

Espaço de Estados

Lane Alencar¹

¹IME-USP

2012

Índice

Motivação

Exemplos

Inferência

Motivação

Propor e ajustar um modelo que apresenta componentes não observadas. Essas componentes podem ser:

- ▶ tendência, sazonalidade;
- ▶ dados faltantes.

Harvey (1961): Filtro de Kalman para obter EMV via decomposição do erro de predição

Modelo

$$y_t = F_t x_t + D_t z_t + v_t, \quad (1)$$

$$x_t = \mu + G_t x_{t-1} + w_t. \quad (2)$$

em que:

1. x_t é um vetor $p \times 1$;
2. z_t é um vetor de variáveis conhecidas $r \times 1$;
3. F_t é a matriz de observação $q \times p$;
4. D_t é uma matriz $q \times r$;
5. G_t é a matriz de transição $p \times p$;
6. $v_t \text{ iid} \sim N_q(\mathbf{0}, V)$, $w_t \text{ iid} \sim N_q(\mathbf{0}, W)$ e assumiremos v_t e w_t não correlacionados apenas para simplificarmos a apresentação do modelo;
7. $x_0 \sim N_p(\nu, \Sigma)$.

Obs

A equação (1) é denominada equação de observação e relaciona y_t com uma variável não observada x_t , de ordem $p \times 1$, denominada variável de estado.

A equação (2) é denominada equação de estado. Note que a variável de estado x_t deve conter toda informação necessária para que sua dinâmica dependa essencialmente de x_{t-1} , assim x_t é cadeia de Markov.

Aplicação

Muitos exemplos de aplicações utilizando modelos espaço de estados são encontrados em Shumway e Stoffer (2000), Harvey (1990) e Kim e Nelson (1999). Dentre diversos exemplos destacamos:

- ▶ análise de parâmetros biomédicos monitorados ao longo do tempo com muitos dados omissos (Jones 1993);
- ▶ modelos ARMA podem ser analisados sob a forma espaço de estados;
- ▶ construção de indicador econômico coincidente extraíndo a componente comum de diversas variáveis macroeconômicas (Stock e Watson 1991).

Exemplo 1 - Modelo AR(2)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + a_t, t = 1, \dots, N$$

$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} x_t$$

$$x_t = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} x_{t-1} + \begin{bmatrix} a_t \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} Z_t \\ Z_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} Z_{t-1} \\ Z_{t-2} \end{bmatrix} + \begin{bmatrix} a_t \\ 0 \end{bmatrix}$$

Parametrização Alternativa - Hyndman (acho estranho ter parâmetro em x_t)

$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} x_t$$

$$x_t = \begin{bmatrix} y_t \\ \phi_2 y_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \phi_2 y_{t-2} \end{bmatrix} + \begin{bmatrix} a_t \\ 0 \end{bmatrix}$$

Exemplo 2 - Modelo AR(p)

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + a_t, t = 1, \dots, N$$

$$y_t = [1 \ 0 \ \dots \ 0] x_t$$
$$x_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix} x_{t-1} + \begin{bmatrix} a_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Exemplo 3 - Modelo MA(1)

$$y_t = \mu + \theta a_{t-1} + a_t, t = 1, \dots, N$$

$$y_t = \mu + [1 \quad \theta] \begin{bmatrix} a_t \\ a_{t-1} \end{bmatrix}$$

$$x_t = \begin{bmatrix} a_t \\ a_{t-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_{t-1} \\ a_{t-2} \end{bmatrix} + \begin{bmatrix} a_t \\ 0 \end{bmatrix}$$

$$V = 0, \quad W = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$

Exemplo 4 - Modelo ARMA(1,1)

$$y_t = \phi y_{t-1} + \theta e_{t-1} + e_t, t = 1, \dots, N$$

$$y_t = [1 \ 0] x_t = [1 \ 0] \begin{bmatrix} y_t \\ \theta e_t \end{bmatrix}$$

$$x_t = \begin{bmatrix} y_t \\ \theta e_t \end{bmatrix} = \begin{bmatrix} \phi & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \theta e_{t-1} \end{bmatrix} + \begin{bmatrix} e_t \\ \theta e_{t-1} \end{bmatrix}$$

$$V = 0 \quad W = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$

Exemplo 5 - Modelo ARMA(p,q)

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t, t = 1, \dots, N$$

Ver em Hyndman e

[http://www-stat.wharton.upenn.edu/
stine/stat910/lectures/14_state_space.pdf](http://www-stat.wharton.upenn.edu/stine/stat910/lectures/14_state_space.pdf)

Modelo de Regressão Linear

$$y_t = z_t \beta_t + a_t, t = 1, \dots, N$$

com y_t variável resposta, z_t variável explicativa e β_t vetor de coeficientes desconhecidos. Podemos considerar

- ▶ Modelo estático: $\beta_t = \beta$;
- ▶ Modelo dinâmico: $\beta_t = \beta_{t-1} + w_t$.

Nível Local

$$\begin{aligned}Z_t &= \mu_t + \mathbf{e}_t, t = 1, \dots, N \\ \mu_t &= \mu_{t-1} + \eta_t\end{aligned}$$

com Z_t variável resposta e μ_t o nível local.

Temos $e_t \sim N(0, \sigma_e^2)$, $\eta_t \sim N(0, \sigma_\eta^2)$ erros independentes.

Tendência Local

$$\begin{aligned}Z_t &= \mu_t + e_t, t = 1, \dots, N \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t &= \beta_{t-1} + \xi_t\end{aligned}$$

com Z_t variável resposta,

μ_t o nível local e β_t a inclinação local.

Temos $e_t \sim N(0, \sigma_e^2)$, $\eta_t \sim N(0, \sigma_\eta^2)$ e $\xi_t \sim N(0, \sigma_\xi^2)$, todos erros independentes.

Tendência Local e Componente Sazonal

$$Z_t = \mu_t + S_t + e_t, t = 1, \dots, N$$

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

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

$$S_t = -S_{t-1} + -S_{t-2} \dots - S_{t-s+1} + a_t$$

com Z_t variável resposta,

S_t a componente sazonal com sazonalidade s ,

μ_t o nível local e β_t a inclinação local.

Temos $e_t \sim N(0, \sigma_e^2)$, $\eta_t \sim N(0, \sigma_\eta^2)$, $\xi_t \sim N(0, \sigma_\xi^2)$ e $a_t \sim N(0, \sigma_a^2)$, todos erros independentes.

Tendência Local e Componente Sazonal Trimestral

$$Z_t = \mu_t + S_t + e_t$$

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

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

$$S_t = -S_{t-1} - S_{t-2} - S_{t-3} + a_t$$

$$Z_t = (1 \ 0 \ 1 \ 0 \ 0)X_t + e_t$$

$$X_t = \begin{pmatrix} \mu_t \\ \beta_t \\ S_t \\ S_{t-1} \\ S_{t-2} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \beta_{t-1} \\ S_{t-1} \\ S_{t-2} \\ S_{t-3} \end{pmatrix} + \begin{pmatrix} \eta_t \\ \xi_t \\ a_t \\ 0 \\ 0 \end{pmatrix}$$

Inferência

A modelagem na forma espaço de estados envolve as seguintes etapas:

- ▶ definição de modelo identificável;
- ▶ fixados os parâmetros do modelo, utiliza-se o Filtro de Kalman para obtenção de previsões lineares de mínimos quadrados para o vetor de estados.
- ▶ estimação dos parâmetros;
- ▶ avaliação do ajuste do modelo.

Identificabilidade

Podemos parametrizar os coeficientes e a estrutura de correlação em função de Θ , vetor $k \times 1$, de modo que: $F_t = F_t(\Theta)$, $D = D(\Theta)$, $\mu = \mu(\Theta)$, $G = G(\Theta)$, $V = V(\Theta)$, $W = W(\Theta)$, $\nu = \nu(\Theta)$ e $\Sigma = \Sigma(\Theta)$.

Identificabilidade

Segundo Rothenberg(1971), um modelo é dito globalmente identificável em um particular parâmetro θ_0 se, para qualquer valor de θ , existe uma possível realização $Y_T = (y_1, \dots, y_T)$ tal que o valor da verossimilhança em θ seja diferente do valor em θ_0 . Quando não temos identificabilidade? É mais fácil entender, considerando, por exemplo:

$$\begin{aligned}y_t &= Fx_t + v_t, \\x_t &= Gx_{t-1} + w_t.\end{aligned}$$

com y_t e x_t escalares.

Note que se considero $(2x_t) = G2x_{t-1} + (2w_t)$ e $F/2$ obtenho exatamente o mesmo modelo. A função de verossimilhança também será a mesma.

Filtro de Kalman

Seja $\tilde{y}_s = (y_1, \dots, y_s)$, o vetor de informações disponíveis até o instante s . Supondo o parâmetro Θ conhecido, o filtro de Kalman consiste no algoritmo que fornece um estimador para x_t baseado em \tilde{y}_s . Adotando a notação apresentada em Shumway e Stoffer (2000), definimos:

$$\begin{aligned}x_{t|s} &= \mathbf{E}(x_t | \tilde{y}_s), \\P_{t|s} &= \mathbf{E}[(x_t - x_{t|s})(x_t - x_{t|s})' | \tilde{y}_s], \\P_{t_1, t_2 | s} &= \mathbf{E}[(x_{t_1} - x_{t_1|s})(x_{t_2} - x_{t_2|s})' | \tilde{y}_s].\end{aligned}$$

Sob a hipótese de normalidade, obteremos o estimador $x_{t|s}$ e sua variância $P_{t|s}$. Mesmo sem a hipótese de normalidade, os estimadores obtidos são os que minimizam o erro quadrático médio na classe dos estimadores lineares (Shumway e Stoffer (2000)).

Filtro de Kalman

A partir dos valores iniciais $x_{0|0} = \mu$ e $P_{0|0} = \Sigma$, o Filtro de Kalman apresenta equações de predição para $s = t - 1$ e de atualização para $s = t$, com $t = 1, \dots, T$.

Equações de Predição

$$\begin{aligned}x_{t|t-1} &= \mu + Gx_{t-1|t-1}, \\P_{t|t-1} &= GP_{t-1|t-1}G' + W.\end{aligned}$$

Filtro de Kalman

Considerando o conhecimento de y_t , atualizamos a predição.

Equações de Atualização

$$\begin{aligned}x_{t|t} &= x_{t|t-1} + K_t(y_t - F_t x_{t|t-1} - D z_t), \\ P_{t|t} &= P_{t|t-1} - K_t F_t P_{t|t-1},\end{aligned}$$

em que K_t é o ganho do filtro,

$$K_t = P_{t|t-1} F_t' (F_t P_{t|t-1} F_t' + V)^{-1}.$$

Filtro de Kalman

Os estimadores de x_t baseados na informação completa da amostra \tilde{y}_T são obtidos a partir das equações conhecidas como equações de suavização. Esses estimadores suavizados correspondem aos estimadores de mínimo erro quadrático médio. A partir dos valores de $x_{T|T}$ e $P_{T|T}$ utilizamos as equações de suavização para $t = T, \dots, 1$.

Equações de Suavização

$$\begin{aligned}x_{t-1|T} &= x_{t-1|t-1} + J_{t-1}(x_{t|T} - Gx_{t-1|t-1} - \mu), \\P_{t-1|T} &= P_{t-1|t-1} + J_{t-1}(P_{t|T} - P_{t|t-1})J'_{t-1},\end{aligned}$$

onde

$$J_{t-1} = P_{t-1|t-1} G'(P_{t|t-1})^{-1}.$$

Filtro de Kalman

A completa derivação das equações do Filtro de Kalman e de suavização é apresentada em Harvey (1990) e Hamilton (1994), onde também são encontradas equações para a obtenção de previsões $x_{t|s}$ e $P_{t|s}$ para $t > s$ e as propriedades de convergência do filtro de Kalman para sistemas estáveis. A obtenção das equações de predição, atualização e suavização segundo a inferência bayesiana é apresentada em West e Harrison (1997).

Verossimilhança

A estimação dos parâmetros será realizada pelo método de máxima verossimilhança. Para o modelo apresentado a função logaritmo da verossimilhança é dada por

$$\ln(L) = \sum_{t=1}^T |\Sigma_t(\Theta)| + \sum_{t=1}^T \epsilon_t(\Theta)' \Sigma_t(\Theta) \epsilon_t(\Theta),$$

onde

$$\begin{aligned}\epsilon_t(\Theta) &= y_t - F_t x_{t|t-1} - D z_t, \\ \Sigma_t(\Theta) &= F_t P_{t|t-1} F_t' + V.\end{aligned}$$

Verossimilhança

A maximização pode ser realizada pelo método Newton-Raphson (Gupta e Mehra (1974)), pelo algoritmo escore (Harvey (1989)). Nesse caso, a variância assintótica de $\hat{\Theta}$ é obtida a partir do inverso da matriz informação de Fisher.

Outro método de estimação utiliza o algoritmo EM (Shumway e Stoffer, 1982). A formulação do algoritmo EM, principais propriedades e alguns exemplos foi pioneiramente apresentada em Dempster, Laird e Rubin (1977). Alguns resultados de convergência para o algoritmo EM são encontrados em Wu (1983), McLahan e Krishnan (1997) que apresentam uma completa discussão do assunto e Shumway e Stoffer (1982) que utilizam o algoritmo EM para modelos espaço de estados.

Verossimilhança Completa

Basicamente, o objetivo do algoritmo é maximizar a função de verossimilhança completa, ou seja, verossimilhança para os dados observados e não observados, $(\tilde{y}_T, \tilde{x}_T) = (y_1, \dots, y_T, x_0, \dots, x_T)$:

$$f_{\Theta}(\tilde{y}_T, \tilde{x}_T) = f_{\nu, \Sigma}(x_0) \prod_{t=1}^T f_{G, W}(x_t | x_{t-1}) f_{F, V}(y_t | x_t).$$

Para o modelo proposto, temos que o $\ln(L(\Theta))$ é:

$$\begin{aligned} & -\frac{1}{2} \ln |\Sigma| - \frac{1}{2} (x_0 - \nu)' \Sigma^{-1} (x_0 - \nu) + \\ & -\frac{n}{2} \ln |W| - \frac{1}{2} \sum_{t=1}^T (x_t - \mu - Gx_{t-1})' W^{-1} (x_t - \mu - Gx_{t-1}) \\ & -\frac{n}{2} \ln |V| - \frac{1}{2} \sum_{t=1}^T (y_t - F_t x_t - Dz_t)' V^{-1} (y_t - F_t x_t - Dz_t). \end{aligned}$$

Algoritmo EM

O algoritmo EM é um algoritmo iterativo e ao ser iniciado o vetor de parâmetros deve ser inicializado com $\Theta = \Theta^{(0)}$. Em cada iteração j , o algoritmo é dividido em 2 passos. No passo de cálculo da esperança (E), é calculada a esperança da função $\ln(L(\Theta))$ dada toda informação disponível Y_T para $\Theta = \Theta^{(j-1)}$. No passo de maximização (M), encontraremos os estimadores que maximizam $H = E_{\Theta^{(j-1)}}(\ln(L(\Theta)) | \tilde{y}_T)$.

Passo E

$$\begin{aligned}
 & \mathbf{E}_{\Theta(\hat{y}_{-1})}(\ln(L(\Theta))|\tilde{y}_T) \\
 & -\frac{1}{2}\ln|\Sigma| - \frac{1}{2}\text{tr}\{\Sigma^{-1}[P_{0|T} + (x_{0|T} - \nu)'(x_{0|T} - \nu)]\} \\
 & -\frac{1}{2}\text{tr}[W^{-1}(C - BG' - GB' + GAG')] - \frac{n}{2}\ln|W| \\
 & -\frac{1}{2}\text{tr}\{\sum_{t=1}^T V^{-1}[(y_t - F_t x_{t|T} - Dz_t)(y_t - F_t x_{t|T} - Dz_t)' + F_t P_{t|T} F_t']\} -
 \end{aligned}$$

$$A = \sum_{t=1}^T P_{t-1|T} + x_{t-1|T} x_{t-1|T}'$$

$$B = \sum_{t=1}^T P_{t,t-1|T} + x_{t|T} x_{t-1|T}'$$

$$C = \sum_{t=1}^T P_{t|T} + x_{t|T} x_{t|T}'$$

Passo M

Em seguida, encontramos o estimador $\Theta^{(j)}$ que maximiza H . Por exemplo, para F_t conhecido e para $\Theta = (G, V, W, \mu, \Sigma)$, temos os estimadores (Shumway e Stoffer 2000):

$$\hat{G}_t = BA^{-1},$$

$$\hat{W} = \frac{1}{T}(C - B\hat{G}' - \hat{G}B' + \hat{G}A\hat{G}'),$$

$$\hat{V} = \frac{1}{T} \sum_{t=1}^T [(y_t - F_t x_{t|T} - Dz_t)(y_t - F_t x_{t|T} - Dz_t)' + F_t P_{t|T} F_t'],$$

$$\hat{\mu} = x_{0|T},$$

$$\hat{\Sigma} = P_{0|T}.$$

Variância

Para avaliar a distribuição e calcular medidas de variabilidade associadas aos estimadores dos parâmetros, Stoffer e Wall (1991) sugerem um método bootstrap para modelos espaço de estados.

Tópicos para Discussão em Shumway e Stoffer

- ▶ Bootstrap;
- ▶ Alterações quando há dados faltantes;
- ▶ Modelo com mudança de regime;
- ▶ Modelos não lineares e não gaussianos usando métodos Monte Carlo;
- ▶ Volatilidade Estocástica.

GARCH(1,1) e Volatilidade Estocástica

$$\begin{aligned}r_t &= \sigma_t e_t, \quad e_t \sim RB \\ \sigma_t^2 &= \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2\end{aligned}$$

Se definimos $h_t = \log \sigma_t^2$ e $y_t = \log r_t^2$, temos a equação de observação $y_t = h_t + \log e_t^2$, com a variável de estado (não obs.) h_t .
No modelo de volatilidade estocástica temos:

$$\begin{aligned}y_t &= h_t \log(e_t^2) \\ h_t &= \phi_0 + \phi_1 h_{t-1} + w_t, \quad w_t \sim NID(0, \sigma_w^2)\end{aligned}$$

Modelo de nível local

$$\begin{aligned}Z_t &= \mu_t + e_t, \\ \mu_t &= \mu_{t-1} + \eta_t.\end{aligned}$$

Com erros independentes $e_t \sim RB(\sigma_e^2)$ e $\eta_t \sim RB(\sigma_\eta^2)$.
Nesse modelo, μ_t^t é uma média móvel das observações passadas com um parâmetro de suavização que é função da razão sinal ruído, $f = \sigma_\eta^2 / \sigma_e^2$. A previsão fica constante.

Modelo de tendência local

$$\begin{aligned}Z_t &= \mu_t + e_t, \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, \\ \beta_t &= \beta_{t-1} + \xi_t.\end{aligned}$$

Com erros independentes $e_t \sim RB(\sigma_e^2)$, $\eta_t \sim RB(\sigma_\eta^2)$ e $\xi_t \sim RB(\sigma_\xi^2)$.

μ_t é o nível local e β_t é a inclinação local.

$$\begin{aligned}Z_t &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + e_t \\ \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \xi_t \end{bmatrix}\end{aligned}$$





R

R: StructTS e Kalmanlike (stats)

Não me fornece erros padrões das estimativas.

R: Programas do cap. 6 de Shumway and Stoffer (2006).

Tinham problema para o R2.8.0 e posteriores, pois não consegue utilizar uma library que trabalha com vetores indexados a partir de 0.

-  Alencar, A. P. (2006). Tese de Doutorado.
-  Hamilton, J. D. (1994). *Time series analysis*. Princeton, N.J.: Princeton University Press.
-  Morettin, P. A. e Toloi, C.M.C. (2006). *Análise de Séries Temporais*. Edgard Blucher.
-  Shumway, R. H. e Stoffer, D. S. (2000). *Time series analysis and its application*. New York: Springer-Verlag.
-  Stoffer, D. S. e Wall, K. (1991). Bootstrapping state space models: Gaussian maximum likelihood estimation and the Kalman filter. *Journal of the American Statistician Association*, 86, 1024-1033.