

**XXXVI SBPO - Simpósio Brasileiro de Pesquisa Operacional**  
São João del-Rei, 23 - 26 nov 2004

## Introdução aos Modelos Dinâmicos Bayesianos

Helio Migon, Dani Gamerman & Romy Rodriguez

Instituto de Matemática  
Universidade Federal de Rio de Janeiro

## Definição e Resultados Principais

- 1 Introdução
- 2 Revisão de Probabilidade e Inferência Bayesiana
- 3 Modelo Linear Dinâmico
- 4 Modelos de Tendência
- 5 Evolução

## Tópicos Especiais

- 6 Análise Retrospectiva
- 7 Superposição de Modelos
- 8 Modelos com Variáveis Causais
- 9 Modelos com Sazonalidade
- 10 Monitoração
- 11 Intervenção
- 12 Análise de Dados

## Extensões

- 13 Modelos Lineares Dinâmicos Generalizados (MLDG)
- 14 Inferência em MLDG: Linear Bayes
- 15 Inferência em MLDG: MCMC
- 16 Modelos Não Lineares Dinâmicos Generalizados (MNLDG)
- 17 Modelos Dinâmicos Generalizados (MDG)
- 18 Inferência em MDG: MCMC seqüencial

- Pole, West & Harrison (1994). *Applied Bayesian Forecasting and Time Series Analysis*. New York: Chapman-Hall.
- West & Harrison (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Varlag.

## Software:

**BATS** <ftp://stat.duke.edu/pub/bats/>

**WinBUGS** <http://www.mrc-bsu.cam.ac.uk/bugs/>

**Bts** <http://lib.stat.cmu.edu/DOS/S/>  
(SPLUS for Windows functions)

**Ox** <http://www.doornik.com/download.html>

# Parte I

## Definição e Resultados Principais

# Introdução

---

- **Série Temporal (ST)** é uma seqüência de observações ao longo do tempo.
- Nos modelos de regressão usuais a ordem das observações é irrelevante, na **ST** a passagem do tempo tem efeito marcante.
- Normalmente numa **ST** as observações são equiespaçadas. Caso não sejam, isso pode ser acomodado com mudanças na escala e observações ausentes. Deve-se, entretanto, tomar cuidado com as escalas de medição. A abordagem Bayesiana (diferentemente de outras) incorpora isso.
- Neste curso, só estudaremos **ST** univariadas (possivelmente com variáveis explicativas ou regressores).



- **Controle** é uma descrição do que já aconteceu
- **Previsão** é uma descrição do que vai acontecer

*Os dois podem ser feitos independentemente.*

- A abordagem aqui é baseada na construção de um modelo.
- Tendo o modelo, pode-se fazer previsão, controle ou ambos.

Numa situação real teremos o ciclo:

... **previsão**  $\Rightarrow$  **observação**  $\Rightarrow$  **análise** ...

Estudos na área de modelos e previsão tiveram um primeiro impulso em engenharia de sistemas nos anos 60.

Lá, o interesse era voltado para sistema de funcionamento de máquinas (por exemplo, satélites) e havia uma ênfase grande em controle.

Embora os desenvolvimentos subseqüentes em Estatística e Engenharia de Sistemas tomaram caminhos distintos, boa parte do curso será voltada para a base comum sobre a qual foram desenvolvidas as extensões.

Três componentes englobam boa parte dos modelos usados na prática:

- Modelos de Tendência
- Modelos de sazonalidade (ou ciclos sistemáticos)
- Modelos de variáveis causais (ou regressores)

A combinação das três componentes fornece modelos para séries:

- financeiras (vendas, estoque);
- industriais (produção, capacidade operativa);
- agrícolas (produção de leite, mercado de carnes);
- médicas (monitoração de órgãos);
- sociais (acidentes, nascimentos).

**Figura: Vendas mensais de um tipo de bala de Janeiro de 1976 a Dezembro de 1981 (CANDY.DAT)**

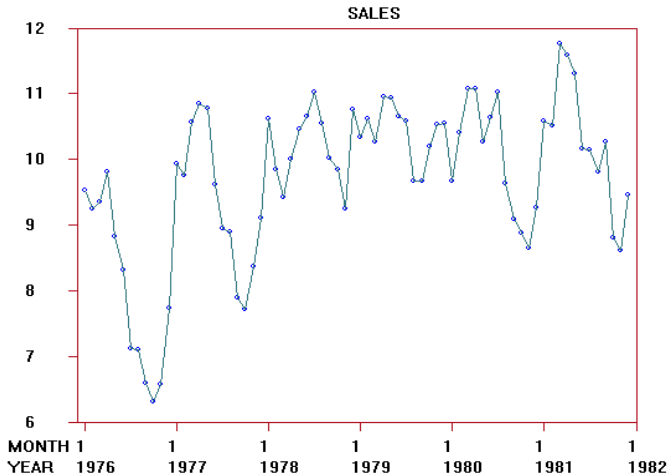
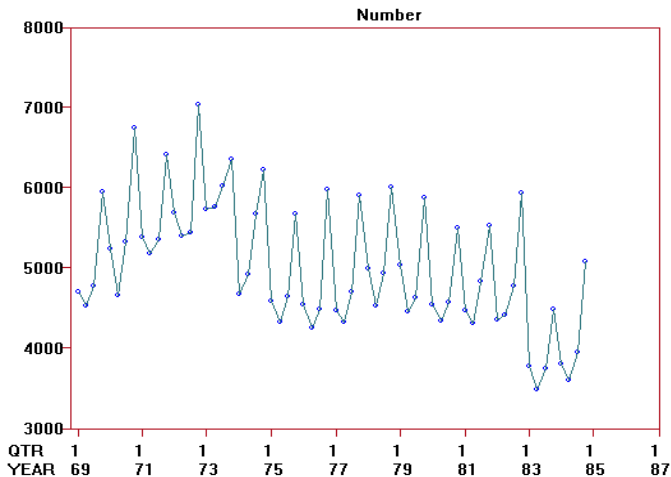


Figura: Acidentes rodoviários graves (1969:1-1984:4)



- A maioria das análises estatísticas utilizam *modelos estáticos*: Modelos com uma descrição fixa (através de parâmetros fixos) ao longo das unidades de observação. Exemplos: análise de regressão, MLG, modelos ARMA.
- Em **ST**, essa hipótese muitas vezes é violada: **as estruturas mudam com a passagem do tempo**.
- **ST** ligados às atividades humanas são alvos de mudanças:
  - ✓ Abruptas - devido a grandes mudanças, hecatombes, novas leis;
  - ✓ Graduais.

- Neste curso todos os **modelos são dinâmicos**: a descrição (os parâmetros) muda com a passagem do tempo. Eles incluem como caso particular os modelos estáticos (onde a mudança é nula).
- Normalmente, a passagem do tempo traz novas observações e aumenta o nosso conhecimento.
- Em modelos dinâmicos temos também perda de informação devido à passagem do tempo.  
Exemplo: o nível de vendas mês passado é mais relevante hoje que o nível de vendas em setembro.
- Construção do modelo dinâmico é feita em duas etapas: 1a. qualitativa e 2a. quantitativa de uma forma local. Em modelos estáticos, a mesma quantificação é válida globalmente.

- A construção de um modelo é uma arte.
- Um **modelo** é uma representação de uma realidade; será tão adequado quanto à sua capacidade de alcançar os objetivos a que ele se destina. Portanto, a construção do modelo traz inerente em si um caráter subjetivo.
- **Previsão** é uma afirmação sobre um futuro incerto.
- A **incerteza** aqui será sempre representada através de probabilidade. Portanto, a previsão será sempre formulada em termos de probabilidade condicionada ao nosso estado de conhecimento. Se ele muda, nossa previsão mudará.



- Nosso conhecimento provém de duas fontes:

- ✓ A série histórica ou dados;
- ✓ Outros conhecimentos (subjetivos)

Exemplos: entrada em vigor de leis, falência de competidor.

- Ambas as fontes são importantes, podem e devem ser utilizadas.
- A abordagem Bayesiana incorpora esses elementos natural e coerentemente.
- Aplicando à previsão, significa que o modelo *default* é posto para funcionar.
- Se acontecimentos não rotineiros intervêm, o modelo os incorpora:
  - ✓ Preparando para mudança e/ou
  - ✓ Alterando o que for necessário

Exemplo: se o competidor vai falir, precisa usar seu conhecimento sobre a divisão do mercado para formular a mudança que ele espera que aconteça.

# Revisão de Probabilidade e Inferência Bayesiana

---

## Principais Resultados

- **Probabilidade** é um número entre 0 e 1 representando a crença numa determinada afirmação. Ex:
  - $\Pr(\text{"cara no lançamento de uma moeda"})=0.5$ ,
  - $\Pr(\text{"chover hoje"})=0.1$ ,
  - $\Pr(\text{"Ibis ser o campeão"})=0.01$
- **Probabilidades Totais** (0 ou 1) representam crença na veracidade ou falsidade de uma afirmação.
- **Probabilidade Condicional** é a probabilidade baseada no conhecimento prévio da veracidade de uma afirmação. Ex:  $\Pr(\text{"dado lançado dar par"} \mid \text{"resultado foi } \leq 5\text{"})$ .
  - É calculada com:  $\Pr(A|B) = \frac{\Pr(AB)}{\Pr(B)}$ .  
Sendo:  $A = \{2, 4, 6\}$ ,  $B = \{1, 2, 3, 4, 5\}$ ,  $AB = \{2, 4\}$ ,
  - Daí:  $\Pr(AB) = 2/6$  e  $\Pr(B) = 5/6$ .  
Logo,  $\Pr(A|B) = \frac{2/6}{5/6} = \frac{2}{5}$

# Teorema de Bayes

Também temos que  $Pr(B|A) = \frac{Pr(AB)}{Pr(A)}$   
assim,  $Pr(AB) = Pr(B|A) \times P(A)$ . Logo,

$$\begin{aligned} Pr(A|B) &= \frac{Pr(AB)}{Pr(B)} \\ &= \frac{Pr(B|A) \times Pr(A)}{Pr(B)} \\ &\propto Pr(B|A)Pr(A) \end{aligned}$$

O resultado acima é conhecido como **Teorema de Bayes**.  
Ele fornece a base da abordagem Bayesiana pois nos *ensina como atualizarmos nossa crença em A após receber novas informações, no caso, B*

- *Variável ou Quantidade Aleatória* é aquela cujo valor nos é incerto. Ex.:
  - $X$  : Número resultando do lançamento de um dado
  - $Y$  : Nível de glicose no sangue de um indivíduo.
- Sua incerteza é representada probabilisticamente, p.e.,  $Pr(X = x) = \frac{1}{6}, x = 1, 2, 3, 4, 5, 6.$
- As Variáveis Aleatórias admitem várias classificações possíveis:
  - *Discreta* ou *Contínua*
  - *Observável* ou *Não Observável*

## Exemplos:

- Discreta e observável:  $X$
- Discreta e não observável: Indicador de doença num indivíduo
- Contínua e observável: Vendas de um produto em larga escala
- Contínua e não observável:  $Y$

**Variáveis Discretas** são representadas pela **função de probabilidade**  $f(x) = Pr(X = x)$ , que caracteriza completamente a incerteza a respeito de  $X$  pois

$$Pr(X \in A) = \sum_{x \in A} Pr(X = x) = \sum_{x \in A} f(x)$$

**Variáveis Contínuas** são representadas pela função de densidade de probabilidade ou **densidade**  $f(x)$ , que caracteriza a incerteza a respeito de  $X$  pois

$$Pr(X \in A) = \int_A f(x) dx$$

Exemplo:  $f(x) = 1/[\pi(1 + x^2)]$  daqui,  $Pr(X \in [-1, 2]) = 0,6$

- A *função de probabilidade* e a *densidade* definem uma **distribuição de probabilidade**.
- Características importantes de uma distribuição de probabilidade:
  - *Medidas de Posição*:
    - Moda: valor mais provável,
    - Média: centro de gravidade, denotado por  $E[X]$
  - *Medidas de dispersão*:
    - Variância: denotado por  $V[X]$ ,
    - Desvio-Padrão:  $\sqrt{V[X]}$

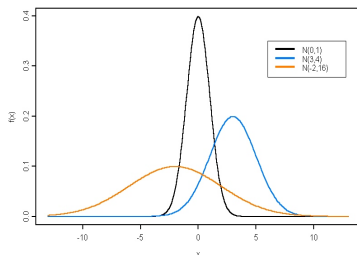
# Distribuição Normal

A v.a (contínua)  $X$  que tem densidade

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right\}$$

é dita ter *distribuição normal com média  $\mu$  e variância  $\sigma^2$* .

Notação:  $X \sim N(\mu, \sigma^2)$





- A distribuição normal é simétrica e surge como resultante de processos onde muitas fontes de incerteza intervêm.
- Infelizmente  $Pr(X \in A) = \int_A f(x)dx$  só pode ser calculada numericamente.
- Resultados importantes para a normal padrão ( $N(0, 1)$ ):
  - $Pr(X \in [-2, 2]) = 0.95$
  - $Pr(X \in [-1.64, 1.64]) = 0.90$

# Teorema de Bayes para Variáveis Aleatórias

Suponha a existência de uma quantidade de interesse (desconhecida) que chamaremos de  $\mu$ . Nossa incerteza sobre  $\mu$  é representada pela densidade  $p(\mu)$ .

Posteriormente, observamos uma outra quantidade  $X$  relacionada a  $\mu$  (por exemplo, uma medição) cuja incerteza é representada pela densidade condicional  $f(x|\mu)$ .

Após observar  $X = x$ , nossa incerteza sobre  $\mu$  passa a ser refletida pela densidade condicional  $p(\mu|x)$ .

Pelo teorema de Bayes,

$$p(\mu|x) = \frac{f(x|\mu)p(\mu)}{f(x)} \propto f(x|\mu)p(\mu)$$

*Importante é que o teorema nos ensina como atualizar nosso conhecimento após receber informação relevante.*

# Teorema de Bayes para Variáveis Normais

- Sabe-se que o nível de glicose em uma pessoa normal pode ser descrito pela distribuição  $N(120, 100)$ , i.e,

$$p(\mu) = \frac{1}{\sqrt{200\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{\mu - 120}{10} \right)^2 \right\}$$

Logo,  $Pr(\mu \text{ estar entre } 100 \text{ e } 140) = 0,95$

- Uma medição,  $y$ , em laboratório é feita. Vamos supor que a  $y|\mu$  é  $N(\mu, 25)$ , i.e.

$$f(y|\mu) = \frac{1}{\sqrt{50\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{y - \mu}{5} \right)^2 \right\}$$

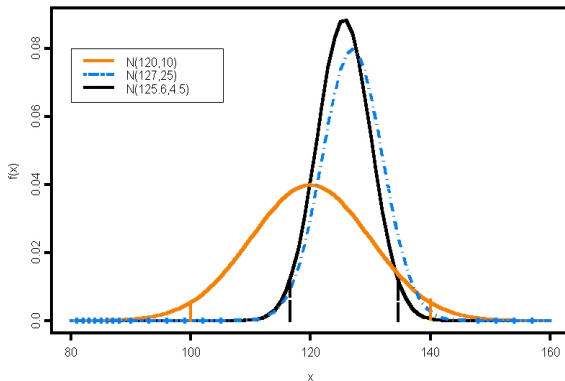
daí, as chances da medição errar o nível em mais de 10 ml são de 5%

- ✓ A amostra é colhida e a **medição observada é de 127 ml**.
- ✓ Agora, nossa incerteza sobre o nível de glicose é dado por

$$\begin{aligned} p(\mu|y = 127) &\propto f(127 - \mu)p(\mu) \\ &\propto \exp\left\{-\frac{1}{2}\left(\frac{127 - \mu}{5}\right)^2\right\} \exp\left\{-\frac{1}{2}\left(\frac{\mu - 120}{10}\right)^2\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left(\frac{\mu - 125.6}{4.5}\right)^2\right\} \end{aligned}$$

Logo,  $(\mu|y=127)$  é  $N(125.6, (4.5)^2)$  e  
 $Pr(\mu \text{ estar entre } 116.6 \text{ e } 134.6) = 0.95$

Figura: Teorema de Bayes para Variáveis Normais



- Uma **coleção de variáveis ou quantidades aleatórias** pode ser agrupada num **vetor aleatório**.
- Se todas as componentes de um vetor aleatório forem discretas, o vetor também será **discreto** com função de probabilidade:

$$f(x_1, \dots, x_p) = Pr(X_1 = x_1, \dots, X_p = x_p)$$

$$\text{e } Pr((X_1, \dots, X_p) \in A) = \sum_{(x_1, \dots, x_p) \in A} f(x_1, \dots, x_p)$$

- Se todas as componentes de um vetor forem contínuas, o vetor também será **contínuo** com densidade  $f(x_1, \dots, x_p)$  e

$$Pr((X_1, \dots, X_p) \in A) = \int_A f(x_1, \dots, x_p) dx_1 \dots dx_p$$

- A generalização multivariada da normal é a distribuição normal multivariada com vetor de médias  $\mu$  e matriz de variâncias-covariâncias  $\Sigma$  denotada por  $N(\mu, \Sigma)$ .
- Se  $(X_1, \dots, X_p) \sim N(\mu, \Sigma)$ , então cada componente  $X_j \sim N(\mu_j, \sigma_{jj})$  onde:  $\mu_j$  é o  $j$ -ésimo componente do vetor  $\mu$   $\sigma_{jj}$  é o  $j$ -ésimo componente da diagonal da matriz  $\Sigma$ .
- Finalmente, o teorema de Bayes também leva a uma normal multivariada se a observação for normal e o parâmetro for normal multivariado.

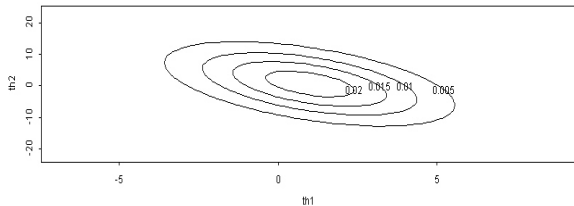
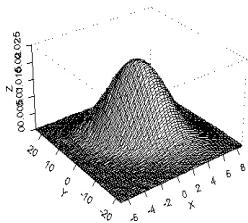


Figura: Caso particular: Normal Bivariada



- Teorema de Bayes para variáveis aleatórias:

$$p(\mu|x) = \frac{f(x|\mu)p(\mu)}{f(x)} \propto f(x|\mu)p(\mu)$$

- Normal Multivariada - Distribuição Condicional

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \Big| \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim N \left[ \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{12} & \Sigma_2 \end{pmatrix} \right]$$

$$Y_1 | Y_2, \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim N[\mu_1 + \Sigma_{12}\Sigma_2^{-1}(Y_2 - \mu_2), \Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{21}]$$

- Normal- Gama Multivariada

Se  $\mathbf{Y} | \boldsymbol{\mu}, \phi, \boldsymbol{\Sigma} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}\phi^{-1})$  e  $\phi \sim \text{Gama}(n/2, d/2)$  então:

$$\mathbf{Y} \sim t_n(\boldsymbol{\mu}, d/n\boldsymbol{\Sigma})$$

# Modelo Linear Dinâmico

---

# O Modelo de Regressão Usual

Num modelo de regressão temos uma variável resposta  $y$  que é explicada por um conjunto de variáveis explicativas  $x_1, \dots, x_p$  através da relação

$$y = \theta_0 + \theta_1 x_1 + \dots + \theta_p x_p + v$$

Em geral, assume-se que o  $v$  tem distribuição  $N(0, \sigma^2)$ .

A equação acima pode ser mais compactamente escrita como:

$$y = \mathbf{x}'\boldsymbol{\theta} + v,$$

$$\text{onde: } \mathbf{x} = \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_p \end{pmatrix} \quad \text{e} \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_p \end{pmatrix}$$

- A natureza das variáveis explicativas ou regressores é bastante ampla. Podendo assim, utilizar-se qualquer variável quantificável.
- Os coeficientes de regressão  $\theta_1, \dots, \theta_p$  informam sobre a influência que os regressores têm sobre a resposta  $y$ .
- Na prática, seus valores são desconhecidos e estimados a partir de uma coleção de observações feitas sobre o modelo acima.

Assim, observamos respostas  $y_1, \dots, y_n$  com seus respectivos regressores  $x_1, \dots, x_n$ . Simbolicamente, temos:

$$y_t = \mathbf{x}'_t \boldsymbol{\theta} + v_t, \quad t = 1, \dots, n.$$

# Definição do Modelo

Em modelos dinâmicos os parâmetros mudam com o passar do tempo. O modelo de regressão é estendido para

$$y_t = x_t' \theta_t + v_t, \quad t = 1, \dots, n.$$

onde a única mudança em relação ao modelo de regressão foi a indexação de  $\theta$

- A formulação acima cria uma profusão de parâmetros a serem estimados.
- O modelo acima necessita de mais informação. Essa informação vem do fato que os parâmetros sucessivos estão intimamente relacionados.
- Em geral, um parâmetro é igual ao seu antecessor mais uma pequena perturbação causada pelas mudanças às quais o sistema está sujeito.

Se o sistema é estático, como em regressão, temos:

$$\theta_t = \theta_{t-1} = \theta.$$

Em modelos dinâmicos, vamos admitir a forma mais geral

$$\theta_t = G_{t-1}\theta_{t-1} + w_t$$

onde  $G_t$  contém valores conhecidos e  $w_t$  é uma perturbação aleatória.

- A equação acima é conhecida como **equação do sistema**.
- A **matriz de evolução**  $G_t$  controla a parte determinística da evolução do sistema e estabelece a propagação do sistema ao longo do tempo.
- A **perturbação**  $w_t$  é responsável pela introdução de incertezas devidas à passagem do tempo e conseqüente perda de informação.
- Note que se  $G_t = I$  e  $w_t = 0$ , o modelo se reduz ao caso estático.

## Exemplo

Se observamos uma série de vendas ( $y$ ) explicada pela respectiva série de preços ( $x$ ) através de uma relação estável teremos:

$$\begin{aligned}y_t &= \mu_t + \beta_t x_t + v_t \\ \mu_t &= \mu_{t-1} + \Delta\mu_t \\ \beta_t &= \beta_{t-1} + \Delta\beta_t\end{aligned}$$

Assim,

- As *vendas* são explicadas pelos *preços* em uma regressão dinâmica.
- A maior (ou menor) estabilidade dessa relação será controlada pela magnitude dos incrementos  $\Delta\mu_t$  e  $\Delta\beta_t$ .

O **modelo linear dinâmico** pode então ser definido como:

- Equação das Observações:

$$y_t = F_t' \theta_t + v_t, \quad v_t \sim N(0, V)$$

- Equação do Sistema:

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

No exemplo acima,  $y_t = \text{venda}$ ,  $x_t = \text{preço}$ ,  $F_t' = (1, x_t)$

$$\theta = \begin{pmatrix} \mu \\ \beta \end{pmatrix}, \quad G_t = I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{e} \quad w_t = \begin{pmatrix} \Delta\mu \\ \Delta\beta \end{pmatrix}$$





Previsões no modelo dinâmico são obtidas pela combinação da informação a priori com a equação das observações.

- A combinação de  $y_t = F_t'\theta_t + v_t$  com a priori  $p(\theta_t|D_{t-1})$  permite a obtenção da **distribuição preditiva**:

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

baseado na qual as previsões serão feitas.

- Em particular, se quisermos
  - uma **previsão pontual**: podemos tomar a média dessa distribuição,  $E[y_t|D_{t-1}]$ .
  - um **intervalo de predição** de 90% de probabilidade, basta tomar  $A$  de forma que

$$0,90 = Pr(y_t \in A|D_{t-1}) = \int_A p(y_t|D_{t-1})dy_t$$

A **Previsão vários passos a frente** é feita de forma similar. Se temos interesse em prever  $y_{t+k}$  no tempo  $t + 1$ , precisamos utilizar a equação do sistema sucessivamente até podermos escrever  $\theta_{t+k}$  como função de  $\theta_{t-1}$ . Por exemplo, combinando

$$\theta_{t-1} = G_{t+1}\theta_t + w_{t+1} \quad \text{e} \quad \theta_t = G_t\theta_{t-1} + w_t$$

pode-se obter:

$$\begin{aligned}\theta_{t+1} &= G_{t+1}[G_t\theta_{t-1} + w_t] + w_{t+1} \\ &= G_{t+1}G_t\theta_{t-1} + G_{t+1}w_t + w_{t+1}\end{aligned}$$

A partir daí, combina-se com a equação das observações no tempo  $t + k$  para obter a distribuição preditiva  $p(y_{t+k}|D_{t-1})$ . **Previsões cumulativas para os próximos  $k$  períodos**, isto é,  $y_t + y_{t+1} + \dots + y_{t+k-1}$  também podem ser obtidas pelo mesmo método.

**Tabela: DLM univariado: variância  $V_t$  conhecida**

Eq. Observ.:	$y_t = F_t' \theta_t + \nu_t$	$\nu_t \sim N[0, V_t]$
Eq. Sistema:	$\theta_t = G_t \theta_{t-1} + \omega_t$	$\omega_t \sim N[0, W_t]$
Informação:	$(\theta_{t-1}   D_{t-1}) \sim N[m_{t-1}, C_{t-1}]$	
Priori:	$(\theta_t   D_{t-1}) \sim N[a_t, R_t]$	$a_t = G_t m_{t-1}$ $R_t = G_t C_{t-1} G_t' + W_t$
Previsão:	$(y_t   D_{t-1}) \sim N[f_t, Q_t]$	$f_t = F_t' a_t$ $Q_t = F_t' R_t F_t + V_t$

## Tabela: DLM univariado: variância $V_t$ conhecida (cont.)

- Relações recursivas de atualização

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

$$m_t = a_t + A_t e_t$$

$$C_t = R_t - A_t A_t' Q_t$$

$$e_t = y_t - f_t$$

$$A_t = R_t F_t / Q_t$$

- Distribuições preditivas para  $k \geq 1$

$$(\theta_{t+k} | D_t) \sim N[a_t(k), R_t(k)]$$

$$(y_{t+k} | D_t) \sim N[f_t(k), Q_t(k)]$$

$$a_t(k) = G_{t+k} a_t(k-1)$$

$$R_t(k) = G_{t+k} R_t(k-1) G_{t+k}' + W_{t+k}$$

$$f_t(k) = F_{t+k}' a_t(k)$$

$$Q_t(k) = F_{t+k}' R_t(k) F_{t+k} + V_{t+k}$$

$$a_t(0) = m_t$$

$$R_t(0) = C_t$$

**Tabela: DLM: variância desconhecida**  $V_t = k_t \phi^{-1}$ ,  $k_t$  conhecido

Eq. Observ.:	$y_t = F_t' \theta_t + \nu_t$	$\nu_t \sim N[0, k_t \phi^{-1}]$
Eq. Sistema:	$\theta_t = G_t \theta_{t-1} + \omega_t$	$\omega_t \sim t_{n_{t-1}}[0, W_t]$
Informação:	$(\theta_{t-1}   D_{t-1}) \sim t_{n_{t-1}}[m_{t-1}, C_{t-1}]$ $(\phi_{t-1}   D_{t-1}) \sim G[n_{t-1}/2, d_{t-1}/2]$	
Priori:	$(\theta_t   D_{t-1}) \sim t_{\delta_t n_{t-1}}[a_t, R_t]$ $a_t = G_t m_{t-1}$ $R_t = G_t C_{t-1} G_t' + W_t$ $(\phi_t   D_{t-1}) \sim G[\delta_t n_{t-1}/2, \delta_t d_{t-1}/2]$	
Previsão:	$(y_t   D_{t-1}) \sim t_{\delta_t n_{t-1}}[f_t, Q_t]$ $f_t = F_t' a_t$ $Q_t = F_t' R_t F_t + k_t S_{t-1}$	

**Tabela: DLM: variância desconhecida**  $V_t = k_t \phi^{-1}$ ,  $k_t$  conhecido

- **Relações recursivas de atualização**

$$(\theta_t | D_t) \sim t_{n_t}[m_t, C_t]$$

$$(\phi_t | D_t) \sim G[n_t/2, d_t/2]$$

$$m_t = a_t + A_t e_t$$

$$C_t = (S_t/S_{t-1})R_t - A_t A_t' Q_t$$

$$e_t = y_t - f_t, A_t = R_t F_t / Q_t$$

$$n_t = \delta_t n_{t-1} + 1, d_t = \delta_t d_{t-1} + S_{t-1} e_t^2$$

$$S_t = d_t / n_t$$

- **Distribuições preditivas para  $k \geq 1$**

$$(\theta_{t+k} | D_t) \sim t_{\delta_t n_t}[a_t(k), R_t(k)]$$

$$(y_{t+k} | D_t) \sim t_{\delta_t n_t}[f_t(k), Q_t(k)]$$

$$a_t(k) = G_{t+k} a_t(k-1)$$

$$R_t(k) = G_{t+k} R_t(k-1) G_{t+k}' + W_{t+k}$$

$$f_t(k) = F_{t+k}' a_t(k)$$

$$Q_t(k) = F_{t+k}' R_t(k) F_{t+k} + k_{t+k} S_t$$

$$a_t(0) = m_t,$$

$$R_t(0) = C_t$$

# Modelos de Tendência

---



O mais simples modelo dinâmico é o modelo de **tendência estável** ou **modelo polinomial de primeira ordem**. Ele é composto apenas de *um nível que varia segundo um passeio aleatório*:

$$y_t = \mu_t + v_t, \quad v_t \sim N(0, V_t)$$
$$\mu_t = \mu_{t-1} + w_t, \quad w_t \sim N(0, W_t)$$

- Segundo esse modelo, o nível permanece localmente constante, mas varia quando se considera longos períodos de tempo.
- Usualmente, a variação das observações em torno dos níveis (medida por  $V$ ) é bem maior que as variações temporais do nível ao longo do tempo (medidas por  $W$ ).
- Ele é obtido ao particularizar o modelo dinâmico com  $F_t = 1$  e  $G_t = 1$ .

O **modelo polinomial de segunda ordem** ou modelo de **tendência linear**, permite um *crescimento no nível*. Isso é quantificado através de um parâmetro adicional e o modelo fica:

$$y_t = \mu_t + v_t$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + w_{1t}$$

$$\beta_t = \beta_{t-1} + w_{2t}$$

Esse modelo é obtido com  $F_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  e  $G_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$

Aqui, o nível permanece localmente linear, mas a forma da reta pode variar com o tempo.

*Tendências de ordem maior são pouco utilizadas.*

- Os dados deste exemplo correspondem as vendas de uma bala (SALES) do arquivo CANDY.DAT do pacote BATS.  
(ver série)
- Nas figuras a seguir apresentam-se alguns resultados do ajuste de modelos polinomiais de primeira e segunda ordem a série de Vendas Mensais de Janeiro de 1976 a Dezembro de 1981.

Figura: Estimativa do nível

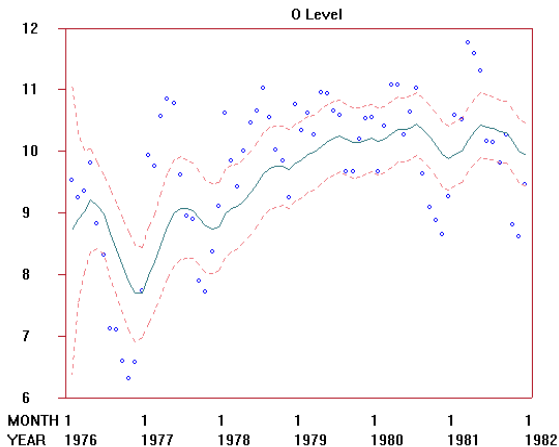


Figura: Estimativa das vendas



Figura: Previsão um passo a frente

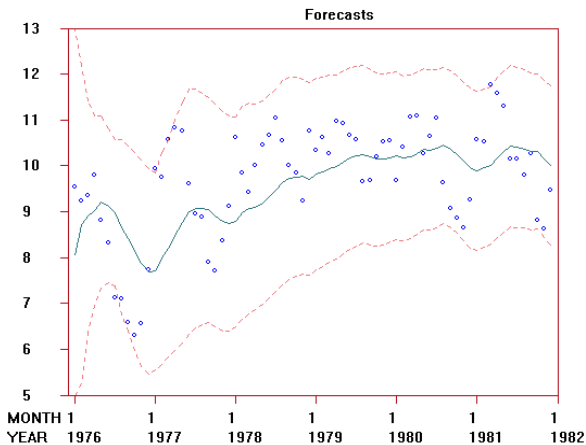


Figura: Estimativa do nível

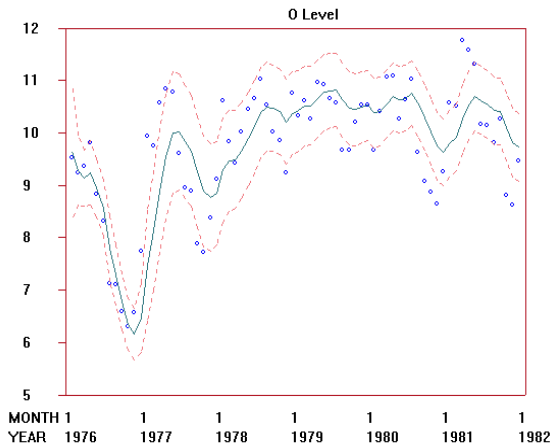


Figura: Estimativa das vendas

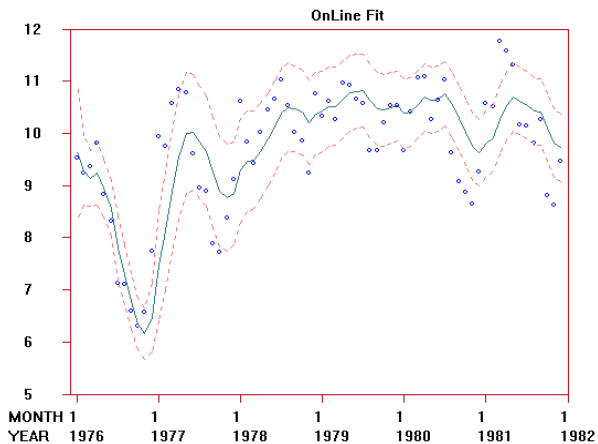
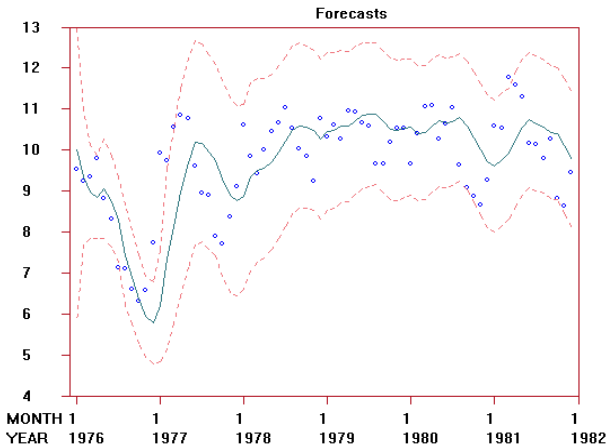




Figura: Previsão um passo a frente



# Modelo Polinomial de Primeira Ordem

$$\begin{aligned} y_t &= \mu_t + \nu_t, & \nu_t &\sim N[0, V_t] & \text{ou} & & y_t | \mu_t &\sim N[\mu_t, V_t] \\ \mu_t &= \mu_{t-1} + \omega_t, & \omega_t &\sim N[0, W_t] & \text{ou} & & \mu_t | \mu_{t-1} &\sim N[\mu_{t-1}, W_t] \end{aligned}$$

onde  $\mu_t$ : nível da série no instante  $t$ ,  $\nu_t$ : erro da equação de observação e  $\omega_t$ : erro da equação de evolução.

Função de previsão:

$$f_t(k) = E[y_{t+k} | D_t] = E[\mu_t | D_t] = m_t, \quad \forall k > 0$$

pois  $E[y_{t+k} | \mu_t] = E[\mu_{t+k} | \mu_t] = \mu_t$

**Tabela:** Equações de atualização para  $\{1, 1, V_t, W_t\}$

Informação inicial	$(\mu_0 D_0) \sim N[m_0, C_0]$
(a) Posteriori para $\mu_{t-1}$	$(\mu_{t-1} D_{t-1}) \sim N[m_{t-1}, C_{t-1}]$
(b) Priori para $\mu_t$	$(\mu_t D_{t-1}) \sim N[m_{t-1}, R_t]$ $R_t = C_{t-1} + W_t$
(c) Previsão um passo à frente	$(y_t D_{t-1}) \sim N[f_t, Q_t]$ $f_t = m_{t-1}$ $Q_t = R_t + V_t$
(d) Posteriori para $\mu_t$	$(\mu_t D_t) \sim N[m_t, C_t]$ $m_t = m_{t-1} + A_t e_t$ $C_t = A_t V_t$ $A_t = R_t / Q_t$ $e_t = y_t - f_t$

- $e_t$  : erro de previsão um passo à frente  $e_t = y_t - f_t = y_t - m_{t-1}$
- $A_t = \frac{R_t}{Q_t} = \frac{R_t}{R_t + V_t} = \frac{C_{t-1} + W_t}{C_{t-1} + W_t + V_t}$
- $A_t : \rho_t^2(y_t, \mu_t)$  ou  $\hat{\beta}$  em  $\mu_t = \alpha + \beta y_t$   $0 \leq A_t \leq 1$
- $m_t = m_{t-1} + \rho_t^2(y_t - m_{t-1}) = A_t y_t + (1 - A_t)m_{t-1}$

### Tabela: Distribuições preditivas

previsão  $k$  passos à frente:

$$(y_{t+k} | D_t) \sim N[m_t, Q_t(k)]$$

$$Q_t(k) = C_t + \sum_{j=1}^k W_{t+j} + V_{t+k}$$

$k$ -step lead time forecast:

$$X_t(k) = y_{t+1} + y_{t+2} + \dots + y_{t+k}, k > 0$$

$$(X_{t+k} | D_t) \sim N[km_t, L_t(k)]$$

$$L_t(k) = k^2 C_t + \sum_{j=1}^k V_{t+j} + \sum_{j=1}^k j^2 W_{t+k+1-j}$$

# Evolução

---

- Até agora, não foi discutido o processamento da incerteza relativa a  $W$ .
- A abordagem Bayesiana para parâmetros desconhecidos é sempre a mesma: **atualização via teorema de Bayes**.
- O tratamento dado à variância das observações é analítico. O mesmo não acontece com a variância do sistema.

Felizmente, existe uma solução baseada em **fatores de desconto** que produz uma alternativa aceitável.

- Como já dissemos, o *valor da informação diminui com o tempo*.
- Essa diminuição é controlada pela evolução do sistema, através do aumento da incerteza do sistema.

No modelo estável, temos que

$$V[\mu_t|D_{t-1}] = V[\mu_{t-1}|D_{t-1}] + W_t$$

Como a **variância mede a dispersão do sistema**, o seu inverso, a **precisão mede a informação do sistema**.

Quanto mais dispersa for a variável, maior será sua variância e menor será sua precisão. Logo, dispomos de menos informação sobre essa variável.

Pensando agora em **percentagem ou fração de informação perdida com a passagem de tempo**, podemos definir um

**fator de desconto**  $\delta \in (0, 1]$ , tal que

$$V^{-1}[\mu_t|D_{t-1}] = \delta V^{-1}[\mu_{t-1}|D_{t-1}]$$

As duas equações acima fornecem uma base para especificação de  $W_t$ .

O **fator de desconto**  $\delta$  é a percentagem de informação que passa de um período a outro.

- Valores típicos para sistemas sem variações bruscas se encontram acima de 90%
- A escolha do valor adequado vai depender da aplicação e sugere-se que alguns valores sejam comparados.
- Valores muito próximos não produzem diferenças perceptíveis.
- Valores muito **baixos** (abaixo de 0,8) tendem a introduzir **muita incerteza** e produzem limites de incerteza para predição muito grandes.
- Valores muito **altos** representam um sistema com **mudanças muito suaves**.
- No limite, quando  $\delta = 1$ , temos o modelo estático onde não há perda de informação.
- A mesma idéia de desconto pode ser estendida a modelos mais gerais com vários descontos aplicados a partes diferentes do modelo. Essa formulação ficará mais clara quando abordamos superposição de modelos.



- Ilustramos o uso de fatores de desconto com a série de vendas de uma bala (SALES) do arquivo CANDY.DAT do pacote BATS. (ver série)
- As figuras a seguir apresentam as previsões um passo a frente resultantes do ajuste de um modelo de tendência constante com os seguintes fatores de desconto: 1,0; 0,9 e 0,8

Figura: Fator de Desconto  $\delta = 1$  (estrutura estática)

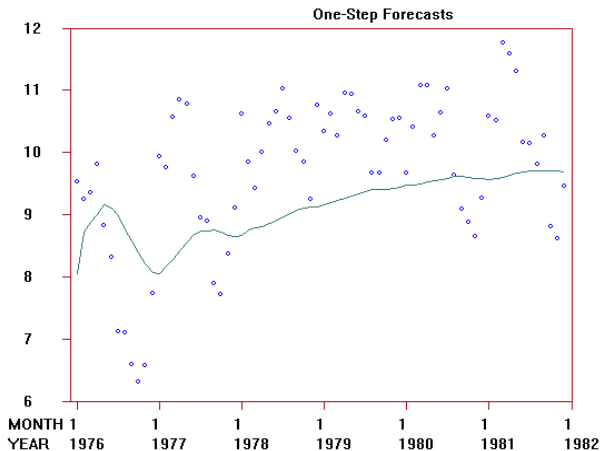
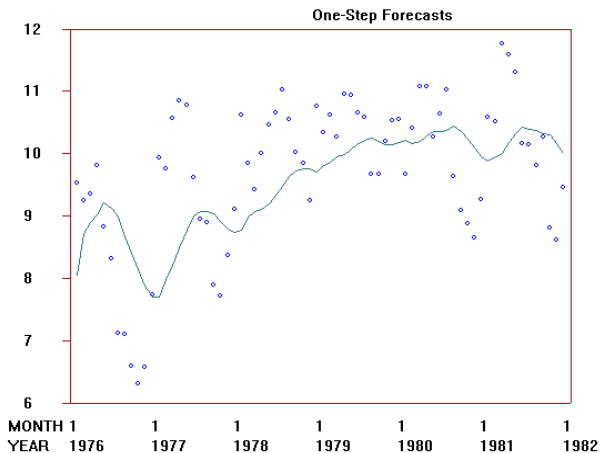
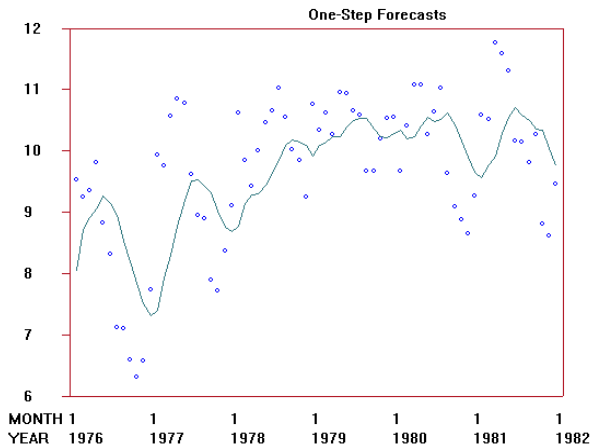


Figura: Fator de Desconto  $\delta = 0,9$



BATS: Trend: Constant, Discount:Trend=0.90,Variance=0.99

Figura: Fator de Desconto  $\delta = 0,8$



BATS: Trend: Constant, Discount:Trend=0.80,Variance=0.99

## Parte II

# Tópicos Especiais

# Análise Retrospectiva

---

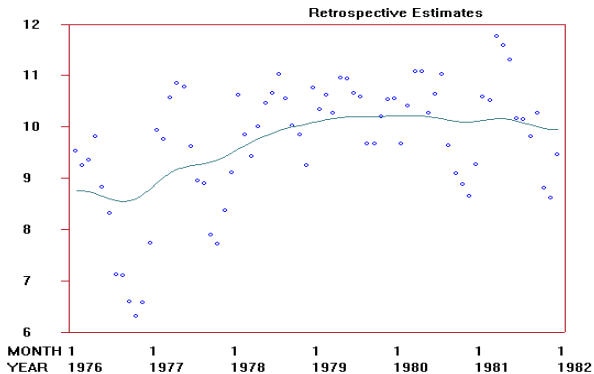
- **Análise retrospectiva** usa toda a série observada para **reavaliar a inferência realizada** durante o procedimento seqüencial.
- Essa reavaliação é devida à utilização de observações colhidas após o período de interesse.
- Com mais informação, sabemos mais e dispomos de mais instrumentos para entender o que se passou.
- Essa operação de passagem de informação para trás no tempo é chamada de **suavização ou análise retrospectiva**.

Da análise seqüencial, obtemos  $p(\theta_t|D_t)$ . Se coletamos observações até o tempo  $t + k$ , a melhor descrição de nossa incerteza sobre  $\theta_t$  é através de  $p(\theta_t|D_{t+k})$ . Observe, no entanto, que só podemos nos beneficiar dessa regra, após serem decorridos  $k$  períodos de tempo.

- Ilustramos o resultado da análise retrospectiva com a série de vendas de uma bala (SALES) do arquivo CANDY.DAT do pacote BATS. (ver série)
- As figuras a seguir apresentam as estimativas suavizadas para as vendas e para o nível de série obtidas a partir do ajuste de um modelo polinomial de primeira ordem.

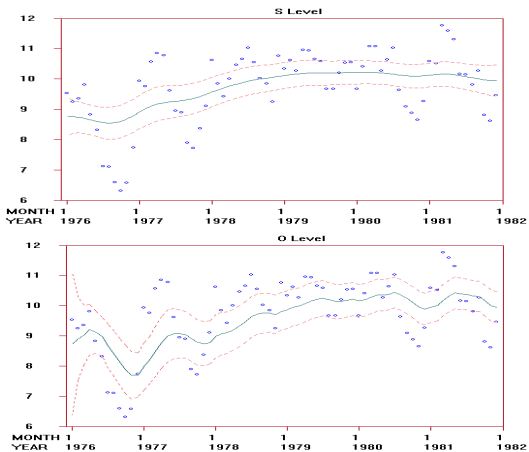


Figura: Estimativas Retrospectivas



BATS: Trend: Constant, Discount:Trend=0.90,Variance=0.99

Figura: Estimativas On-Line e Suavizada do Nível



Pode-se ver que com a filtragem tem-se trajetórias mais suaves e limites de incerteza mais próximos.

## Modelo com variância conhecida:

Para  $1 \leq k \leq t$

$$(\theta_{t-k}|D_t) \sim N[a_t(-k), R_t(-k)]$$

$$a_t(-k) = m_{t-k} - B_{t-k}[a_{t-k+1} - a_t(-k+1)]$$

$$R_t(-k) = C_{t-k} - B_{t-k}[R_{t-k+1} - R_t(-k+1)]B'_{t-k}$$

$$B_t = C_t G'_{t+1} R_{t+1}^{-1}$$

## Modelo com variância desconhecida:

Para  $1 \leq k \leq t$

$$(\theta_{t-k}|D_t) \sim t_{n_t(-k)}[a_t(-k), R_t(-k)]$$

$$(\phi_{t-k}|D_t) \sim G[n_t(-k)/2, d_t(-k)/2]$$

$$a_t(-k) = m_{t-k} - B_{t-k}[a_{t-k+1} - a_t(-k+1)]$$

$$R_t(-k) = C_{t-k} - B_{t-k}[R_{t-k+1} - R_t(-k+1)]B'_{t-k}$$

$$B_t = C_t G'_{t+1} R_{t+1}^{-1}$$

$$n_t(-k) = n_{t-k} + \delta_{t-k+1}(n_t(-k+1) - \delta_{t-k+1}n_{t-k})$$

$$S_t^{-1} = S_{t-k}^{-1} + \delta_{t-k+1}(S_t^{-1}(-k+1) - S_{t-k}^{-1})$$

$$d_t(-k) = n_t(-k)S_t(-k)$$

# Superposição de Modelos

---

Antes de apresentar o tratamento para variáveis causais e sazonalidade, é útil termos uma forma geral para estruturar e acomodar as várias *componentes intervenientes num modelo dinâmico*.

Muitas séries temporais exibem um comportamento bastante complexo. Ao identificarmos as características mais marcantes, estamos caminhando na direção de formular um modelo. A série de acidentes é um exemplo típico. (ver série)

- A tendência global parece ser de uma **variação suave do nível**.
- Se agora nos concentramos na variação em torno desse nível, podemos detectar um **comportamento cíclico**.

Essa inspeção permitiu identificar os *dois componentes* de um modelo: um componente para a **tendência** e outro para a **sazonalidade**.

- A estrutura dos modelos dinâmicos é apropriada, pois permite que as componentes sejam modeladas separadamente e depois integradas num modelo.
- No caso mais comum de duas componentes: tendência e sazonalidade, estruturamos a equação das observações com dois termos.

$$y_t = y_{Nt} + y_{St} + v_t$$

Cada um dos termos é descrito através de um modelo dinâmico

$$\begin{aligned} y_{Nt} &= F'_{Nt} \theta_{Nt} \\ \theta_{Nt} &= G_{Nt} \theta_{Nt-1} + w_{Nt} \end{aligned}$$

$$\begin{aligned} y_{St} &= F'_{St} \theta_{St} \\ \theta_{St} &= G_{St} \theta_{St-1} + w_{St} \end{aligned}$$

Se agora integramos esses termos, obtemos a **equação das observações**

$$y_t = \mathbf{F}'\boldsymbol{\theta}_t + v_t$$

$$\text{onde } \mathbf{F} = \begin{pmatrix} F_{Nt} \\ F_{St} \end{pmatrix} \text{ e } \boldsymbol{\theta} = \begin{pmatrix} \theta_{Nt} \\ \theta_{St} \end{pmatrix}$$

Similarmente, a **equação do sistema (integrado)** fica

$$\boldsymbol{\theta}_t = \mathbf{G}\boldsymbol{\theta}_{t-1} + \mathbf{w}_t; \quad \mathbf{w}_t \sim N(0, \mathbf{W}_t)$$

$$\text{onde } \mathbf{G}_t = \begin{pmatrix} G_{Nt} & 0 \\ 0 & G_{St} \end{pmatrix} \text{ e } \mathbf{W}_t = \begin{pmatrix} W_{Nt} & 0 \\ 0 & W_{St} \end{pmatrix}$$

- Modelos com mais componentes são construídos da mesma forma: cada termo contribui para a equação das observações e com um bloco de parâmetros para a equação do sistema.
- A especificação da variância do sistema através do método dos descontos segue o mesmo caminho, i.é, é feita componente a componente. Nesse caso, são agrupados conjuntos de parâmetros cujo comportamento é julgado similar em termos de variação temporal.
- Exemplo: no modelo de *vendas* explicadas pelo preço temos dois parâmetros,  $\mu_t$  e  $\beta_t$  (coeficiente de *preço*), que evoluirão segundo descontos  $\delta_N$  e  $\delta_P$  tais que

$$V^{-1}[\mu_t|D_{t-1}] = \delta_N V^{-1}[\mu_{t-1}|D_{t-1}]$$

$$V^{-1}[\beta_t|D_{t-1}] = \delta_S V^{-1}[\beta_{t-1}|D_{t-1}]$$



# Modelos com Variáveis Causais

---

- O modelo genérico usado para introduzir os modelos dinâmicos foi obtido a partir de uma generalização dos modelos de regressão.
- Se a série de *vendas*,  $(y_t)$  é explicada pela série de *preços*  $(x_t)$  temos:

$$y_t = \mu_t + \beta_t x_t + v_t$$

$$\mu_t = \mu_{t-1} + w_{1t}$$

$$\beta_t = \beta_{t-1} + w_{1t}$$

*Nada na estrutura acima impede que outras variáveis sejam incluídas como regressores.*

No caso de uma série  $y_t$  com variáveis explicativas  $x_{1t}, \dots, x_{pt}$ , temos

$$y_t = \beta_{0t} + \beta_{1t}x_{1t} + \dots + \beta_{pt}x_{pt} + v_t$$

$$\beta_{it} = \beta_{i,t-1} + w_{it}; \quad i = 0, 1, \dots, p$$

chamado de **modelo de regressão dinâmica**

A estrutura de modelo dinâmico é evidente com:

- $F'_t = (1, x_{1t}, \dots, x_{pt})$
- $G_t = I_{p+1}$ , a matriz identidade de ordem  $p + 1$  e
- $w'_t = (w_{01}, w_{1t}, \dots, w_{pt})$

## MLD de Regressão Múltipla

---

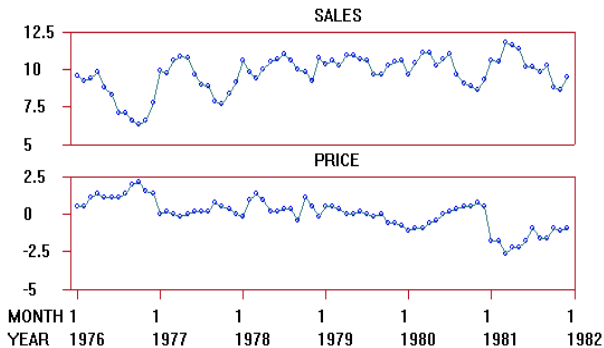
$$\begin{aligned} \text{Eq. observação:} \quad & y_t = \mathbf{F}_t' \boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim N[0, V_t] \\ \text{Eq. sistema:} \quad & \boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N[\mathbf{0}, \mathbf{W}_t] \end{aligned}$$

onde

- $\mathbf{F}_t = (X_{t1}, \dots, X_{tn})'$  : vetor de regressoras
- $X_1, \dots, X_n$  : variáveis independentes
- $X_{ti}$  : valor da  $i$ -ésima variável  $X$  no instante  $t$
- $\boldsymbol{\theta}_t$  :  $n \times 1$  vetor de parâmetros da regressão
- $\boldsymbol{\omega}_t$  : matriz da variância de  $\boldsymbol{\theta}_t$ .

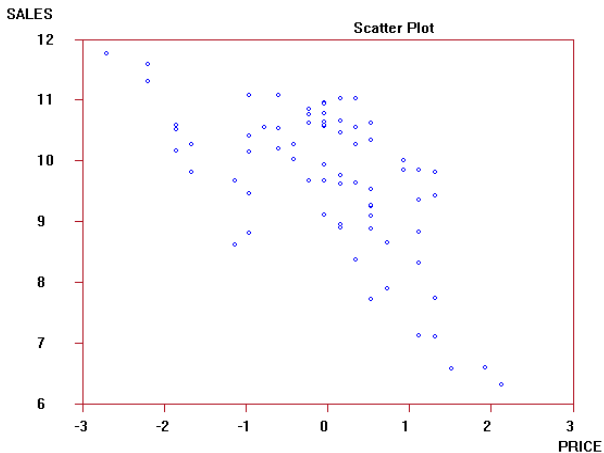
- Os dados deste exemplo correspondem as vendas e preços mensais de uma bala do arquivo CANDY.DAT do pacote BATS.
- Espera-se que série de vendas (SALES) esteja relacionada à série de preços (PRICE).
- Nas figuras a seguir apresentam-se alguns resultados do ajuste de um modelo de regressão dinâmica com tendência estável, utilizando o preço como variável explicativa.

Figura: MPlot de Vendas e Preços



Observa-se um aparente movimento comum das 2 séries.

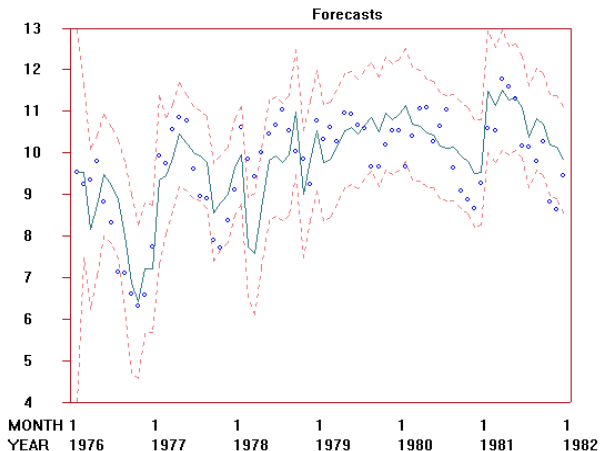
Figura: X-Y Plot de Vendas e Preços



O gráfico evidencia a relação existente entre as séries (Correlação=-0.63)

As previsões exibem melhoras consideráveis em comparação com os modelos sem preços.

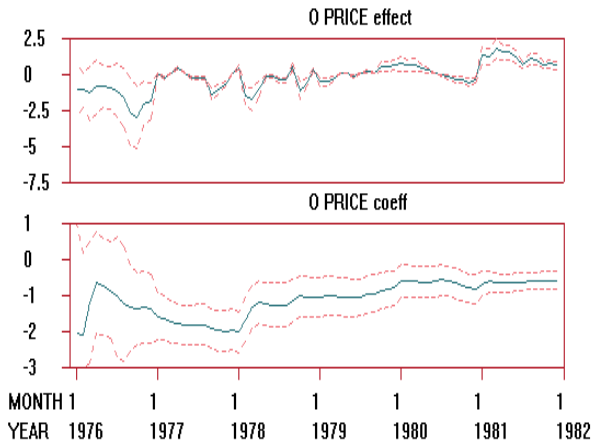
Figura: Previsão



BATS: Trend: Constant, Discount: Trend=0.90, Variance=0.99



Figura: Coeficiente e efeito dos preços



Este gráfico contém as trajetórias de  $E[\beta_t|D_t]$  e de  $E[\beta_t|D_t] \times x_t$ .

Figura: Estimativas Retrospectivas para vendas

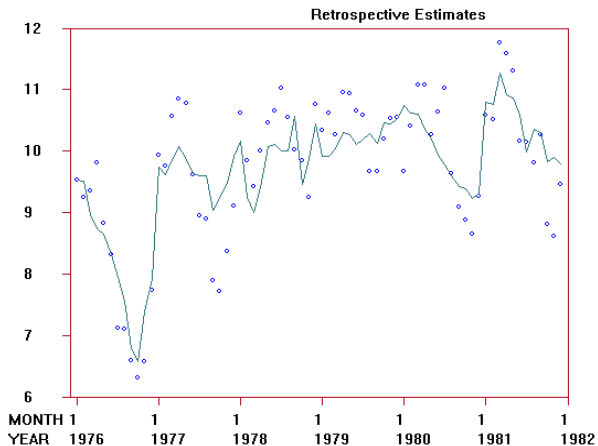
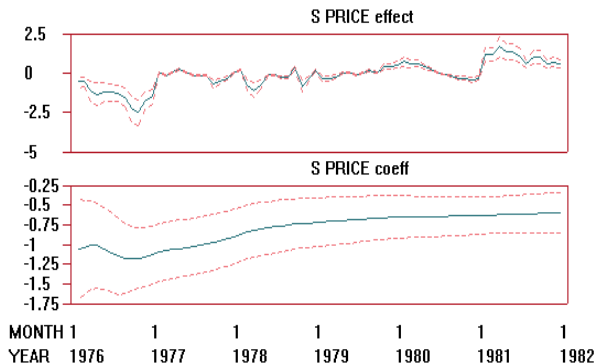


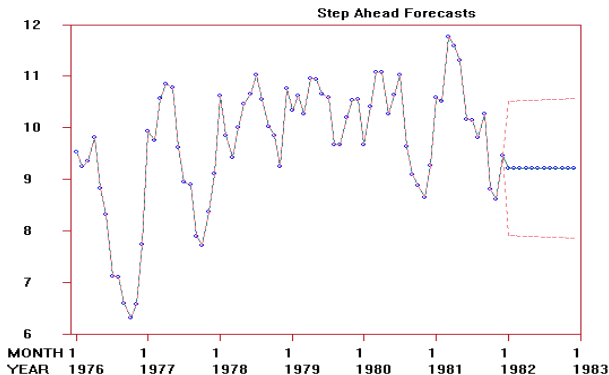
Figura: Coeficiente e efeito de preço suavizados



Este gráfico contém as trajetórias de  $E[\beta_t|D_t]$  e de  $E[\beta_t|D_t] \times x_t$  após a filtragem. Pode-se ver que o coeficiente da regressão varia de  $-1$  a  $-0.6$ . Esse movimento é permitido pelo MLD através do  $\delta$ .

- Para fazer previsão em modelos com regressores, é *necessário ter o valores dos regressores ao longo do horizonte de previsão.*
- Normalmente, esses valores também são incertos e o tratamento a ser dado é muito mais complicado.
- Uma alternativa intermediária é fazer previsão sob vários cenários plausíveis.

Figura: CANDY.DAT: Previsão para o período 1982/1 - 1982/12



Neste caso todos os valores de preço para o período de predição são zero, portanto, as previsões têm a forma constante do modelo estável.

# Modelos com Sazonalidade

---

# Modelagem de Forma Livre

- Modelos sazonais *requerem uma componente periódica* no modelo.
- A representação mais simples é através de **fatores ou indicadores de cada período no ciclo**. Para dados trimestrais, são usados quatro indicadores.
- Uma pequena alteração envolve o uso de efeitos indicando a variação sazonal em torno de um nível. Nesse caso, os efeitos estão restritos a ter soma zero.
- Fatores trimestrais de 100, 140, 80 e 120 equivalem a um nível de 110 e efeitos trimestrais de - 10, 30, - 30 e 10.
- A última formulação é mais atraente pois permite a separação entre sazonalidade e tendência.
- A restrição deve ser mantida em todas as afirmações probabilísticas mas é facilmente incorporável ao método de inferência utilizado.

Para dados trimestrais, os modelos dinâmicos utilizam quatro indicadores. A passagem do tempo faz com que eles experimentem uma rotação. Assim,

$$\theta_{t-1} = \begin{pmatrix} \text{trim4} \\ \text{trim1} \\ \text{trim2} \\ \text{trim3} \end{pmatrix}$$

Essa rotação pode ser efetuada pela matriz de evolução

$$G = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

O modelo é completado por uma equação de observações que considera apenas a primeira componente do vetor paramétrico, ou seja,  $F'_t = (1, 0, 0, 0)$ . A extensão para um ciclo de  $p$  períodos é análoga.



# Modelagem por harmônicos

- Uma outra modelagem de padrões cíclicos pode ser feita usando **funções trigonométricas**.
- A função  $\cos(\omega(t - 1))$  é periódica com período  $\frac{2\pi}{\omega}$ . Se  $\omega = \pi/6$ , o período é 12 e o máximo ocorre para  $t = 1$ .
- Dados mensais com ciclo anual podem ser concisamente modelados via

$$y_t = a_t \cos\left(\frac{\pi(t - 1)}{6}\right) + v_t$$

onde  $a_t$  é um parâmetro que controla a amplitude e o máximo ocorre em *janeiro*. Observe a redução drástica na dimensão do vetor paramétrico de 11 para 1.

- Defasagens no ponto de máximo do ciclo podem ser acomodadas com um parâmetro extra segundo

$$y_t = a_t \cos \left( \frac{\pi(t-1)}{6} \right) + b_t \sin \left( \frac{\pi(t-1)}{6} \right) + v_t$$

A formulação dinâmica dessa função harmônica utiliza 2 parâmetros,

$$F_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, G_t = \begin{pmatrix} \cos \omega & \sin \omega \\ -\sin \omega & \cos \omega \end{pmatrix}$$

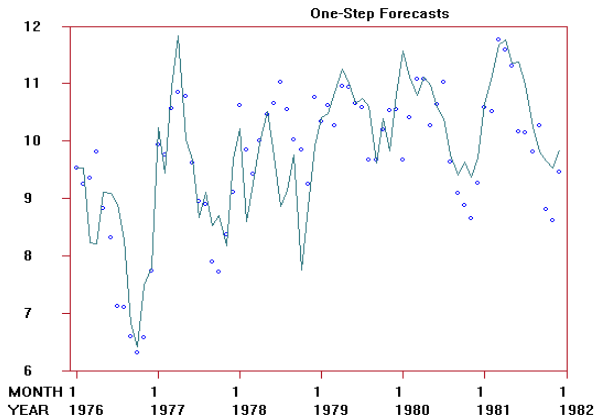
onde  $\omega = \pi/6$ .

Esse modelo descreve um ciclo segundo uma função *coseno*.

- Padrões cíclicos mais complicados podem ser modelados com a inclusão de formas harmônicas de frequência maior. A função  $\cos(2\omega(t - 1))$  é similar porém completa 2 ciclos durante um período de tempo  $p = 2\pi/\omega$  ou  $\omega = 2\pi/p$ .
- O resultado fundamental aqui informa que qualquer padrão cíclico de período  $p$  pode ser reproduzido com a soma de, no máximo,  $p/2$  harmônicos de períodos  $p/j, j = 1, \dots, [p/2]$  harmônicos.
- A vantagem desse resultado reside em podemos fazer **economia no número de parâmetros utilizados** e conseqüentemente aumentar nossa capacidade de aprendizado sobre o sistema e melhorar nossas previsões.

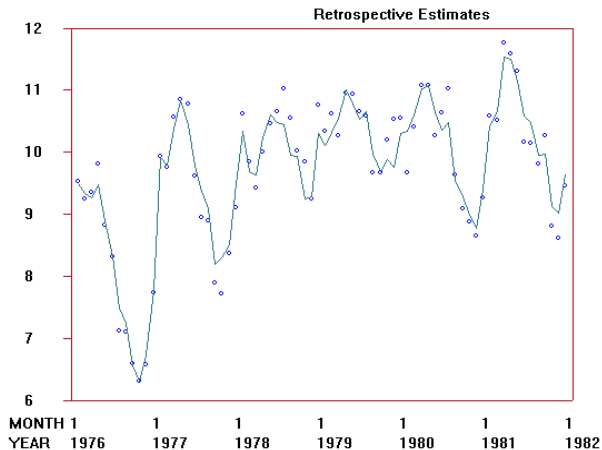
- A série de vendas do arquivo CANDY.DAT do BATS exibe um comportamento cíclico que até agora não havia sido tratado.  
(ver série)
- Nas figuras a seguir apresentam-se os resultados da análise da série CANDY.DAT considerando um modelo de tendência constante e um regressor (preço), incluindo a componente sazonal representada de forma livre e com harmônicos.

Figura: Previsão um passo a frente



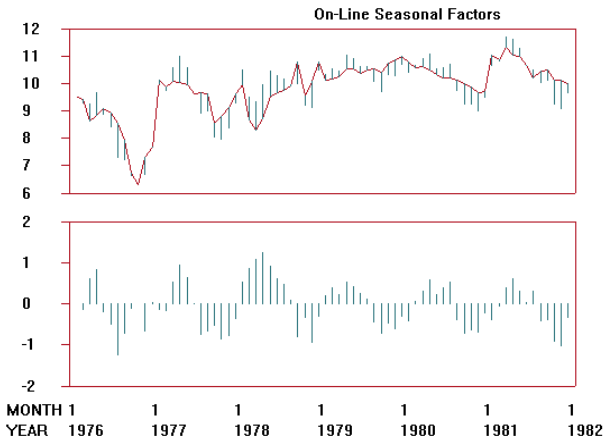
BATS: Trend: Constant, Regression: PRICE, Seasonal: FREE-FORM

## Figura: Estimativas Retrospectivas



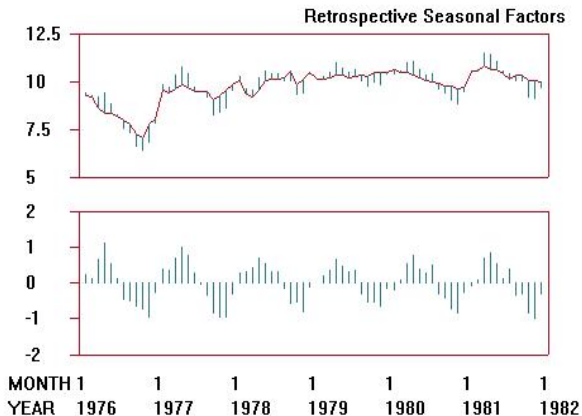
BATS: Trend: Constant, Regression: PRICE, Seasonal: FREE-FORM

Figura: Estimativa On-Line da Sazonalidade



BATS: Trend: Constant, Regression: PRICE, Seasonal: FREE-FORM

Figura: Estimativa Suavizada da Sazonalidade

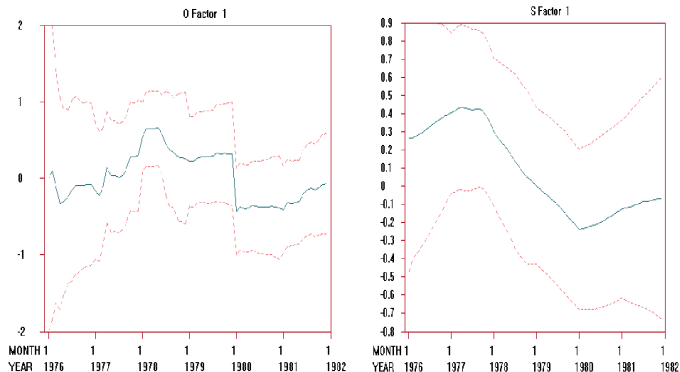


BATS: Trend: Constant, Regression: PRICE, Seasonal: FREE-FORM



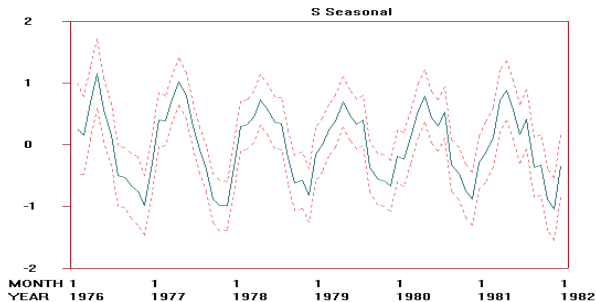
Este gráfico mostra claramente variação do efeito de janeiro ao longo do tempo ressaltando a importância da modelagem dinâmica.

Figura: Estimativas On-line e suavizadas do efeito de Janeiro



BATS: Trend: Constant, Regression: PRICE, Seasonal: FREE-FORM

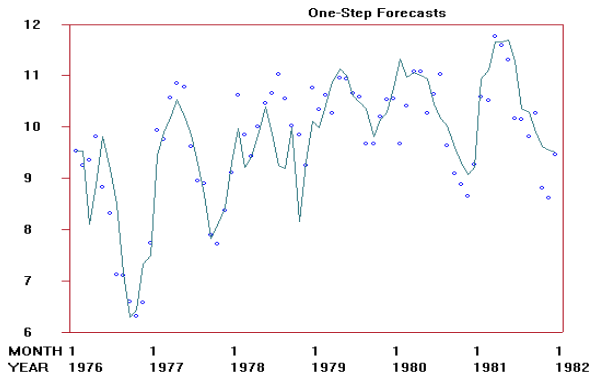
Figura: Estimativa da sazonalidade com limites de incerteza: forma livre



Esta figura mostra um padrão sazonal bastante similar ao de uma função *seno* embora variando no tempo, daí sugere que a flexibilidade fornecida pela modelagem em forma livre pode ter sido desnecessária. Pode-se contemplar a possibilidade de redução do tamanho do modelo através do uso de formas harmônicas para a componente sazonal.

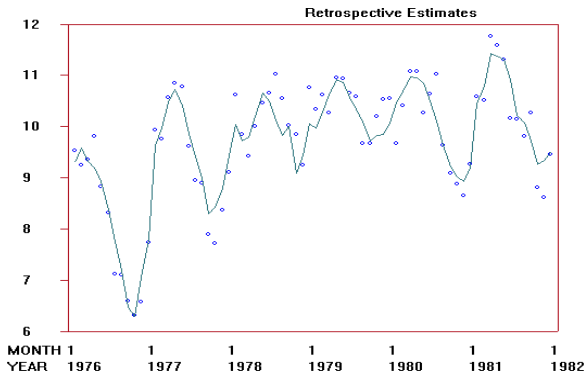
- Como os dados são mensais, temos que o período tem tamanho  $p = 12$ , havendo portanto até 6 harmônicos de períodos. O primeiro ou fundamental, de período  $12 = 12/1$ ; o segundo, de período  $6 = 12/2 \dots$  até o último, de período  $2 = 12/6$ .
- A diminuição da dimensão do vetor paramétrico é importante pois embora não altere as previsões pontuais, diminui a incerteza e facilita a monitoração da performance do modelo. Além disso, uma modelagem mais parcimoniosa acelera o tempo de processamento da análise.
- Esta modelagem da sazonalidade permite que o período do ciclo sazonal seja diferente do período natural da série e que o modelo escolhido contenha apenas o harmônico fundamental, ou seja, uma única função senoidal.

Figura: Previsão um passo a frente



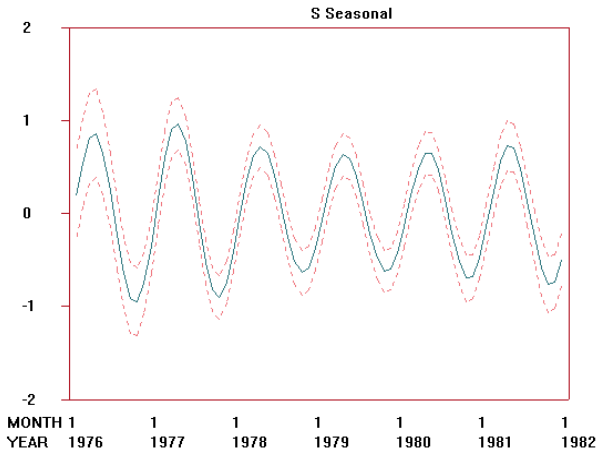
BATS: Trend: Constant, Regression: PRICE, Seasonal: RESTRICTED-HARMONICS

Figura: Estimativas Retrospectivas



BATS: Trend: Constant, Regression: PRICE, Seasonal: RESTRICTED-HARMONICS

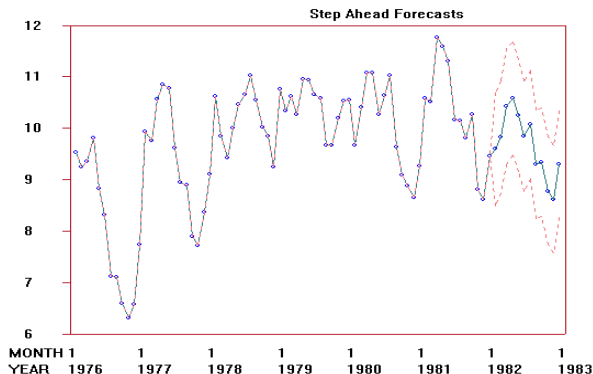
Figura: Estimativa da sazonalidade com limites de incerteza: 1<sup>o</sup> harmônico



As estimativas são bastantes parecidas com a modelagem de forma livre.

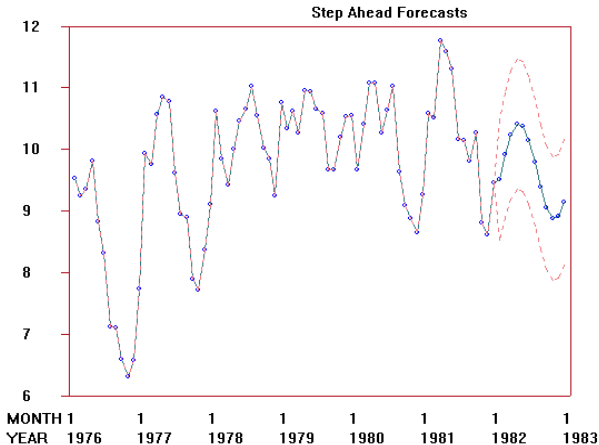
BATS: Trend: Constant, Regression: PRICE, Seasonal: RESTRICTED-HARMONICS

Figura: Previsão com limites de incerteza: Forma livre



O horizonte de previsão é 1982/1 a 1982/12. Os regressores estão zerados.

Figura: Previsão com limites de incerteza: 1<sup>o</sup> harmônico



Da comparação com a figura anterior, conclui-se que as previsões são similares mas têm maior incerteza na modelagem em forma livre.



- Até agora, foram realizadas análises utilizando as distribuições a priori de referência do BATS. Nada impede usar prioris informativas.
- Para a componente sazonal, só é necessário a especificação de incertezas a respeito dos fatores sazonais.
- No caso de modelagem por forma livre, BATS se encarrega de garantir que a restrição de soma zero seja respeitada. No caso de modelagem via harmônicos, BATS trata de ajustar essa incerteza especificada pelo usuário da melhor forma possível mesmo que o modelo tenha sido especificado apenas com alguns dos harmônicos.

# Monitoração

---

- Após fazer a previsão e observar o valor correspondente, podemos avaliar a precisão de um modelo.
- A idéia básica é compará-lo perante alternativas. Essa comparação pode servir para sinalizar acontecimentos inesperados
- Como a previsão é baseada numa distribuição de probabilidade, quanto *mais na cauda* cair a observação, *mais extrema e inesperada* (para o modelo) é a observação. Isso pode ocorrer devido a uma série de motivos:
  - uma mudança passageira e ocasional na estrutura dos dados;
  - uma mudança persistente e estrutural;
  - uma deterioração da performance do modelo.

O importante é que o sistema tenha capacidade de “soar o alarme”

Um esquema de monitoração funciona assim:

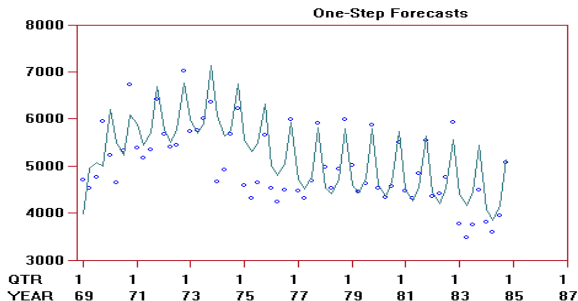
- quando a observação cair muito na cauda, “soa o alarme”.
- Nesse momento, o preditor terá de refletir sobre a adequação de seu modelo e, se for o caso, modificá-lo.
- Para auxiliá-lo, é útil termos uma coleção de alternativas e possíveis direções de ação.

# Aplicação: Acidentes rodoviários graves

- Os dados deste exemplo correspondem ao número de acidentes rodoviários graves (NUMBER) do arquivo QACCIDS.DAT do pacote BATS. *(ver série)*
- Na série de acidentes rodoviários pode-se notar 3 intervalos de tempo distintos dentro dos quais o comportamento da série é estável mas o padrão sazonal permanece estável ao longo da série. Podemos analisar os 3 intervalos separadamente mas estaríamos assim perdendo informação, por exemplo, sobre a componente sazonal.
- Nas figuras a seguir apresentam-se alguns resultados do ajuste de um modelo de tendência linear com componente sazonal de forma livre utilizando a opção de monitoração do BATS.

Ajustando um modelo com 2 componentes: **uma tendência linear e uma forma livre para a sazonalidade** e especificando **descontos constantes de 0.98 para tendência e sazonalidade** tem-se:

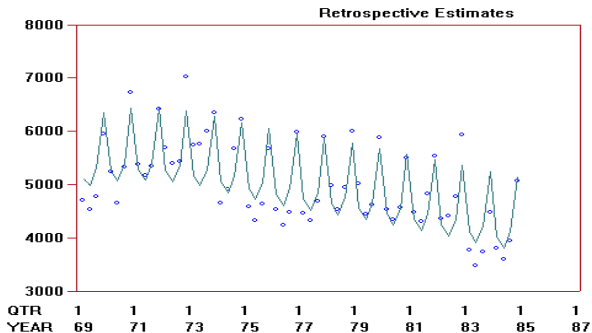
Figura: Previsões sem monitoração



Na figura pode se ver que **o modelo aprende rápido sobre o padrão sazonal mas se comporta muito mal no início dos intervalos onde ocorre mudança**. O motivo da demora a se ajustar à mudança é a especificação de **descontos altos**.

Na figura temos o resultado da análise retrospectiva. Podemos ver um comportamento sazonal bastante estável sobre uma tendência linear que varia suavemente. A estrutura do modelo impossibilita mudanças bruscas em quaisquer de suas componentes.

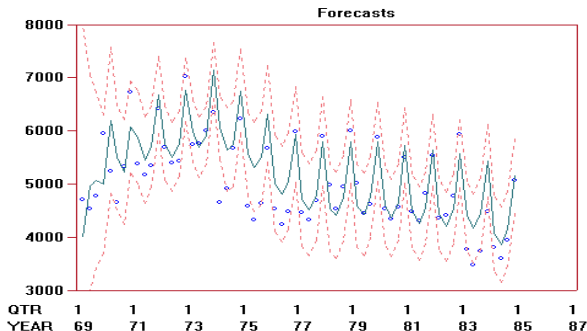
Figura: Estimativas suavizadas



BATS:Trend: Linear, Seasonal: Free-form, Discount:Constant, Fit:Reference.

Para melhorar a performance do modelo, **ligamos o monitor**. Pelo menos, esperamos que ele sinalize os tempos onde existe deterioração do modelo.

Figura: Previsões com monitoração



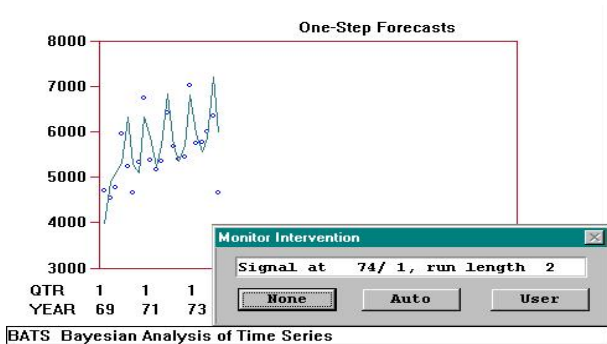
BATS:Trend: Linear, Seasonal: Free-form, Discount:Constant, Interrupt: Level Decrease  
Monitor, Fit:Reference.



- A figura anterior mostra que, mesmo com o monitor ligado, a análise permaneceu a mesma!
- Os grosseiros erros de 1974 não foram suficientemente grandes para fazer o monitor soar. Em 1974 o sistema ainda está incerto devido à inicialização de referência.
- Observe que os limites de 90% de incerteza não estão muito longe das observações de 1974 e que os limites aumentaram em seguida a elas.
- Para confirmar o ponto acima, vamos repetir a análise com menores incerteza a priori.

O monitor agora sinaliza no **1º trimestre de 1974**. A **redução na incerteza inicial** foi suficiente para tornar essa observação extrema o suficiente para **“soar o alarme”**.

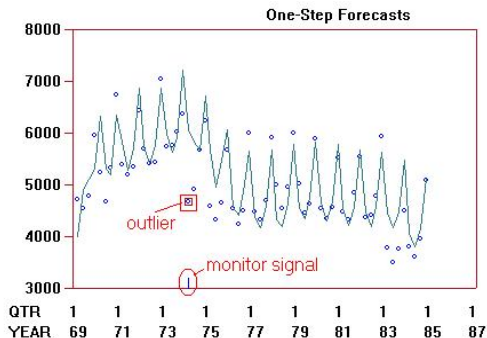
Figura: Previsões com monitoração



**Neste momento temos três opções:** Não fazer nada (já vimos não ser apropriado); deixar que o BATS atue automaticamente ou fazer algo nós mesmos.

Se deixamos ao BATS solucionar o problema temos que a análise continua tratando essa observação como aberrante e não a incorpora à análise. A adaptação ao novo nível é bem mais rápida.

Figura: Previsões com monitoração e menos incerteza a priori

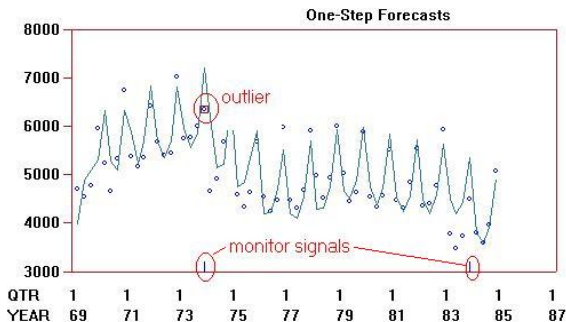


O **monitor do BATS** funciona baseado nas seguintes regras:

- Se a sinalização foi **devida apenas à observação mais recente**, ele a ignora e aumenta a incerteza a respeito dos parâmetros
- Se a sinalização foi **devida às últimas  $k$  observações**, há **indicação de mudança estrutural na série e apenas ocorre um aumento na incerteza a respeito dos parâmetros**.
- O aumento da incerteza é feito através da diminuição momentânea dos descontos para: **0.1 para tendência e componente sazonal, 0.8 para regressores e 0.9 para variância das observações**. Esses valores e a sensibilidade do monitor também *podem ser mudados*.

Mudando as especificações do monitor de queda no nível (desvio padrão para -2.5 e limiar do fator de Bayes para 0.3) temos:

Figura: Previsões com monitor mais sensível



Agora, o monitor sinaliza em 1973/4 e 1983/4, só que no último caso baseado numa série de 4 observações. O monitor não sinalizou em 1974/1 pois o modelo estava preparado para mudanças com incertezas aumentadas e nenhuma observação de 1983 foi dispensada.

# Intervenção

---

- Normalmente, um sistema de previsão funciona sem mudanças ao longo de sua análise. Existem momentos, entretanto, que é preciso fazer mudanças no seu funcionamento.
- Anteriormente, havíamos discutido a possibilidade do conhecimento de eventos excepcionais intervirem na série estudada, usando como exemplo a falência de um competidor. Embora esse tipo de informação não seja parte da série histórica, é fundamental para o sucesso do modelo, que seja incorporado à análise.
- Na notação utilizada, **ao invés de evoluirmos de  $p(\theta_{t-1}|D_{t-1})$  para  $p(\theta_t|D_{t-1})$**  devemos fazê-lo para  $p(\theta_t|D_{t-1}, I_t)$  onde  $I_t$  consiste na informação relevante. Note que essa mudança é essencialmente subjetiva.

- Até agora, só interviemos no modelo após a sinalização do monitor. Obviamente, essa restrição é desnecessária. Na prática, muitas vezes temos informação que nos sugere possíveis pontos de mudança na série.
- No caso da série de acidentes, possuímos tais informações:
  - em **1974/1** a crise do petróleo forçou um aumento significativo no preço da gasolina e
  - em **1983/1** passou a ser obrigatório o uso de cinto de segurança nos carros.



# Aplicação: Acidentes rodoviários graves

Na série de acidentes (ver série) indicamos ao BATS que vamos a intervir em 1973/4 e 1983/1. O programa para em esses dois pontos antes de incorporar as observações. Foram mudadas as prioridades do nível e do crescimento.

Figura: Previsões com intervenção antecipada em 1973/4 e 1983/1

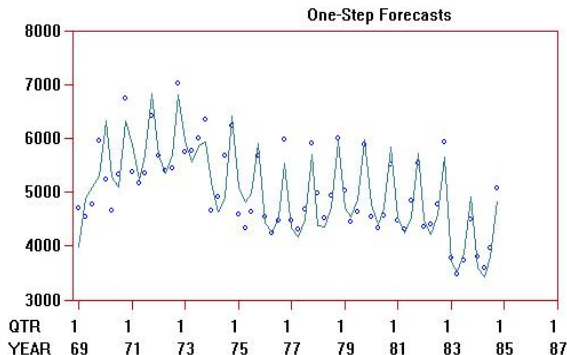


Figura: Análise Retrospectiva com intervenção antecipada

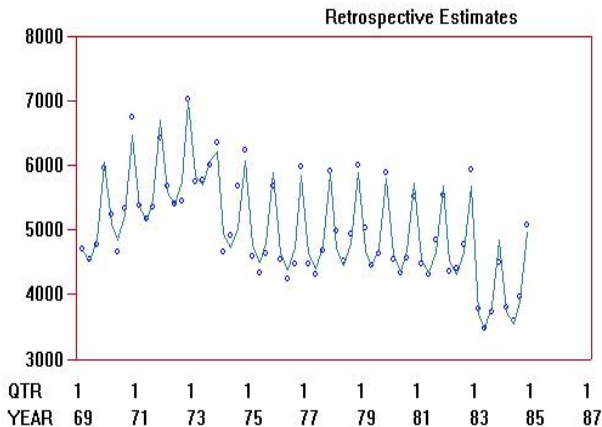


Figura: Nível Suavizado com intervenção antecipada

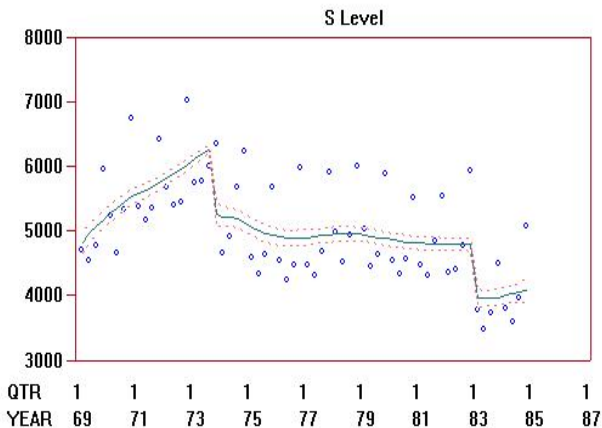


Figura: Crescimento Suavizado com intervenção antecipada

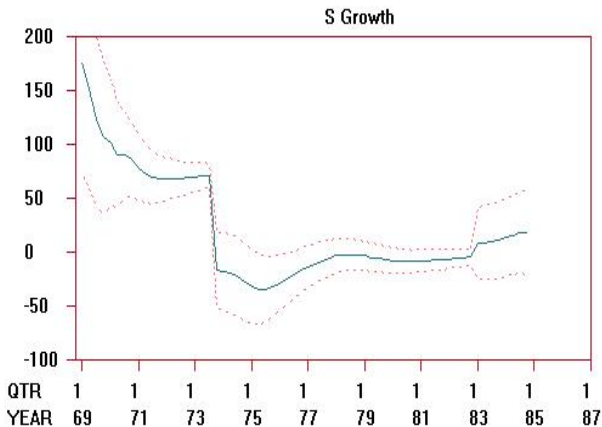
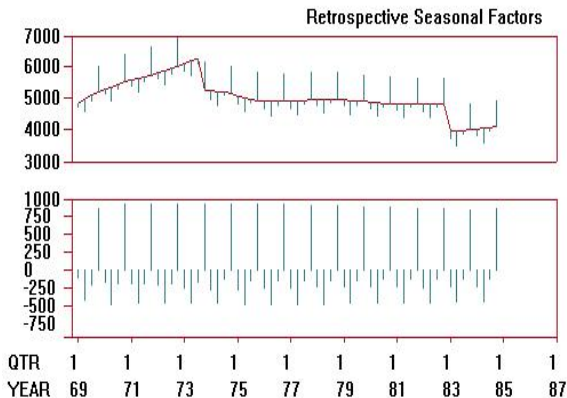


Figura: Padrão Sazonal da análise com intervenção antecipada

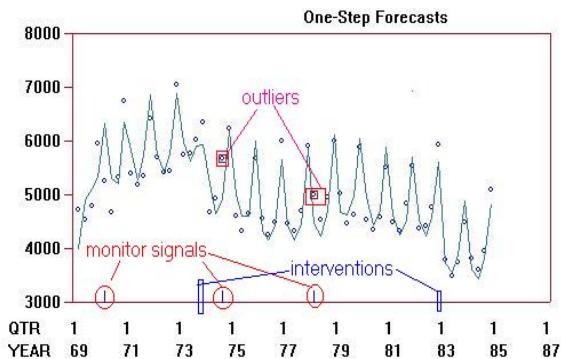


O comportamento sazonal é bastante estável ao longo de toda a série.

# Acidentes: Todas as intervenções

Normalmente **podemos monitorar e intervir numa mesma análise**

Figura: Previsões com todas as intervenções



Na figura anterior:

- O monitor sinaliza em 1970/1 mas a incerteza ainda é grande é melhor não intervir.
- A análise para no período pré-especificado 1973/4 onde fazem-se mudanças na priori.
- O monitor sinaliza em 1974/3 e 1978/1, em ambos permite-se intervenção automática.
- Finalmente a análise para em 1983/1 onde repetem-se as mudanças na priori.

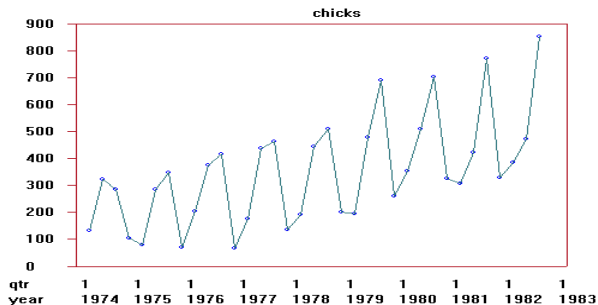
# Análise de Dados: transformações

---



O gráfico mostra as vendas trimestrais de filhotes de peru na Irlanda.

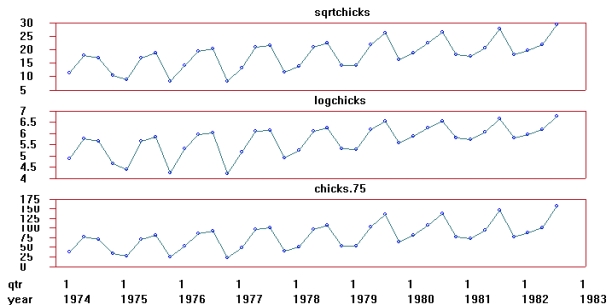
Figura: Vendas trimestrais de filhotes de peru de 1 dia na Irlanda



Além de uma **tendência de crescimento**, a série exibe **amplitude sazonal também crescente**.

Para controlar essa **não-linearidade**, procedem-se as transformações nos dados mostradas na figura:

**Figura:** Diferentes transformações na série de filhotes de peru



A potência  $3/4$  parece dar os melhores resultados.

Outro aspecto interessante da série é a mudança no padrão sazonal de baixo, alto, alto, baixo para baixo, médio, alto, baixo.

As séries transformadas foram analisadas com **modelo de tendência linear e componente sazonal de forma livre e priori informativa**. Os critérios de avaliação da performance foram:

- MSE: erro médio quadrático de previsão 1 passo a frente.
- MAD: erro médio absoluto de previsão 1 passo a frente.
- Logaritmo da verossimilhança do modelo.

A verossimilhança do modelo é a densidade dos dados condicionada ao modelo. Se o modelo  $M_1$  tem densidade mais alta que  $M_2$ , ele tem mais chances de ter gerado os dados e deve ser preferido. Essa regra pode ser formalizada probabilisticamente através da inferência bayesiana: se os dois modelos tem a mesma probabilidade a priori  $M_1$  tem verossimilhança maior que  $M_2$  então pelo Teorema de Bayes segue que a probabilidade a posteriori de  $M_1$  é maior que a de  $M_2$

- Tanto na transformação  $\sqrt{y_t}$  quanto na  $y_t^{3/4}$ , os modelos dinâmicos sempre foram superiores aos modelos estáticos para várias combinações de descontos utilizadas.
- Na transformação  $y_t^{3/4}$  o modelo dinâmico com  $\delta_T = 0.9$  e  $\delta_S = 0.7$  e o modelo estático forneceram:

Critério	M. Dinâmico	M. Estático
MSE	111.0	153.6
MAD	7.9	9.6
log(L)	-134.7	-144.2

- Essa comparação pode ser feita também com as previsões mostradas nas próximas figuras.

Figura: Previsões para a transformação  $y^{3/4}$ : Modelo Estático

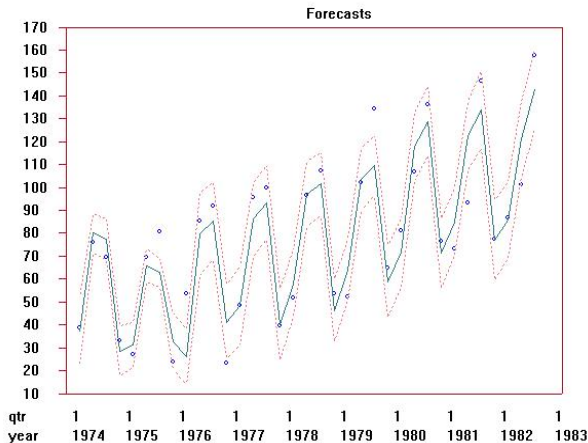
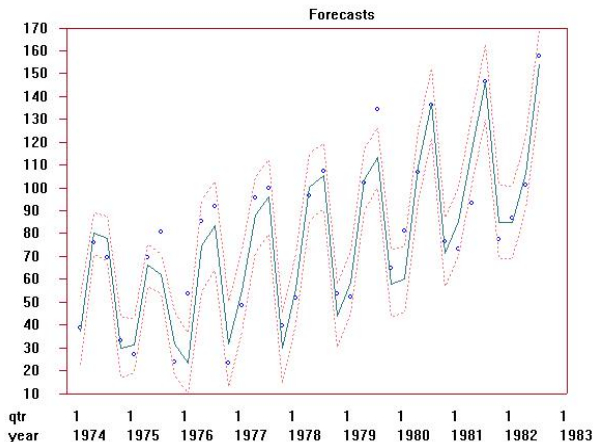


Figura: Previsões para a transformação  $y^{3/4}$ : Modelo Dinâmico



Para **comparar séries transformadas**  $z_t = y_t^\lambda$ , a verossimilhança deve ser ajustada pelo fator  $\lambda^n \prod_t y_t^{\lambda-1}$ . Com esse ajuste, as log-verossimilhanças dos modelos dinâmicos ficam:

Transformação $\sqrt{y_t}$ , $\lambda = 1/2$		
$\delta$	Log-L ( $\ell$ )	Log-L ajustado
(0.9,0.8)	-71.1	-194.8
(1.0,1.0)	-77.6	-201.3
Transformação $y_t^{3/4}$ , $\lambda = 3/4$		
$\delta$	Log-L ( $\ell$ )	Log-L ajustado
(0.9,0.7)	-134.7	-194.5
(1.0,1.0)	-144.2	-204.0

Com este ajuste, as log-verossimilhanças estão indicando uma ligeira preferência pela transformação  $y_t^{3/4}$

As previsões em séries transformadas devem ser transformadas à escala original para melhor comunicação. Uma previsão com média 87 e limites de 90% de incerteza  $(87 - 15)$  e  $(87 + 15)$  na escala  $y_t^{3/4}$  corresponde a previsões com média  $87^{4/3} = 382$  e limites de 90% de incerteza  $(87 - 15)^{4/3} = 299$  e  $(87 + 15)^{4/3} = 469$

Transformação $\sqrt{y_t}$ , $\lambda = 1/2$		
Período	Moda Prevista	Intervalo 90%
1982 IV	380	(278,498)
1983 I	422	(317,507)

Transformação $y_t^{3/4}$ , $\lambda = 3/4$		
Período	Moda Prevista	Intervalo 90%
1982 IV	382	(299,469)
1983 I	424	(317,507)



## Parte III

# Extensões

# Modelos Lineares Dinâmicos Generalizados

---

- Suponha agora que as observações são não normais
- A distribuição das observações é membro da família exponencial, onde  $\eta_t$  é o *parâmetro natural* e  $\phi_t = V_t^{-1}$  é a *precisão* da distribuição.
- $\eta_t$  tem uma relação linear com o vetor de estados  $\theta_t$
- A equação de evolução para o vetor de estados é igual ao modelo normal
- A distribuição normal faz parte dessa família.

O modelo é especificado por:

$$\text{Equação de Observação : } p(y_t | \eta_t) = \exp\left\{ \phi_t [y_t \eta_t - a(\eta_t)] \right\} b(y_t, V_t)$$

$$\text{Função da Média : } \mu_t = E(y_t | \eta_t) = a'(\eta_t)$$

$$\text{Função de Ligação : } g(\mu_t) = \mathbf{F}'_t \boldsymbol{\theta}_t$$

$$\text{Equação do sistema : } \boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$$

$$\text{Erro do sistema : } \boldsymbol{\omega}_t \sim (\mathbf{0}, \mathbf{W}_t)$$

$$\text{Informação a priori : } \boldsymbol{\theta}_1 \sim (\mathbf{a}_1, \mathbf{R}_1)$$

★ No caso da normal,  $\eta_t = \mu_t$

# Exemplo 01:

## Modelo de Regressão Log-linear Dinâmico

Se  $y_t \sim \text{Poisson}(\lambda_t)$ , então:

$$p(y_t | \lambda_t) \propto \exp\{y_t \log(\lambda_t) - \lambda_t\}$$

$$\phi = 1$$

$$\eta_t = \log \lambda_t$$

$$a(\eta_t) = \exp(\eta_t) = \lambda_t$$

$$\eta_t = \theta_{1t} + \theta_{2t}x_t = (1, x_t)\boldsymbol{\theta}_t$$

$$\mathbf{F}'_t = (1, x_t)$$

$$\boldsymbol{\theta}_t = (\theta_{1t}, \theta_{2t})' = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t$$

$$\boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}), \quad \boldsymbol{\omega}_1 \sim N(\mathbf{0}, \mathbf{R})$$

## Exemplo 02: Modelo de Regressão Logístico Dinâmico

Se  $y_t \sim \text{Binomial}(n_t, \pi_t)$ , então:

$$p(y_t | n_t, \pi_t) \propto \exp \left\{ y_t \text{logit}(\pi_t) + n_t \log(1 - \pi_t) \right\}$$

$$\phi = 1$$

$$\eta_t = \text{logit}(\pi_t)$$

$$a(\eta_t) = -n_t \log(1 - \pi_t)$$

$$\eta_t = \theta_{1t} + \theta_{2t}x_t = (1, x_t)\theta_t$$

$$\mathbf{F}'_t = (1, x_t)$$

$$\theta_t = (\theta_{1t}, \theta_{2t})' = \theta_{t-1} + \omega_t$$

$$\omega_t \sim N(\mathbf{0}, \mathbf{W}), \quad \omega_1 \sim N(\mathbf{0}, \mathbf{R})$$

# Inferência em MLDG: Linear Bayes

---

# Inferência em MLDG: Linear Bayes

Re-lembrando:  $p(y_t | \eta_t) \quad \eta_t = \mathbf{F}'_t \boldsymbol{\theta}_t;$   
 $\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim (\mathbf{0}, \mathbf{W}_t)$

- Seja  $(\boldsymbol{\theta}_{t-1} | D_{t-1}) \sim (\mathbf{m}_{t-1}, \mathbf{C}_{t-1})$ . Então  $(\boldsymbol{\theta}_t | D_{t-1}) \sim (\mathbf{a}_t, \mathbf{R}_t)$  onde  $\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}$  e  $\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}'_t + \mathbf{W}_t$
- **Priori para  $\eta_t$ :**  $(\eta_t | D_{t-1}) \sim (f_t, q_t)$ .  
Se a priori de  $\eta_t$  é conjugada, tem-se que priori e posteriori pertencem à mesma família, logo  $(\eta_t | D_t) \sim (f_t^*, q_t^*)$ .
- **Estrutura Condicional de  $(\boldsymbol{\theta}_t | \eta_t, D_{t-1})$ :**

$$\left( \begin{pmatrix} \eta_t \\ \boldsymbol{\theta}_t \end{pmatrix} \middle| D_{t-1} \right) \sim \left( \begin{pmatrix} f_t \\ \mathbf{a}_t \end{pmatrix}, \begin{pmatrix} q_t & \mathbf{F}'_t \mathbf{R}_t \\ \mathbf{R}_t \mathbf{F}_t & \mathbf{R}_t \end{pmatrix} \right).$$

Linear Bayes  $\Rightarrow \quad \widehat{E}(\boldsymbol{\theta}_t | \eta_t, D_{t-1}) = \mathbf{a}_t + \mathbf{R}_t \mathbf{F}_t (\eta_t - f_t) / q_t$   
 $\widehat{Var}(\boldsymbol{\theta}_t | \eta_t, D_{t-1}) = \mathbf{R}_t - \mathbf{R}_t \mathbf{F}_t \mathbf{F}'_t \mathbf{R}_t / q_t$



- Atualização para  $\theta_t$ .

$$p(\theta_t | D_t) = \int \underbrace{p(\theta_t | \eta_t, D_{t-1})}_{\text{não é conhecida}} \underbrace{p(\eta_t | D_t)}_{\text{conjugação}} d\eta_t$$

Temos então que:

$$\theta_t \sim (m_t, C_t)$$

$$\begin{aligned} m_t &= E[\theta_t | D_t] \\ &= E[\widehat{E}\{\theta_t | \eta_t, D_{t-1}\} | D_t] \\ &= \mathbf{a}_t + \mathbf{R}_t \mathbf{F}_t (f_t^* - f_t) / q_t \end{aligned}$$

$$\begin{aligned} C_t &= V[\theta_t | D_t] \\ &= V[\widehat{E}\{\theta_t | \eta_t, D_{t-1}\} | D_t] + E[\widehat{V}\{\theta_t | \eta_t, D_{t-1}\} | D_t] \\ &= \mathbf{R}_t - \mathbf{R}_t \mathbf{F}_t \mathbf{F}_t' \mathbf{R}_t (1 - q_t^* / q_t) / q_t \end{aligned}$$

# Aplicação: MLDG Poisson-Gama de 2<sup>o</sup> ordem

- **Objetivo:** Análise temporal do número mensal de crimes violentos, ocorridos entre janeiro de 1998 e agosto de 2001 numa região de Belo Horizonte.
- **Modelo:**

$$(y_t | \lambda_t) \sim \text{Poisson}(\lambda_t), t = 1, \dots, 45; \quad (1a)$$

$$\log \lambda_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \mu_t \quad (1b)$$

$$\begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix} \quad (1c)$$

$$\mu_0 \sim N(m_{10}, C_{10})$$

$$\beta_0 \sim N(m_{20}, C_{20})$$

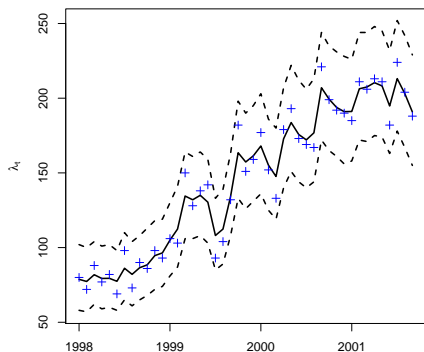
- $w_t \sim (0, W) \rightarrow W$  especificado via descontos ( $\delta = 0.98$ )

```

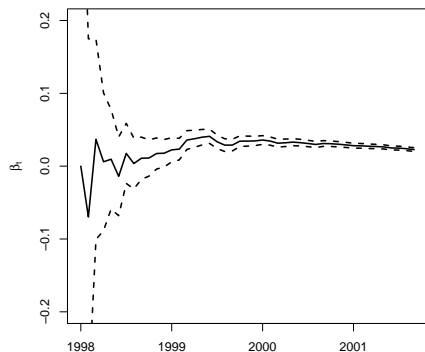
# include <oxstd.h> main() {
// Dados e Variáveis
    decl data = loadmat("cia6.dat"); decl nobs = (rows(data)+1);
    decl Ft = <1; 0>; decl Gt = <1, 1; 0, 1>; decl delta = 0.99;
    ...
// Informação inicial
    mt[0] = <4.30; 0.00>; Ct[0] = <0.10, 0.00; 0.00, 0.01>;
// Linear Bayes
    for(i=1; i<(nobs); i++){
        at[i] = Gt*mt[i-1];
        Rt[i] = (Gt*Ct[i-1]*Gt')*(1/delta);
        ft[i] = Ft'*at[i];
        qt[i] = Ft'*Rt[i]*Ft;
        rt[i] = 1/qt[i];
        st[i] = 1/(qt[i]*exp(ft[i]));
        rs[i] = rt[i] + y[i];
        ss[i] = st[i] + 1;
        fs[i] = log(rs[i]) - log(ss[i]);
        qs[i] = 1/(rs[i]);
        mt[i] = at[i] + Rt[i]*Ft*(fs[i]-ft[i])*(1/qt[i]);
        Ct[i] = Rt[i] - (Rt[i]*Ft*Ft'*Rt[i])*(1-(qs[i]/qt[i]))*(1/qt[i]);
    }... }

```

# Aplicação: número de crimes



(a)  $\lambda_t|D_t$



(b)  $\beta_t|D_t$

Figura: IC 95% para o Nível e Parâmetro de crescimento de (1) estimado com Linear Bayes

# Inferência em MLDG: MCMC

---

- Métodos Monte Carlo (MC): métodos de inferência baseados em simulação.
- MC possibilita a implementação eficiente de métodos de simulação com modelos complexos como MLD e MLDG.
- Necessário se queremos estimar  $W$  ou elementos desconhecidos de  $F$  e  $G$ .
- Podemos usar MC para gerar valores da distribuição a posteriori dos parâmetros dos MLD e MLDG.
- Em MLD e MLDG não se consegue gerar diretamente da posteriori
- Solução: MCMC.

# Monte Carlo via cadeia de Markov (MCMC)

- Um algoritmo Monte Carlo via cadeia de Markov (MCMC) para simular de uma distribuição  $\pi$  é qualquer método que produza uma cadeia de Markov homogênea, ergódica e irredutível cuja distribuição estacionária seja  $\pi$ . (Uma cadeia é ergódica se ela é aperiódica e recorrente positiva)
- MCMC: amostras dependentes são geradas de uma cadeia de Markov cuja distribuição de equilíbrio é a distribuição de interesse. Aviso: (1) amostras do MCMC geralmente são muito correlacionadas; (2) estimativas de amostras correlacionadas tendem a ter variâncias maiores do que amostras independentes.
- Algumas questões: qual o tamanho do *burn-in*? as cadeias estão passeando por todo o espaço paramétrico? quantas iterações?
- Exemplos de MCMC:
  - Amostrador de Gibbs
  - Metropolis Hasting

## ● Amostrador de Gibbs

Para obter uma amostra de  $p(\theta_1, \theta_2)$  passeamos aleatoriamente pelo espaço paramétrico de acordo com a seguinte regra de transição:

- Cada iteração produz um movimento numa direção somente, simulando de  $p(\theta_1|\theta_2)$  e  $p(\theta_2|\theta_1)$ .
- Depois do período de *burn-in* (quando se perde a influência do ponto inicial, atingindo a convergência) começamos a ter uma amostra de  $p(\theta_1, \theta_2)$

A idéia se generaliza para  $\theta$  com mais de 2 componentes gerando das condicionais completas  $p(\theta_i|\theta_{-i})$

## ● Metropolis Hasting

Técnica que produz uma cadeia com regra de transição em duas etapas:

- Movimentos são propostos de uma regra de transição arbitrária
  - Etapa de aceitação
- Para MLD usamos Gibbs. Para MDLG usamos Metropolis.



# Sobre as regras de transição propostas

- Principais diferenças:
  - Escolha das regras de transição propostas,
  - Forma de acelerar a convergência
- Atualização do vetor de estados:
  - $\theta_t$  a cada passo (*Single move*),
  - $\Theta_t$  a cada passo (*Multi move*),
  - $(\theta_r, \theta_s)$  a cada passo (*Block move*).

- **Objetivo:** Análise temporal do número mensal de crimes violentos, ocorridos entre janeiro de 1998 e agosto de 2001 numa região de Belo Horizonte.
- **Modelo:**

$$y_t | \lambda_t \sim \text{Poisson}(\lambda_t) \quad (2a)$$

$$\log \lambda_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \mu_t \quad (2b)$$

$$\begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}; \quad \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \sim \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix} \right]$$

$$\mu_0 \sim N(m_{10}, C_{10})$$

$$\beta_0 \sim N(m_{20}, C_{20})$$

$$W_1^{-1} \sim \text{Gama}(0.01, 0.01)$$

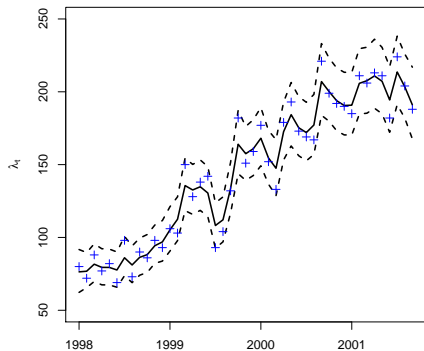
$$W_2^{-1} \sim \text{Gama}(0.01, 0.01)$$

```

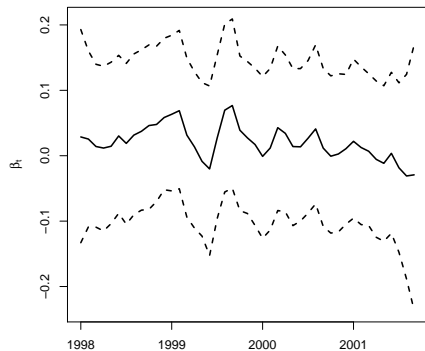
model "modelo dinâmico poisson de 2-ordem";
{
# Equações do modelo
  for(t in 2:46)
  {
    beta[t] ~ dnorm(beta[t-1],iw2);
    mean.mu[t] <- mu[t-1]+beta[t-1];
    mu[t] ~ dnorm(mean.mu[t],iw1);
    log(lambda[t]) <- mu[t]
    y[t] ~ dpois(lambda[t])
    y.rep[t] ~ dpois(lambda[t])I(0,10000)
  }
# Informação inicial
  mu[1] ~ dnorm(a[1], iR10)
  beta[1] ~ dnorm(a[2], iR20)
# Distribuição a priori
  iw1 ~ dgamma(0.01,0.01)
  iw2 ~ dgamma(0.01,0.01)
  w1 <- 1/iw1;    w2 <- 1/iw2;
# Hiperparâmetros
  a[1] <- 4.5;  a[2] <- 0.2;
  iR10 <- 2.0; iR20 <- 1;
}

```

# Aplicação: número de crimes



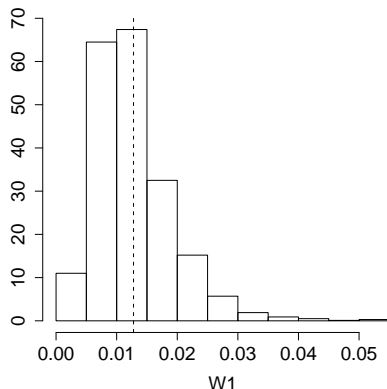
(a)  $\lambda_t | \mathbf{y}$



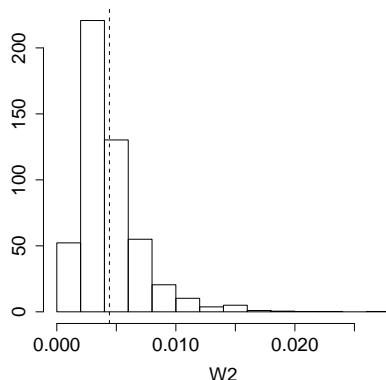
(b)  $\beta_t | \mathbf{y}$

Figura: IC 95% para o Nível e Parâmetro de crescimento de (2) estimado com MCMC

# Aplicação: número de crimes (cont.)



(a)  $W_1 | \mathbf{y}$



(b)  $W_2 | \mathbf{y}$

**Figura:** Amostras das distribuições a posteriori de  $W_1$  e  $W_2$  de (2) obtidas com MCMC (A linha pontilhada indica a média a posteriori)

# Modelos Não Lineares Dinâmicos Generalizados

---

# Modelos Não Lineares Dinâmicos Generalizados (MNLDG)

Seja  $y_t$  uma série temporal observada de  $t = 1$  a  $t = T$ .  
O modelo não linear dinâmico generalizado é dado por:

$$\text{Eq. de Observação :} \quad y_t = f_t(\theta_t) + v_t \quad (3a)$$

$$\text{Eq. do Sistema :} \quad \theta_t = g_t(\theta_{t-1}) + w_t \quad (3b)$$

onde  $f_t(\cdot)$  e  $g_t(\cdot)$  são funções conhecidas

Exemplo: [Modelo de Função de Transferência](#)

$$y_t = \mu + E_t + \nu_t, \quad \nu_t \sim N(0, V) \quad (4a)$$

$$E_t = \rho E_{t-1} + \gamma X_t \quad (4b)$$

# Aplicação: dados de chuva e vazão

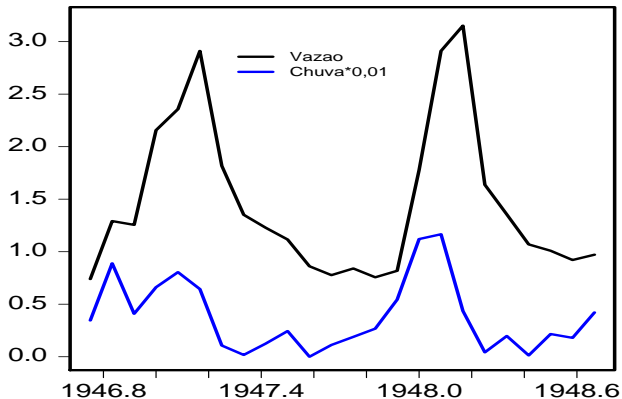
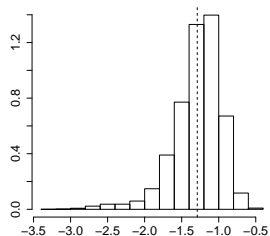


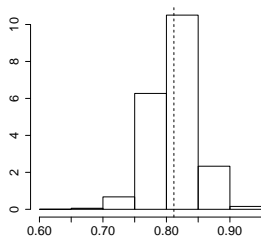
Figura: Vazão e Precipitação de Outubro de 1946 a Setembro de 1948 na Bacia do Riberao Pinheirinho - SP



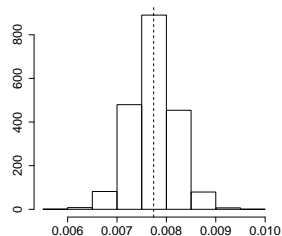
# Aplicação: dados de chuva e vazão (cont.)



(a)  $\mu|y$



(b)  $\rho|y$



(c)  $\gamma|y$

**Figura:** Amostras das distribuições a posteriori de  $\mu$ ,  $\rho$  e  $\gamma$  de (2) obtidas com MCMC (A linha pontilhada indica a média a posteriori)

# Aplicação: dados de chuva e vazão (cont.)

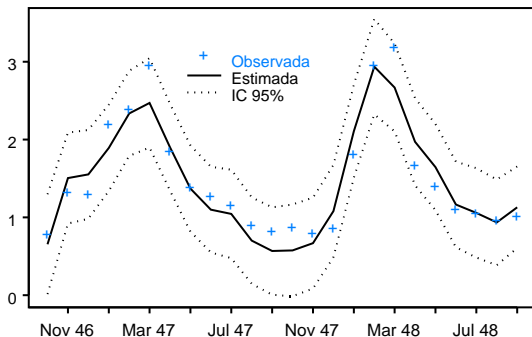


Figura: Vazão Observada e IC 95 % para a vazão estimada com o modelo (4)

# Modelos Dinâmicos Generalizados (MDG)

---

Seja  $y_t$  uma série temporal observada de  $t = 1$  a  $t = T$ .  
O modelo dinâmico generalizado é dado por:

$$\text{Eq. das Observações :} \quad y_t = f_t(\theta_t, \nu_t, V) \quad (5a)$$

$$\text{Eq. do Sistema :} \quad \theta_t = g_t(\theta_{t-1}, \omega_t, W) \quad (5b)$$

onde

- $f_t(\cdot)$  e  $g_t(\cdot)$  são funções conhecidas;
- $\nu_t$  e  $\omega_t$  são os erros observacional e do sistema não correlacionados e mutuamente independentes, com variâncias  $V$  e  $W$ , respectivamente.

# Inferência em MDG: MC seqüencial

---

# Monte Carlo seqüencial

- Algumas vezes as observações chegam de forma seqüencial no tempo e pode-se estar interessado na obtenção de inferências *on line*, sendo necessária a atualização das distribuições tão logo os dados se tornarem disponíveis.
- Quando a variável de interesse é modelada por um sistema dinâmico gaussiano e linear é possível acessar de forma analítica a seqüencia de posterioris. Entretanto, em diversas situações reais, normalidade e linearidade não se adequam aos dados. Nesses casos, pode-se usar MCMC mas a análise deixa de ser seqüencial.
- Métodos de Monte Carlo Seqüenciais aparecem como uma solução para o problema de determinação de distribuições a posteriori. Tais métodos proliferaram nos últimos anos e suas aplicações se tornaram cada vez mais acessíveis devido à melhora na capacidade computacional disponível. Dentre esses métodos estão os *filtros de partículas*.

- Os filtros de partículas são estratégias de simulação que aproximam a distribuição a posteriori de  $\theta_t, p(\theta_t|D_t)$ , por *partículas*  $\theta_t^{(1)}, \dots, \theta_t^{(M_t)}$  com respectivas probabilidades discretas  $w_t^{(1)}, \dots, w_t^{(M_t)}$ . Em outras palavras, o conjunto  $\{\theta_t^{(j)}, w_t^{(j)}\}_{j=1}^{M_t}$  aproxima  $p(\theta_t|D_t)$ .
- Todas as operações de evolução, previsão e atualização são realizadas por MC e aplicadas às partículas utilizando técnicas como *Sampling Importance Resampling* (SIR).
- Cuidados especiais devem ser tomados para dar conta das mudanças observadas para evitar degeneração das partículas.
- Alguns filtros incluem inferência para os hiperparâmetros.

# Muito obrigado!

migon@im.ufrj.br

dani@im.ufrj.br

[dme.ufrj.br/~hsmigon](http://dme.ufrj.br/~hsmigon)   [acd.ufrj.br/~dani](http://acd.ufrj.br/~dani)

romy@dme.ufrj.br

[dme.ufrj.br/~romy](http://dme.ufrj.br/~romy)