. Caso contrário, aplique (ii) componente a componente.

Para (v), desde que  $Y_1, \dots, Y_t$  tenha distribuição conjunta Normal,  $f_t = E(Y_t | Y_{1:t-1})$  é uma função linear de  $Y_1, \dots, Y_{t-1}$ . Como consequência,  $e_t$  é uma função linear de  $Y_1, \dots, Y_t$ .

Por fim, para provar (vi), para qualquer  $t$ , tendo em vista (v),  $(e_1, \dots, e_t)$  é uma transformação linear de  $(Y_1, \dots, Y_t)$ , que tem uma distribuição conjunta Normal. Então,  $(e_1, \dots, e_t)$  também tem distribuição conjunta Normal. Desde que todas as distribuições de dimensão finita sejam Gaussianas, o processo  $(e_t)_{t \geq 1}$  é também Gaussiano.

□

Em  $Y_t = f_t + e_t$ , podemos pensar em  $Y_t$  como a soma de um componente,  $f_t$ , que é previsível a partir de observações passadas, com outro componente,  $e_t$ , que é independente do passado, então contém informações realmente novas fornecidas pela observação  $Y_t$ . Por este motivo, os erros de previsão podem ser chamados de “inoações”, algumas vezes, pode ser conveniente trabalhar com este termo em um MLD.

A inoação é obtida pela eleição dos vetores  $a_t = E(\theta_t | D_{t-1})$ , como novas variáveis de estado. Então, a partir de  $e_t = Y_t - f_t = Y_t - F_t a_t$ , a equação das observações é dada por

$$Y_t = F_t a_t + e_t \quad (2.8)$$

e, tomando  $a_t = G_t m_{t-1}$ , onde  $m_{t-1}$  é obtido pelo filtro de Kalman, então:

$$\begin{aligned} a_t &= G_t m_{t-1} \\ &= G_t a_{t-1} + G_t R_{t-1} F_{t-1}' Q_{t-1}^{-1} e_t; \end{aligned}$$

então, a nova equação do sistema ou dos estados é assim representada:

$$a_t = G_t a_{t-1} + w_t^*, \quad (2.9)$$

onde  $w_t^* = G_t R_{t-1} F_{t-1}' Q_{t-1}^{-1} e_t$ . O sistema (2.8) e (2.9) é a forma de inovação do MLD. Note que nesta forma, os erros observacionais e do sistema não são mais independentes, isto é, a dinâmica dos estados se torna dependente das observações.

A principal vantagem da forma de inovação é que todos os componentes do vetor de estados, para os quais não podemos obter nenhuma informação a partir das observações, são automaticamente removidos, se tornando então, um modelo mínimo.

Quando as observações são univariadas, espera-se que a sequência de inovações ou erros padronizados sejam variáveis aleatórias normais, com média zero, independentes e identicamente distribuídas. Esta sequência pode ser chamada de ruído branco Gaussiano e é definida por  $\tilde{e}_t = \frac{e_t}{\sqrt{Q_t}}$ . Para checar se o modelo está correto, esta propriedade é útil para

explorar as suposições deste, onde a sequência  $e_1, \dots, e_t$  deve ser um ruído branco. Existem diversos testes não paramétricos de normalidade, como por exemplo, Lilliefors, Shapiro-Wilk e Kolmogorov-Smirnov e, também, de independência, como o Qui-Quadrado, para mais informações consulte Conover (1971). Há outra maneira para checar a normalidade, através de análises gráficas, usando o histograma comum, ou o diagrama conhecido como QQ-Plot. O gráfico das inovações padronizadas distribuídas ao longo tempo, pode mostrar *outliers*, pontos de mudanças e outros padrões inesperados.

Para observações multivariadas, as ferramentas gráficas geralmente são as mesmas, aplicadas componente a componente.

## Capítulo 3

### Modelos com componentes desconhecidas

Quando as matrizes  $F_t$ ,  $G_t$ ,  $V_t$  e  $W_t$  são desconhecidas, não é possível utilizar somente o Filtro de Kalman e a inferência é feita através de métodos de simulação estocástica como, por exemplo, os métodos Monte Carlo via cadeias de Markov (MCMC), que são uma alternativa em problemas complexos em que a distribuição a *posteriori* não pode ser explicitada (COSTA, 2011).

O que foi visto até agora foram análises de séries temporais considerando os MLD com as matrizes  $F_t, G_t, V_t$  e  $W_t$  conhecidas, para melhor compreender seu comportamento e propriedades gerais. Porém na prática, isto raramente ocorre. Neste capítulo, vamos assumir que estas matrizes dependem de um vetor de parâmetros desconhecidos  $\psi$ . Estes parâmetros desconhecidos, geralmente são constantes ao longo do tempo, mas para respeitar a estrutura linear Gaussiana dos MLDs, indexaremos  $\psi_t$ .

Sob o ponto de vista clássico, geralmente estima-se  $\psi$  (por máxima verossimilhança) e, assumindo que  $\psi$  é conhecido, aplica-se as técnicas apresentadas no capítulo anterior para filtragem, suavização ou previsão. No entanto, tal procedimento desconsidera a incerteza inerente a  $\psi$  no procedimento de inferência para os parâmetros de estado. Sob o paradigma bayesiano,  $\psi$  é tratado, juntamente com os parâmetros de estado, como quantidade desconhecida e a incerteza é contemplada integralmente através da distribuição conjunta a posteriori do vetor de parâmetros de estado e  $\psi$ .

A inferência Bayesiana pode envolver cálculos de difícil manejo analítico ou até mesmo integrais sem solução analítica. No entanto, os métodos de Monte Carlo via Cadeia de Markov são bem eficientes na aproximação da distribuição *posteriori* de interesse. Mas antes de detalharmos mais estas técnicas, vamos tratar, especificamente, o caso no qual as variâncias dos erros de evolução são desconhecidas, usando uma técnica conhecida como fator de desconto.

### 3.1. Fator de Desconto

Até em modelos estruturais simples, onde  $F_t$  e  $G_t$  são conhecidos, raramente as matrizes de covariância  $V_t$  e  $W_t$  são totalmente conhecidas, desse modo, o problema é estimá-las. Quando  $V_t$  é conhecido,  $W_t$  pode ser especificado por uma técnica chamada Fator de Desconto.

A estrutura e a grandeza da matriz de covariância de parâmetros de estado  $W_t$  tem um papel crucial na determinação da função das observações passadas na estimação e previsão dos estados. Pensando em  $W_t$  como uma matriz diagonal, por simplicidade. Grandes valores na diagonal de  $W_t$ , implicam em grandes incertezas na evolução do estado, então muita informação da amostra é perdida passando de  $\theta_{t-1}$  para  $\theta_t$ , isto é,  $D_{t-1}$  nos dá informações sobre  $\theta_{t-1}$ , o qual, no entanto, se torna de pouca relevância para previsão de  $\theta_t$ , na verdade o que realmente é determinante na estimação de  $\theta_t | D_t$  são as observações atuais  $y_t$ . Nas recursões do filtro de Kalman, as incertezas sobre  $\theta_t$  dado  $D_{t-1}$  são resumidas na matriz de covariância condicional  $Var(\theta_{t-1} | D_{t-1}) = C_{t-1}$ . Atualizando  $\theta_{t-1}$  para  $\theta_t$ , através da equação de estados  $\theta_t = G_t \theta_{t-1} + w_t$ , a incerteza aumenta e temos  $Var(\theta_t | D_{t-1}) = R_t = G_t C_{t-1} G_t' + W_t$ . Então se  $W_t = 0$ , isto é, considerando que não haja erro na equação de estados, digamos que  $R_t = Var(G_t \theta_{t-1} | D_{t-1}) = P_t$ ; caso contrário  $P_t$  é acrescido em  $R_t = P_t + W_t$ . Neste caso  $W_t$  expressa a perda de informação na passagem de  $\theta_{t-1}$  para  $\theta_t$  devido ao erro estocástico na evolução do estado, a perda depende da grandeza de  $W_t$  em relação a  $P_t$ . Por esse motivo, pode-se expressar  $W_t$  como proporção de  $P_t$ :

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

onde  $0 < \delta \leq 1$ . O parâmetro  $\delta$  é chamado de fator de desconto, m a e q e e “desconta” a matriz  $P_t$  que por sua vez, é determinística na evolução de estado da matriz  $R_t$ .

Se  $\delta = 1$ , então  $W_t = 0$  e não existe perda de informação na passagem de  $\theta_{t-1}$  para  $\theta_t$ , sendo  $Var(\theta_t | D_t) = Var(G_t \theta_{t-1} | D_{t-1}) = P_t$ , isto é, o modelo permanece estático. Porém para  $\delta < 1$ , por exemplo  $\delta = 0,8$ , temos  $Var(\theta_t | D_t) = \left( \frac{1}{0,8} \right) Var(G_t \theta_{t-1} | D_{t-1}) = 1,25 P_t$ , o que

revela um aumento na incerteza. E quanto menor for o fator de desconto maior será a perda de informação. Na prática, o valor de  $\delta$  é fixado entre 0,9 e 0,99 ou é escolhido observando o comportamento preditivo do modelo para diferentes  $\delta$ . Valores muito próximos de  $\delta$  provocam mudanças tão sutis no sistema, que são quase imperceptíveis. Além disso, podem ser considerados diferentes fatores de desconto  $\delta_i$  para diferentes componentes do vetor de estados.

### 3.2. Métodos de Monte Carlo via Cadeias de Markov

Em inferência Bayesiana, a distribuição a *posteriori* dos parâmetros, frequentemente é desconhecida, portanto uma alternativa é um método que simule esta distribuição. Os métodos de Monte Carlo baseados na simulação de variáveis aleatórias de uma Cadeia de Markov, são, atualmente, a maneira usual de fazer uma análise Bayesiana dos dados, chamados de Métodos de Monte Carlo via Cadeias de Markov (MCMC).

Dois desses métodos serão apresentados, são os algoritmos recursivos mais utilizados, Amostrador de Gibbs e Metroplis-Hastings. Nas próximas seções, vamos apresentar técnicas para simulação de um vetor aleatório qualquer  $(\theta_1, \dots, \theta_k)$  que, num contexto bayesiano, representa os parâmetros desconhecidos de um modelo. Sendo assim, os métodos de simulação que iremos apresentar são utilizados em inferência bayesiana para simular da distribuição a posteriori de  $(\theta_1, \dots, \theta_k)$ .

#### 3.2.1. Amostrador de Gibbs

Seja um vetor aleatório  $(\theta_1, \dots, \theta_k)$ , as distribuições de  $(\theta_1 | \theta_2, \theta_3, \dots, \theta_k), \dots, (\theta_k | \theta_1, \theta_2, \dots, \theta_{k-1})$  são chamadas de condicionais completas de  $(\theta_1, \dots, \theta_k)$ . E, pela teoria de probabilidades, a densidade condicional  $f(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$  pode ser assim obtida:

$$f(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k) = \frac{f(\theta_1, \dots, \theta_k)}{\int f(\theta_1, \dots, \theta_k) d\theta_i}.$$

Naturalmente, para distribuições discretas, a integral na expressão acima cede lugar ao somatório.

O Amostrador de Gibbs (AG) é um dos métodos MCMC mais utilizados, onde seu núcleo de transição é formado pelas distribuições condicionais completas, isto é, a partir de valores simulados sucessivamente das distribuições condicionais completas, seu objetivo é chegar à distribuição conjunta a posteriori.

Desde que saibamos simular as condicionais completas, o Amostrador de Gibbs é descrito pela rotina abaixo para simular o vetor aleatório  $(\theta_1, \dots, \theta_k)$ :

1º Passo: Inicializar o contador de iterações  $j=1$  e definir um vetor de valores iniciais

$$\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_k^{(0)}).$$

2º Passo: Obter o vetor  $\theta^{(1)} = (\theta_1^{(1)}, \dots, \theta_k^{(1)})$ , simulando:

$$\begin{aligned} \theta_1^{(j)} & \text{ de } p\left(\theta_1 \mid \theta_2^{(j-1)}, \theta_3^{(j-1)}, \dots, \theta_k^{(j-1)}\right) \\ \theta_2^{(j)} & \text{ de } p\left(\theta_2 \mid \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_k^{(j-1)}\right) \\ & \vdots \\ \theta_k^{(j)} & \text{ de } p\left(\theta_k \mid \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_{k-1}^{(j-1)}\right) \end{aligned}$$

3º Passo: Se o contador for menor que  $L$ , dado que  $L$  é o número desejado de iterações após a convergência, mude o contador de  $j$  para  $j+1$  e repita o passo 2.

Então o algoritmo do AG produz sequências  $\theta_i^{(1)}, \theta_i^{(2)}, \dots, \theta_i^{(L)}$ , para  $i=1, \dots, K$ . Na prática, o procedimento utilizado para obter amostras independentes da distribuição *posteriori*, a partir destas sequências, consiste em descartar um número  $b$  de valores iniciais, e em seguida, escolher valores com espaçamento igual a  $t$ . As quantidades  $b$  e  $t$ , são chamadas respectivamente de aquecimento (*burning*) e espaçamento (*thinning*). Mais informações podem ser obtidas em Gamerman e Lopes (2006).

A simulação da distribuição *posteriori* pelo amostrador de Gibbs envolve então a simulação das condicionais completas. Porém, quando esta simulação não é acessível, podemos usar outras técnicas para simular de alguma (ou todas) as distribuições condicionais. Uma destas outras técnicas será apresentada na próxima seção: o Metropolis-Hastings.

### 3.2.2. Metropolis-Hastings

No algoritmo Metropolis-Hastings, é utilizado um valor gerado de uma distribuição auxiliar conhecida como distribuição proposta e aceito com uma probabilidade  $\alpha$ . Esse

mecanismo de correção garante a convergência da cadeia para a distribuição de equilíbrio, a *posteriori*. A rotina do Metropolis-Hastings é assim descrita:

1º Passo: Inicializar o contador de iterações  $j=1$  e definir um vetor de valores iniciais

$$\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_n^{(0)}).$$

2º Passo: Obter um novo um novo vetor  $\theta^* = (\theta_1^*, \dots, \theta_n^*)$ , simulando uma densidade proposta  $q(\theta^* | \theta^{(j-1)})$ , que pode ou não depender de  $\theta^{(j-1)}$ .

3º Passo: Simular uma amostra  $u$ , vinda da distribuição  $U[0,1]$ .

4º Passo: Calcular a razão

$$r(\theta^{(j-1)}, \theta^*) = \frac{f(\theta^*)q(\theta^{(j-1)} | \theta^*)}{f(\theta^{(j-1)})q(\theta^* | \theta^{(j-1)})}.$$

5º Passo: Se  $u < r(\theta^{(j-1)}, \theta^*)$ , o valor proposto é aceito fazendo  $\theta^{(j)} = \theta^*$ , caso contrário, o valor proposto é rejeitado fazendo  $\theta^{(j)} = \theta^{(j-1)}$ .

6º Passo: Se o contador for menor que  $L$ , dado que  $L$  é o número desejado de iterações após a convergência, mude o contador de  $j$  para  $j+1$  e repita o passo 2.

É importante ressaltar que este método está diretamente ligado à distribuição analisada. Caso a variância da distribuição seja muito pequena, a cadeia de Markov irá convergir lentamente, uma vez que seus incrementos serão pequenos. Se a variância for grande, a taxa de rejeição dos valores propostos será alta e a cadeia tenderá a não se mover. Alguns autores sugerem que a taxa de aceitação do AG deva estar entre 20% e 50% (COSTA,2011).

Note que o Amostrador de Gibbs é um caso particular do Metropolis-Hastings, no qual os elementos de  $\theta$  são atualizados individualmente ou em blocos, onde a distribuição condicional total é a proposta e a probabilidade de aceitação é igual a 1. Como a probabilidade de aceitação é igual a 1, não existe mecanismo de aceitação-rejeição, deste modo a cadeia irá sempre se mover para um novo valor. As transições de estado são feitas de acordo com as distribuições condicionais completas.

Os dois algoritmos AG e MH produzem sequências  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(k)}$  e para obter amostras aproximadamente independentes da distribuição, basta analogamente, retirar as observações iniciais (*burning*) e usar um espaçamento (*thinning*).

### 3.2.3. Amostrador de Gibbs para MLD

Para simular da distribuição a posteriori de um MLD contendo um vetor de parâmetros desconhecidos  $\psi$  (conforme descrito anteriormente), com distribuição a priori  $f(\psi)$ , geralmente usa-se os métodos de MCMC. Quase todos amostradores de cadeia de Markov para análises a posteriori de MLDs estão em uma destas três categorias: amostradores de Gibbs, que incluem os estados como variáveis latentes; amostradores marginais; e amostradores híbridos, que combina aspectos dos dois anteriores. Neste trabalho, falaremos apenas do amostrador de Gibbs, para mais informações veja Petris, Petroni e Campagnoli (2009).

Para um MLD, com observações  $D_T$ , a distribuição *posteriori* de  $\psi$  e dos estados não observados, é dada por

$$f(\psi, \theta_{0:T} | D_T). \quad (3.1)$$

Geralmente é impossível calcular esta distribuição de forma fechada, então, a solução é recorrer a métodos computacionais de simulação estocástica: Métodos MCMC. A abordagem MCMC habitual para analisar a distribuição *posteriori* (3.1) é retirar uma amostra dela e avaliar os resultados gerados pela amostra simulada. Mesmo quando a intenção é estudar apenas a distribuição do vetor de parâmetros desconhecidos  $f(\psi | D_T)$ , a inclusão dos parâmetros de estado na distribuição *posteriori* simplifica a rotina de um amostrador eficiente. Na verdade, simular uma variável ou vetor aleatório a partir de  $f(\psi | \theta_{0:T}, D_T)$  é bem mais simples que a partir de  $f(\psi | D_T)$ . Isto é, uma amostra de (3.1) pode ser obtida através do amostrador de Gibbs, alternando os sorteios entre  $f(\psi | \theta_{0:T}, D_T)$  e  $f(\theta_{0:T} | \psi, D_T)$ . Esta amostra, por sua vez, pode ser utilizada como entrada para gerar uma amostra da distribuição preditiva dos estados e observações,  $f(\theta_{T+1:T+k}, y_{T+1:T+k} | D_T)$ . Então,

$$f(\theta_{T+1:T+k}, y_{T+1:T+k}, \psi, \theta_T | D_T) = f(\theta_{T+1:T+k}, y_{T+1:T+k} | \psi, \theta_T) f(\psi, \theta_T | D_T).$$

Portanto, para cada par  $(\psi, \theta_T)$  formado de  $f(\psi, \theta_{0:T} | D_T)$  pode-se gerar o futuro,  $\theta_{T+1:T+k}, y_{T+1:T+k}$ , de  $f(\theta_{T+1:T+k}, y_{T+1:T+k} | \psi, \theta_T)$  para obter uma amostra da distribuição preditiva.

A aproximação mostrada acima resolve todos os problemas da previsão, suavização e filtragem de MLD com vetor de parâmetros desconhecidos  $\psi$ . Porém, para séries em que novos dados são obtidos sequencialmente, para atualizar a distribuição *posteriori* é necessário

rodar o amostrador de Gibbs novamente, o que se torna uma tarefa ineficiente. Para séries deste tipo, existem as análises online e simulações baseadas em atualizações sequenciais da distribuição *posteriori* dos parâmetros de estado e desconhecidos, que são também técnicas de Monte Carlo.

### 3.2.4. FFBS

Em uma amostra de Gibbs retirada de  $f(\theta_{0:T}, \psi | D_T)$ , é necessário simular a partir das distribuições condicionais completas  $f(\psi | \theta_{0:T}, D_T)$  e  $f(\theta_{0:T} | \psi, D_T)$ . Embora a primeira seja um problema específico, algoritmos eficientes para amostragem da segunda estão disponíveis. A suavização recursiva fornece um algoritmo para calcular a média e variância da distribuição de  $\theta_t$  condicionada a  $D_T$  e  $\psi$ , para  $t = 1, 2, \dots, T$ . Já que todas as distribuições envolvidas são Normais, é fácil determinar a distribuição marginal *posteriori* de  $(\theta_t | D_T, \psi)$ . Entretanto, quando  $f(\theta_{0:T} | \psi, D_T)$  tem o papel de condicional completa em uma amostra de Gibbs retirada de  $f(\theta_{0:T}, \psi | D_T)$ , o desafio é gerar um esquema a partir da distribuição de  $\theta_{0:T}$  dado  $(D_T, \psi)$ . A solução para esse problema foi desenvolvida por Carter e Khon (1994), Frühwirth-Schnatter (1994) e Shepard (1994), um algoritmo conhecido como FFBS, em inglês, *Forward Filtering Backward Sampling*, e em português, filtragem para frente e amostragem para trás.

Pode-se representar a distribuição conjunta de  $\theta_{0:T}$  dado  $D_T$ , como:

$$f(\theta_{0:T} | D_T) = \prod_{t=0}^T f(\theta_t | \theta_{t+1:T}, D_T), \quad (3.2)$$

onde o último fator do produto é simplesmente  $f(\theta_T | D_T)$ , isto é, a distribuição da filtragem de  $\theta_T$ , que é  $N(m_T, C_T)$ . A equação (3.2) sugere que, ao invés de obter um esquema de  $f(\theta_{0:T} | D_T)$ , pode-se começar desenhando  $\theta_T$  a partir de sua distribuição  $N(m_T, C_T)$  e então recursivamente para  $\theta_t$  a partir de  $f(\theta_t | \theta_{t+1:T}, D_T)$ , para  $t = T-1, T-2, \dots, 0$ . Note que o FFBS é essencialmente uma versão simulada das recursões de suavização. Vimos na prova da seção 2.7.2, que  $f(\theta_t | \theta_{t+1:T}, D_T) = f(\theta_t | \theta_{t+1}, D_t)$  e demonstramos que essa distribuição é  $N(h_t, H_t)$ , onde

$$\begin{aligned} h_t &= m_t + C_t G_{t+1}' R_{t+1}^{-1} (\theta_{t+1} - a_{t+1}), \\ H_t &= C_t - C_t G_{t+1}' R_{t+1}^{-1} G_{t+1} C_t. \end{aligned}$$

Além disso, obtendo já  $(\theta_{t+1}, \dots, \theta_T)$ , o próximo passo consiste em apenas desenhar  $\theta_t$  a partir de  $N(h_t, H_t)$ . Note que  $h_t$  depende apenas de  $\theta_{t+1}$  que já foi gerado.

O FFBS é frequentemente utilizado como um bloco na construção do amostrador de Gibbs, no entanto, se o MLD estudado não tiver parâmetros desconhecidos, ainda assim, o uso do FFBS pode ser interessante. Neste caso, a distribuição marginal suavizada de cada  $\theta_t$ , geralmente é suficiente para avaliar as probabilidades *posterioris* de interesse. A distribuição *posteriori* de uma função não linear dos estados pode ser complicada ou até impossível de ser derivada, mesmo quando os parâmetros do modelo são conhecidos. Neste caso, o FFBS fornece um jeito simples de gerar uma amostra da *posteriori* desta função, então a aplicação da parte “*foward filtering*” (fi tragem para frente) do algoritmo só precisa ser realizada uma vez.

## Capítulo 4

### Conclusão

O objetivo deste trabalho foi revisar e apresentar a modelagem dinâmica, onde foi possível entender seu funcionamento para análise de séries temporais e, também, o porquê de propor uma nova técnica de análise e previsão Bayesiana, uma vez que as clássicas poderiam parecer mais simples. Vimos que os modelos de espaço de estados podem ser usados para modelar séries temporais uni ou multivariadas, podendo ser não estacionárias ou apresentar mudanças estruturais ou padrões irregulares, como séries econômicas, por exemplo.

Durante o trabalho a teoria matemática foi apresentada, mostrando casos gerais e especiais, quando o modelo deve ser mais complexo, necessitando de mais variáveis dinâmicas e quando ele pode ser mais simples e direto, como o passeio aleatório. Foram expostas, estudadas e demonstradas técnicas de previsão um passo à frente, utilizando o renomado filtro de Kalman que possibilita a evolução, a atualização e, por fim, a previsão dos parâmetros de estado e observações ainda não coletadas e, como vimos, pode ser muito bem aplicado em séries sequenciais, onde novas observações são coletadas entre períodos de tempo fixos. A partir do filtro de Kalman como base, utiliza-se da teoria de previsão um passo à frente para realizar previsões mais distantes,  $k$  instantes após a última observação coletada, porém sabendo que mais incertezas são inseridas no modelo. Foram apresentados, também, os métodos de suavização que aproveitam de todas as informações possíveis para analisar e compreender a estrutura do modelo. As provas destes métodos, suposições e como eles foram manipulados matematicamente estão descritas no decorrer do trabalho.

Estudamos os modelos com a pressuposição de que alguns parâmetros eram conhecidos, para, depois, poder adequar seu uso na prática, onde frequentemente nos deparamos com todos os parâmetros desconhecidos. Por esse motivo foram apresentados alguns algoritmos recursivos baseados nos métodos de Monte Carlo via cadeias de Markov. Vimos que o Amostrador de Gibbs pode estimar a distribuição *posteriori* dos parâmetros de estado e o FFBS é basicamente uma versão simulada das recursões de suavização, onde em teoria, os parâmetros são conhecidos e pode ser utilizado na construção do amostrador de Gibbs. Nos dias atuais, onde temos total acesso aos meios computacionais, estes algoritmos se mostram de extrema importância e é um dos motivos de a modelagem dinâmica estar sendo difundida com maior frequência atualmente.

## Referências Bibliográficas

ANDERSON, Brian D. O.; MOORE, John B.. **Optimal Filtering**. Englewood Cliffs: Prentice-Hall, 1979.

CARTER, C. e KHON, R.. **On Gibbs Sampling for State Space Models**. Oxford: Biometrika, 1994.

CHARNET, Reinaldo; FREIRE, Clarice Azevedo de Luna; CHARNET, Eugênia M. Reginato; BONVINO, Heloísa. **Análise de Modelos de Regressão Linear: Com Aplicações**. Campinas: Editora da Unicamp, 2008.

CONOVER, W. J.. **Practical Nonparametric Statistics**. New York: Wiley, 1971.

COSTA, Ana Carolina Carioca da. **Modelos Dinâmicos Hierárquicos Espaço-Temporais para Dados na Família Exponencial**. Rio de Janeiro: IM / UFRJ, 2011.

EHLERS, Ricardo S.. **Análise de Séries Temporais**. São Carlos, SP: Universidade de São Paulo, 2009. [Consultado em 07 de Maio de 2014]. Disponível em <URL:<http://www.icmc.usp.br/~ehlers/stemp/stemp.pdf>>.

FRÜWIRTH-SCHNATTER, S.. **Data Augmentation and Dynamic Linear Models**. Journal of Time Series Analysis, 1994.

GAMERMAN, Dani e LOPES, Hedibert F.. **Monte Carlo Markov Chain: Stochastic Simulation for Bayesian Inference**. London: Chapman & Hall, 2006.

KALMAN, R.. **A New Approach to Linear Filtering and Prediction Problems**. Trans. Of the AMSE – Journal of Basic Engineering (Series D), 1960.

KALMAN, R. e BUCY, R.. **New Results in Linear Filtering and Prediction Theory**. Trans. Of the AMSE – Journal of Basic Engineering (Series D), 1963.

MIGON, Helio; GAMERMAN, Dani; RODRIGUEZ, Romy. **Introdução aos Modelos Dinâmicos Bayesianos**. In: XXVI Simpósio Brasileiro de Pesquisa Operacional – SBPO, São João Del Rey, 2004.

PETRIS, Giovanni; PETRONE, Sonia; CAMPAGNOLI, Patrizia. **Dynamic Linear Models with R**. New York: Springer, 2009.

SHEPARD, N.. **Partial Non-Gaussian State Space Models**. Oxford: Biometrika, 1994.

WEST, Mike; HARRISON, Jeff. **Bayesian Forecasting and Dynamic Models**. New York: Springer, 1997.