Modelos Dinâmicos: Teoria e Aplicações

Paloma Vaissman Uribe

I Encontro de Modelagem Estatística, Maringá, UEM.

05/12/2017

Modelos Lineares Dinâmicos Gaussianos: definição básica

- De forma geral, são modelos cujos **parâmetros variam no tempo**.
- São casos especiais dos **Modelos de Espaço de Estados** em que a especificação é linear e as perturbações são Gaussianas.

Modelos de Espaços de Estados: origem

- Originaram-se no início dos anos sessenta na área de engenharia de controle (Kalman et al., 1960).
- Fornecem uma estrutura geral para analisar sistemas dinâmicos determinísticos e estocásticos que são medidos ou observados através de um processo estocástico.
- Permitem uma interpretação natural de uma série temporal como sendo a combinação de vários componentes estruturais, como componentes de tendência, sazonalidade, ciclo, etc.
- Aparecem na literatura da série temporal nos anos setenta (Akaike, 1974; Harrison and Stevens, 1976) e se estabeleceu durante os anos oitenta (Harvey, 1989;West and Harrison,1997).

Modelos de Espaços de Estados: definição formal

Formalmente, um **modelo de espaço de estado** (MEE) consiste em um vetor de séries de tempo não observado $\mathbb{R}^q$-dimensional $\{\theta_t\}_{t\geq 0}$ (**os estados**) e séries de tempo $\mathbb{R}^m$-dimensionais $\{y_t\}_{t\geq 1}$ (**as observações**) que satisfazem os seguintes pressupostos:

(A.1) $\{\theta_t\}_{t\geq 0}$ é uma cadeia de Markov;

(A.2) Condicionalmente em $\{\theta_t\}_{t\geq 0}$, $\{y_t\}_{t\geq 1}$ é independente e y_t depende de θ_t apenas.

Dependência Markoviana

Dizemos que $\{y_t\}_{t\geq 1}$ é uma cadeia de Markov se, para qualquer $t > 1$,

$$p(y_t|y_{1:t-1}) = p(y_t|y_{t-1}),$$

i.e, **o processo de hoje só depende do processo no instante de tempo imediatamente anterior.**

Outra maneira de expressar a dependência Markoviana é dizer que y_t e $y_{1:t-2}$ são condicionalmente independentes, dado y_{t-1}. Assim, a distribuição conjunta de $(y_1, y_2, .., y_t)$ é

$$p(y_{1:t}) = p(y_1) \prod_{t>1} p(y_t|y_{t-1}).$$

Distribuição conjunta dos estados e observações

A partir da definição, concluímos que um MEE é completamente especificado pela distribuição inicial de $p(\theta_0)$ e as densidades condicionais $p(\theta_t|\theta_{t-1})$ e $p(y_t|\theta_t)$.

Assim, para $t > 0$, a distribuição conjunta é a seguinte

$$p(\theta_{0:t}, y_{1:t}) = p(\theta_0) \prod_{t \geq 1} p(\theta_t|\theta_{t-1}) p(y_t|\theta_t). \tag{1}$$

Também pela definição, temos que

$$p(\theta_t|\theta_{0:t-1}, y_{1:t-1}) = p(\theta_t|\theta_{t-1}).$$

Representação em grafos dos MEEs

$$
\begin{tikzcd}
\theta_0 \arrow[r] & \theta_1 \arrow[r] \arrow[d] & \theta_2 \arrow[r] \arrow[d] & \cdots \arrow[r] & \theta_{t-1} \arrow[r] \arrow[d] & \theta_t \arrow[r] \arrow[d] & \theta_{t+1} \arrow[d] \\
 & y_1 & y_2 & & y_{t-1} & y_t & y_{t+1}
\end{tikzcd}
$$

Interpretação Bayesiana

Ou seja, os problemas de *estimação* e *previsão* são resolvidos através da computação recursiva da distribuição condicional dos estados, dada a informação disponível. Nesse sentido, eles são naturalmente tratados dentro da **abordagem Bayesiana**.

Previsão, suavização e filtragem

Para um MEE, as principais tarefas são fazer inferências sobre os estados não observados ou prever futuras observações. Para estimar o vetor de estados, é necessário calcular as densidades condicionais $p(\theta_s|y_{1:t})$ quando:

- $s = t$ (densidade de **filtragem**), ou seja $p(\theta_{t-1}|y_{1:t-1})$,
- $s < t$ (densidade de **suavização**) ou seja, $p(\theta_t|y_{1:T})$,
- $s > t$ (densidade de **previsão**), ou seja, $p(\theta_t|y_{1:t-1})$,

em que T é o total de pontos no tempo observados.

Previsão, suavização e filtragem

Filtragem

- O problema de filtragem consiste em atualizar nossa inferência atual do vetor de estado à medida que novos dados se tornam disponíveis, ou seja, queremos estimar as densidades de filtragem $p(\theta_t|y_{1:t})$, então $p(\theta_{t+1}|y_{1:t+1})$, e assim por diante.

Previsão, suavização e filtragem

Suavização

- O problema de suavização ou análise retrospectiva consiste em estimar a seqüência de estados nos instantes $1, ..., t$ dada toda informação disponível $y_{1:T}$, isto é, estimar $p(\theta_{1:t}|y_{1:T})$.

Previsão, suavização e filtragem

Previsão

- O problema de previsão consiste em estimar y_{t+h} com base nos dados $y_{1:t}$, onde h é o número de passos à frente. Para isso, usa-se a densidade preditiva $p(\theta_{t+h}|y_{1:t}$ do vetor de estado e, com base nesta, computa-se a previsão de y_{t+1}.

Modelos Lineares Dinâmicos: definição formal

O Modelo Linear Dinâmico Gaussiano (MLD) é especificado por uma distribuição à priori Normal para o vetor de estado de q-dimensional no tempo $t = 0$,

$$\theta_0 \sim \mathcal{N}(m_0, C_0),$$

e pelo par de equações para $t \geq 1$,

$$y_t = F_t' \theta_t + \nu_t, \quad \nu_t \sim \mathcal{N}(0, V_t) \tag{2}$$

$$\theta_t = G_t \theta_{t-1} + \omega_t, \quad \omega_t \sim \mathcal{N}(0, W_t), \tag{3}$$

onde G_t e F_t' são matrizes $(q \times q)$ e $(m \times q)$ conhecidas e $\{\nu_t\}_{t \geq 1}$ e $\{\omega_t\}_{t \geq 1}$ são sequências independentes de vetores Gaussianos independentes com média zero e variâncias $\{V_t\}_{t \geq 1}$ e $\{W_t\}_{t \geq 1}$.

Modelos Lineares Dinâmicos: definição formal

Temos que (2) é conhecida como **equação de observação**, e (3) é a **equação dos estados** ou **equação de evolução**.

Assume-se em geral que θ_0 é independente dos erros ν_t e ω_t, para qualquer t.

Em resumo, o MLD Gaussiano é completamente caracterizado pelo conjunto $\{F, G, V, W\}_t = \{F_t, G_t, V_t, W_t\}$ para cada instante t.

MLD como um caso especial de MEE:

Pode-se demonstrar que o MLD satisfaz as suposições (A.1) and (A.2) que caracterizam um MEE, com

$$\begin{aligned} y_t|\theta_t &\sim \mathcal{N}(F_t'\theta_t, V_t), \\ \theta_t|\theta_{t-1} &\sim \mathcal{N}(G_t\theta_{t-1}, W_t). \end{aligned} \tag{4}$$

Proposição: Filtro de Kalman

Considere a definição do MLD e $\theta_{t-1}|y_{1:t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$.

1. A **distribuição preditiva dos estados** um passo à frente é

$$\theta_t|y_{1:t-1} \sim \mathcal{N}(a_t, R_t),$$

   onde $a_t = G_t m_{t-1}, \quad R_t = G_t C_{t-1} G_t' + W_t.$

2. A **distribuição preditiva de** y_t um passo à frente é

$$y_t|y_{1:t-1} \sim \mathcal{N}(f_t, Q_t)$$

   onde $f_t = F_t' a_t, \quad Q_t = F_t' R_t F_t + V_t.$

3. A **densidade de filtragem** dos estados é

$$\theta_t|y_{1:t} \sim \mathcal{N}(m_t, C_t),$$

   onde $m_t = a_t + R_t F_t Q_t^{-1} e_t, \quad C_t = R_t - R_t F_t Q_t^{-1} F_t' R_t$, sendo $e_t = y_t - f_t$ (erro de previsão).

Interpretação Bayesiana: fitro de Kalman

Pode-se interpretar como um problema de inferência Bayesiana do modelo linear Gaussiano

$$y_t = F_t'\theta_t + \nu_t, \quad \nu_t \sim \mathcal{N}(0, V_t),$$

em que o vetor de parâmetros da regressão tem a seguinte **priori**:

$$\theta_t | y_{1:t-1} \sim \mathcal{N}(a_t, R_t),$$

com V_t conhecida.

Pode-se demonstrar que $p(\theta_t | y_{1:t}) \propto p(y_t | \theta_t) p(\theta_t | y_{1:t-1})$, onde $p(y_t | \theta_t)$ é a verossimilhança. No caso, todas as distribuições são Gaussianas, de modo que a **posteriori** é, portanto,

$$\theta_t | y_{1:t} \sim \mathcal{N}(m_t, C_t).$$

Aprendizado Sequencial

$$\begin{tikzcd}
\text{Posteriori: } \theta_{t-1}|y_{1:t-1} \arrow[r] & \text{Priori: } \theta_t|y_{1:t-1} \arrow[r] \arrow[d] & \text{Posteriori: } \theta_t|y_{1:t} \\
& \text{Preditiva: } y_t|\theta_{t-1} &
\end{tikzcd}$$

Previsão em Modelos Dinâmicos

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

- A combinação da equação $y_t = F_t'\theta_t + \nu_t$ com a priori $p(\theta_t|y_{1:t-1})$ permite obter a distribuição preditiva $p(y_t|y_{1:t-1})$, que gerará as previsões.
- **Previsão pontual 1 passo à frente**: $E(y_t|y_{1:t-1})$
- **Previsão intervalar**: $p(y_t \in A|y_{1:t-1}) = \int_A p(y_t|y_{1:t-1})dy_t = \gamma\%$

Previsão em Modelos Dinâmicos

- **Previsão h passos à frente**: Se temos interesse em prever Y_{t+h} no tempo $t-1$, precisamos utilizar a equação do sistema sucessivamente até podermos escrever θ_{t+h} em função de θ_{t-1}.
- Por exemplo:

$$\theta_{t+2} = G_{t+2}\theta_{t+1} + \omega_{t+2},$$

$$\theta_{t+1} = G_{t+1}\theta_t + \omega_{t+1}, \text{ e}$$

$$\theta_t = G_t\theta_{t-1} + \omega_t, \text{ tal que}$$

$$\theta_{t+2} = G_{t+2}(G_{t+1}\theta_t + \omega_{t+1}) + \omega_{t+2}$$
$$\theta_{t+2} = G_{t+2}(G_{t+1}(G_t\theta_{t-1} + \omega_t) + \omega_{t+1}) + \omega_{t+2}.$$

Feito isso, combina-se com a equação das observações no tempo $t+h$ para obter preditiva $p(y_{t+h}|y_{1:t-1})$.

Suavizador de Kalman

Além das recursões de filtragem, as **recursões de suavização** também podem ser definidas em termos de médias e variâncias com a suposição de normalidade, conforme se segue.

Considere a definição do MLD e $\theta_{t+1}|y_{1:T} \sim \mathcal{N}(s_{t+1}, S_{t+1})$. Então,

$$\theta_t|y_{1:T} \sim \mathcal{N}(s_t, S_t),$$

onde

$$s_t = m_t + C_t G'_{t+1} R^{-1}_{t+1}(s_{t+1} - a_{t+1}),$$
$$S_t = C_t - C_t G'_{t+1} R^{-1}_{t+1}(R_{t+1} - S_{t+1}) R^{-1}_{t+1} G_{t+1} C_t.$$

Tipos de MLD: Modelo de Nível Local

O **Modelo de Nível Local** (MNL), ou MLD de primeira ordem possui apenas o componente de nível (μ_t) e do erro (ϵ_t), e é dado por

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, V_t)$$

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, W_t),$$

onde ϵ_t e η_t não correlacionados.

Simplificadamente, supõe-se que $V_t = V$ e $W_t = W$ são conhecidos.

Representação de Espaços de Estados e Filtro de Kalman para um MNL

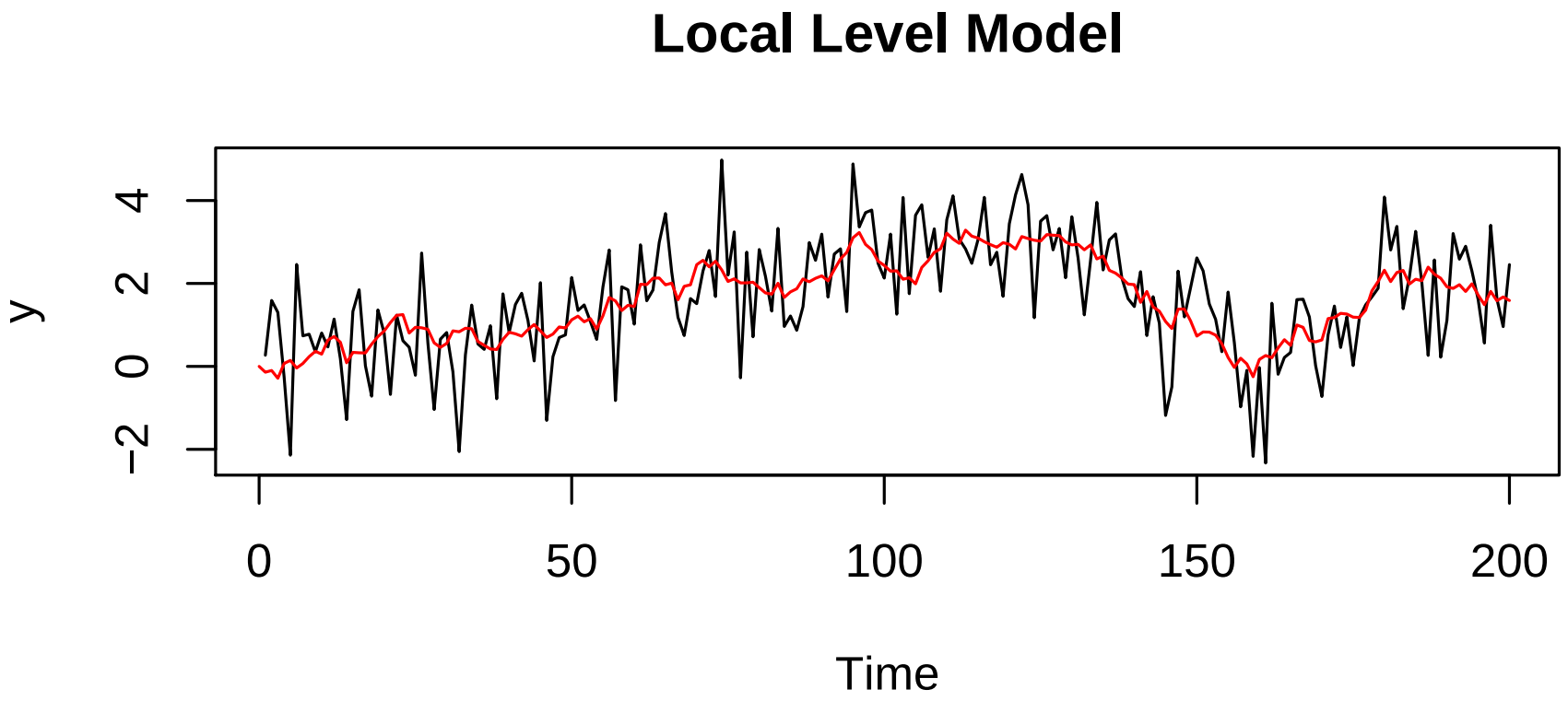
Considerando as equações de observação e de estado de um MLD, temos que

- $G_t = 1$
- $F_t' = 1$, assim temos que

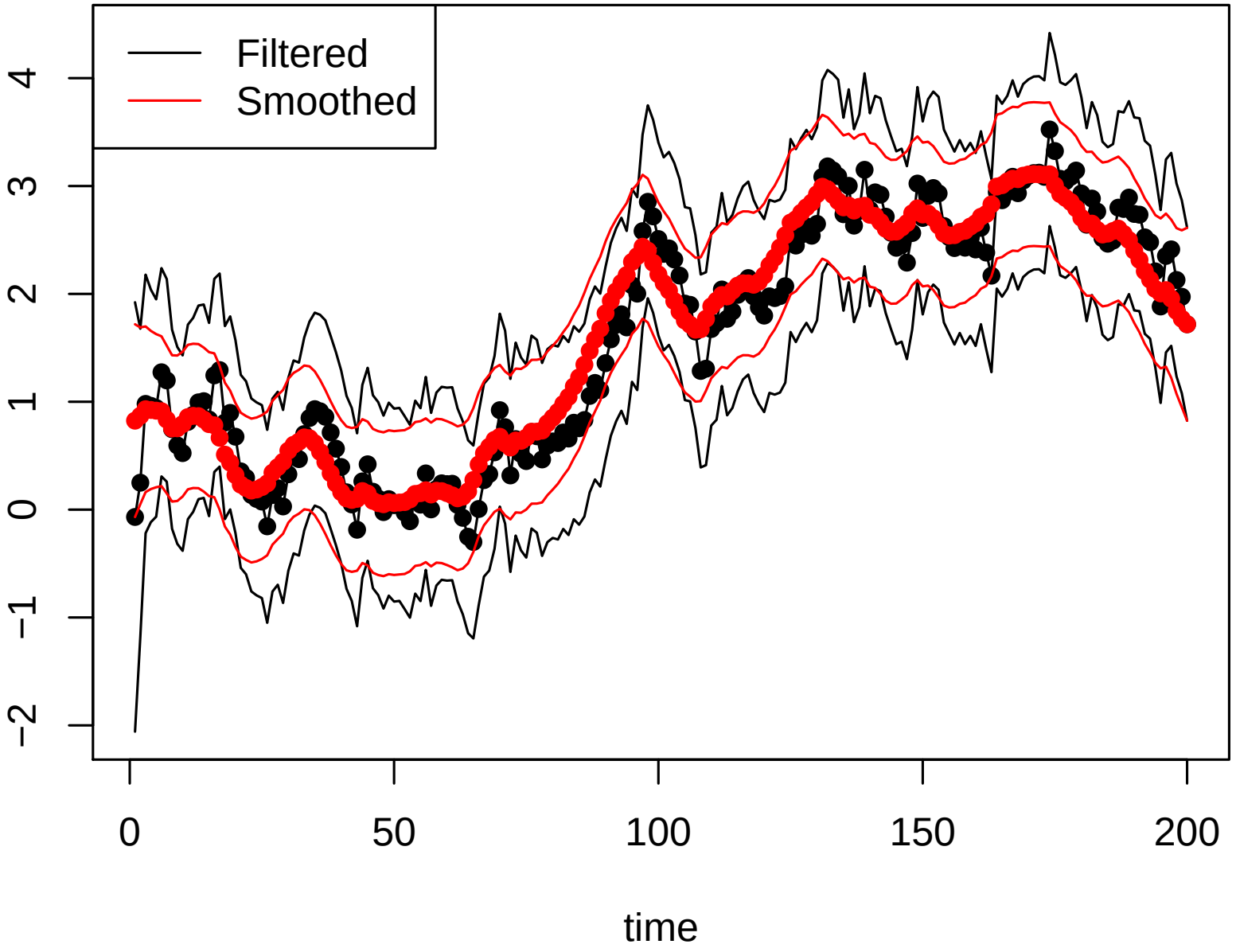
1. Posteriori para μ_{t-1}: $\mu_{t-1}|y_{1:t-1} \sim \mathcal{N}(m_{t-1}, C_{t-1})$
2. Priori para μ_t: $\mu_t|y_{1:t-1} \sim \mathcal{N}(m_{t-1}, R_t)$, com $R_t = C_{t-1} + W_t$
3. Preditiva: $y_t|y_{1:t-1} \sim \mathcal{N}(f_t, Q_t)$, com $f_t = m_{t-1}$ e $Q_t = R_t + V_t$
4. Posteriori para μ_t: $\mu_t|y_{1:t} \sim \mathcal{N}(m_t, C_t)$, com $m_t = m_{t-1} + A_t e_t$ e $C_t = A_t V_t$, onde $A_t = R_t/Q_t$ e $e_t = y_t - f_t$.

Simulação de um MNL

```
set.seed(1)
mu <- arima.sim(n=200,list(order=c(0,1,0)),sd=sqrt(0.05))
y <- mu + as.ts(rnorm(n=200,mean=0,sd=1))
par(mfrow=c(1,1))
ts.plot(y,main="Local Level Model")
lines(mu,col=2)
```



Filtro e Suavizador de Kalman para um MNL



Filtro de Kalman - código em R

```r
# initial conditions
m0 = 0
C0 = 100
m = m0
C = C0
mf = rep(0,n)
Cf = rep(0,n)
af = rep(0,n)
Rf = rep(0,n)

# Kalman filter
for (t in 1:n){
  # Prior at time t
  a  = m
  R = C+W
  af[t] = a
  Rf[t] = R
  # One-step ahead forecast
  f = a
  Q = R + V
  # Posterior at time t
  A = R/Q
  m = a+A*(y[t]-f)
  C = R-A*A*Q
  # Storing means and variances
  mf[t] = m
  Cf[t] = C
}
```

Suavizador de Kalman - código em R

```
# Kalman smoother recursions
mb=rep(0,n)
Cb=rep(0,n)
mb[n]=mf[n]
Cb[n]=Cf[n]
for (t in (n-1):1){
  mb[t] = mf[t] + Cf[t]/Rf[t+1]*(mb[t+1]-af[t+1])
  Cb[t] = Cf[t] - Cf[t]/Rf[t+1]*(Rf[t+1]-Cb[t+1])*Cf[t]/Rf[t+1]
}
```

Tipos de MLD: Modelo de Crescimento Linear

O **Modelo de Crescimento Linear** (MCL), ou MLD de segunda ordem possui dois componentes estruturais (μ_t e δ_t) e o erro (ϵ_t), e é dado por

$$
\begin{aligned}
y_t &= \mu_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, V_t) \\
\mu_t &= \mu_{t-1} + \delta_{t-1} + \omega_{1t}, \quad \omega_{1t} \sim \mathcal{N}(0, W_{1t}) \\
\delta_t &= \delta_{t-1} + \omega_{2t}, \quad \omega_{2t} \sim \mathcal{N}(0, W_{2t}),
\end{aligned}
$$

onde

- $\mu_0 \sim \mathcal{N}(\mu_\mu, \Omega_{10})$,
- $\delta_0 \sim \mathcal{N}(\mu_\delta, \Omega_{20})$, e
- V_t, W_{1t} e W_{2t} são não correlacionados.

Simplificadamente, supõe-se que $V_t = V$, $W_{1t} = W_1$ e W_{2t} são conhecidos.

O modelo propõe criar um parâmetro extra que descreve o crescimento do nível do processo.

Representação de Espaços de Estados e Filtro de Kalman para um MNL

Considerando as equações de observação e de estado de um MLD, temos que

$$G_t = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, F_t' = \begin{pmatrix} 1 & 0 \end{pmatrix}, \theta_t = \begin{pmatrix} \mu_t & \delta_t \end{pmatrix}$$

As distribuições a posteriori, priori e preditiva podem ser obtidas via filtro de Kalman.

Tipos de MLD: Regressão Dinâmica

Um caso importante do MLD Gaussiano é a regressão múltipla dinâmica através da origem (RD) que liga a resposta y_t aos q regressores $X_t = (X_{1t}, ..., X_{qt})$ no tempo t

$$\begin{aligned} y_t &= X_t\beta_t + \nu_t, \qquad & \nu_t \sim \mathcal{N}(0, \sigma_t^2), \\ \beta_t &= \beta_{t-1} + \omega_t, \qquad & \omega_t \sim \mathcal{N}(0, W_t), \end{aligned} \tag{5}$$

para $t = 1, ..., n$, em que

- X_t é o vetor de regressores $(1 \times q)$,
- $\beta_t' = (\beta_{1t}, ..., \beta_{qt})$ is the $(q \times 1)$ é o vetor de coeficientes e
- ν_t and ω_t são duas sequeências independentes de erros Gaussianos com média zero e variâncias σ_t^2 and W_t, respectivamente.

Representação de Espaços de Estados e Filtro de Kalman para um RD

Ou seja, usando a definição formal do MLD, temos

- $F_t' = X_t$,
- $\theta_t = \beta_t$ (os estados),
- $G_t = G = I_q$ e
- $V_t = \sigma_t^2$.

Note que se $W_t = 0$ para todo t, $\beta_t = \beta$, i.e., o caso de **regressão estática**.

Aplicação: Regressão Dinâmica

Uma aplicação frequente é sobre dados financeiros, por exemplo, o modelo CAPM, que relaciona o retorno de uma determinada empresa em função do retorno de mercado.

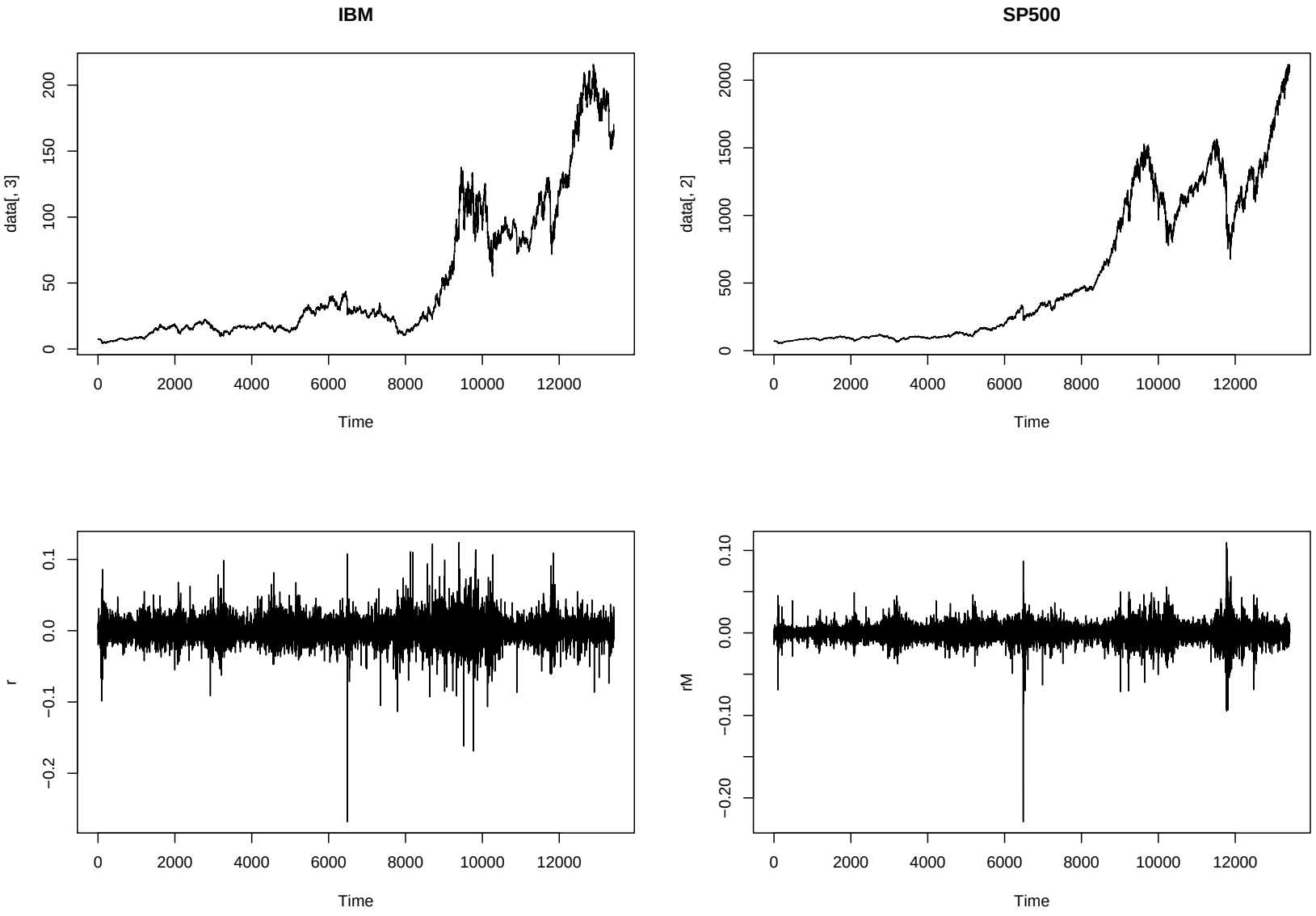
$r_t = \alpha_t + \beta_t r_{M,t} + \epsilon_t, \quad \epsilon_t \sim mathcalN(0, \sigma^2)$

Ou seja, na representação de um MLD

- $F_t' = (1, r_{M,t})$
- $\theta_t = (\alpha_t, \beta_t)'$
- $V = \sigma^2$
- $W = diag(\sigma_\alpha^2, \sigma_\beta^2)$

com $\theta_0 \sim \mathcal{N}(0, kI_2)$.

IBM x SP500



IBM x SP500: regressão estática

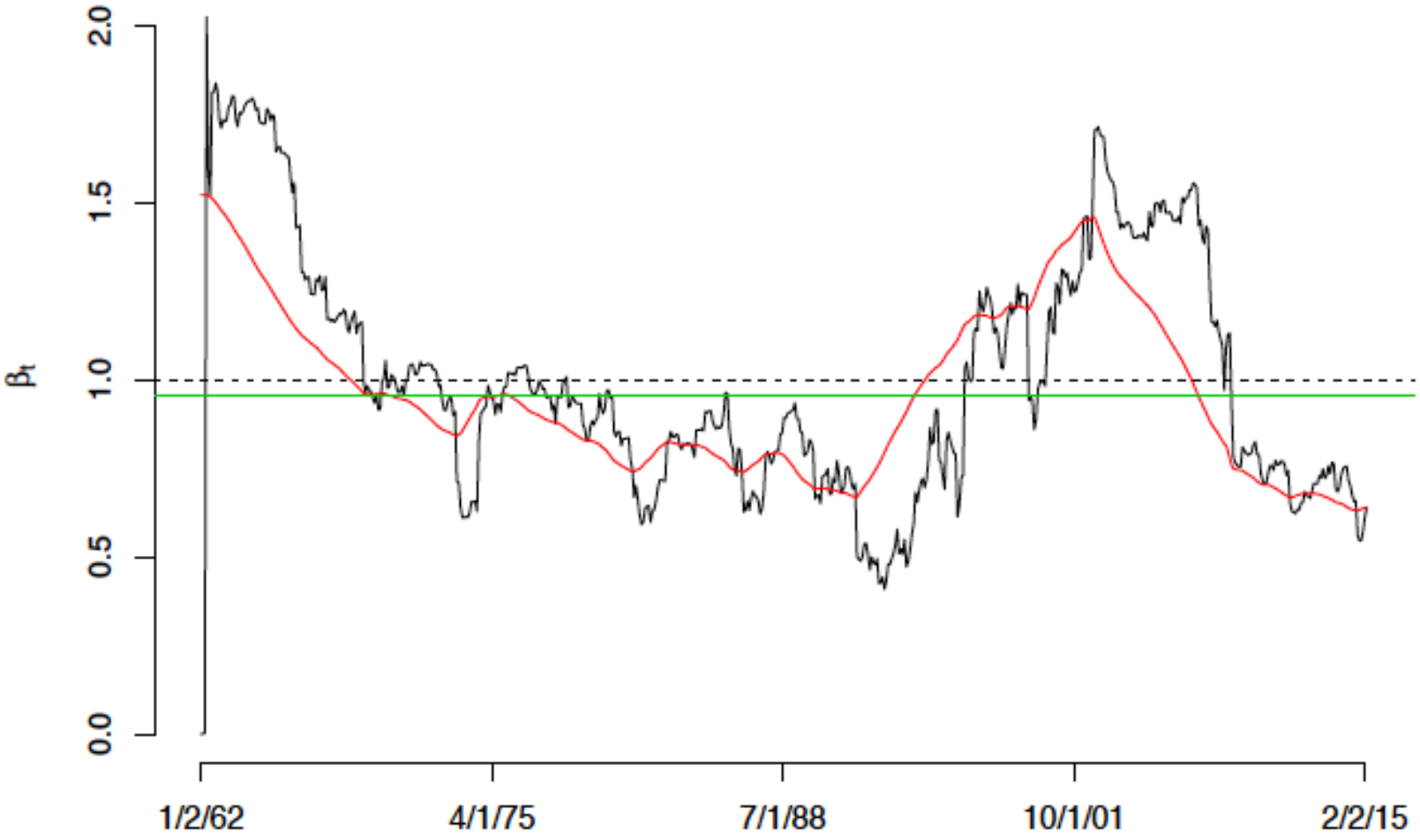
```r
library(readr)
data <- as.data.frame(read_csv("/Volumes/NO NAME 1/sp500.csv"))
```

```
## Parsed with column specification:
## cols(
##   Date = col_character(),
##   SP500 = col_double(),
##   IBM = col_double()
## )
```

```r
r = diff(log(data[,3]))
rM = diff(log(data[,2]))
capm = lm(r~rM)
summary(capm)
```

```
##
## Call:
## lm(formula = r ~ rM)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.163087 -0.005968 -0.000129  0.005764  0.114565
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.076e-05  1.065e-04  -0.195    0.845
## rM           9.972e-01  1.043e-02  95.637   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01233 on 13416 degrees of freedom
## Multiple R-squared:  0.4054, Adjusted R-squared:  0.4053
## F-statistic:  9146 on 1 and 13416 DF,  p-value: < 2.2e-16
```

IBM x SP500: regressão dinâmica



MLD com parâmetros desconhecidos

Na maioria dos casos, as matrizes F_t' e G_t são conhecidas (especialmente para modelos estruturais, como o MCL e MNL). Porém, raramente são as variâncias V_t e W_t.

Na **abordagem Bayesiana**, assume-se distribuições a priori para estas variâncias, em geral, da família Gama Inversa pela conveniência da conjugação.

Na abordagem clássica, os parâmetros são estimados por **Máxima Verossimilhança**.

Algoritmo FFBS

Considere Φ o vetor de parâmetros desconhecidos. Enquanto que $p(\Phi|\theta_{0:t}, y_{1:t})$ é específica ao problema, $p(\theta_{0:t}|\Phi, y_{1:t})$ tem uma expressão geral.

Vimos que o suavizador de Kalman fornece $\theta_t|y_{1:T}$. Assim, pode-se demonstrar que $\theta_{0:T}$ dado $y_{1:T}$ é

$$p(\theta_{0:T}|y_{1:T}) = \prod_{t=0}^{T} p(\theta_t|\theta_{t+1:T}, y_{1:T}),$$

em que o último fator $p(\theta_T|y_{1:T})$ é a distribuição de filtragem de θ_T.

Pelo filtro de Kalman, temos $\theta_T|y_{1:T} \sim \mathcal{N}(m_T, C_T)$. Assim, podemos começar usando o Filtro de Kalman sob θ_T e depois recursivamente amostrar $t = T-1, T-2, ..., 0$ usando o suavizador de Kalman.

Algoritmo FFBS

Esse algoritmo é conhecido como forward filtering backward sampling (FFBS) e é devido a Carter & Kohn (1994) e Frühwirth-Schnatter (1994).

Algoritmo:

1. Rodar 1 passo do filtro de Kalman
2. Amostrar $\theta_T | y_{1:T} \sim \mathcal{N}(m_T, C_T)$.
3. Para $t = T-1, ..., 0$, amostrar $(\theta_t | \theta_{t+1:T}, y_{1:T}) \sim \mathcal{N}(h_t, H_t)$.

Implementação do FFBS para o MNL

Exemplo prático

Obrigada!

Bibliografia

- Chris K Carter and Robert Kohn. On gibbs sampling for state space models. Biometrika, 81 (3):541–553, 1994.
- Sylvia Frühwirth-Schnatter. Data augmentation and dynamic linear models. Journal of time series analysis, 15(2):183–202, 1994.
- Mike West and Jeff Harrison. Bayesian Forecasting and Dynamic Models (2Nd Ed.). Springer-Verlag New York, Inc., New York, NY, USA, 1997.
- AC Harvey. Forecasting, structural time series models and the kalman filter. 1989.