

3

Modelos de Espaço de Estado e o filtro de Kalman

3.1

Introdução

O intuito deste capítulo é fazer uma breve introdução sobre a metodologia de espaço de estado, adotada pelo modelo proposto por De Jong & Zehnwirth[2]. O entendimento desta abordagem é de extrema importância para a análise e discussão do referido modelo.

A forma em *espaço de estado* de um modelo estatístico é uma ferramenta poderosa que permite a manipulação de uma grande variedade de modelos de séries temporais. Como afirmado por Harvey[13], uma vez que o modelo seja colocado na forma de espaço de estado, o filtro de Kalman pode ser aplicado, abrindo possibilidade de utilização de algoritmos de previsão, filtragem, suavização (*smoothing*) e estimação dos parâmetros fixos do modelo.

3.2

A forma de Espaço de Estado

A forma genérica de espaço de estado (SSF) é aplicada a uma série temporal multivariada, \mathbf{y}_t , contendo N elementos. Estas variáveis observáveis estão relacionadas com um vetor de dimensão $m \times 1$, $\boldsymbol{\alpha}_t$, conhecido como *vetor dos estados*, através da *equação das observações*

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{d}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T \quad (3-1)$$

onde \mathbf{Z}_t é uma matriz $N \times m$, \mathbf{d}_t é um vetor $N \times 1$ e $\boldsymbol{\varepsilon}_t$ é um vetor $N \times 1$ de erros serialmente não correlacionados de média zero e matriz covariância \mathbf{H}_t , ou seja

$$E(\boldsymbol{\varepsilon}_t) = \mathbf{0} \text{ e } \text{Var}(\boldsymbol{\varepsilon}_t) = \mathbf{H}_t$$

Em geral, os elementos de $\boldsymbol{\alpha}_t$ são não-observáveis. Entretanto, eles são gerados por um processo Markoviano de primeira ordem,

$$\boldsymbol{\alpha}_t = \mathbf{T}_t \boldsymbol{\alpha}_{t-1} + \mathbf{c}_t + \mathbf{R}_t \boldsymbol{\eta}_t, \quad t = 1, \dots, T \quad (3-2)$$

onde \mathbf{T}_t é uma matriz $m \times m$, \mathbf{c}_t é um vetor $m \times 1$, \mathbf{R}_t é uma matriz $m \times g$ e $\boldsymbol{\eta}_t$ é um vetor $N \times 1$ de erros serialmente não correlacionados de média zero e matriz covariância \mathbf{Q}_t , ou seja

$$E(\boldsymbol{\eta}_t) = \mathbf{0} \text{ e } \text{Var}(\boldsymbol{\eta}_t) = \mathbf{Q}_t$$

A equação (3-2) é a *equação de transição* ou de *estado* do modelo. A inclusão da matriz \mathbf{R}_t multiplicando o ruído $\boldsymbol{\eta}_t$ é, até certo ponto, arbitrária. O ruído pode ser sempre redefinido de forma a ter uma matriz covariância na forma $\mathbf{R}_t \mathbf{Q}_t \mathbf{R}_t'$.

Em Durbin & Koopman[4] a equação de transição é definida de uma forma alternativa:

$$\boldsymbol{\alpha}_{t+1} = \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{c}_t + \mathbf{R}_t \boldsymbol{\eta}_t, \quad t = 1, \dots, T \quad (3-3)$$

Apesar da forma apresentada na equação (3-3) ser a mais moderna, estaremos utilizando a notação apresentada em Harvey[13] e em De Jong & Zehnwirth[2].

De forma a se completar a especificação do sistema de espaço de estado, é necessário assumir duas premissas:

- o vetor de estado inicial, $\boldsymbol{\alpha}_0$, possui média \mathbf{a}_0 e matriz covariância \mathbf{P}_0 conhecidos, isto é

$$E(\boldsymbol{\alpha}_0) = \mathbf{a}_0 \text{ e } \text{Var}(\boldsymbol{\alpha}_0) = \mathbf{P}_0$$

- os erros $\boldsymbol{\varepsilon}_t$ e $\boldsymbol{\eta}_t$ não são correlacionados entre si em todos os instantes de tempo, e também não são correlacionados com o estado inicial, isto é

$$E(\boldsymbol{\varepsilon}_t \boldsymbol{\eta}_s') = 0 \text{ para todo } s, t = 1, \dots, T \quad (3-4a)$$

$$E(\boldsymbol{\varepsilon}_t \boldsymbol{\alpha}_0') = 0 \text{ para todo } t = 1, \dots, T \quad (3-4b)$$

$$E(\boldsymbol{\eta}_t \boldsymbol{\alpha}_0') = \mathbf{0} \text{ para todo } t = 1, \dots, T \quad (3-4c)$$

A condição (3-4a) pode ser relaxada.

As matrizes \mathbf{Z}_t , \mathbf{d}_t e \mathbf{H}_t da equação das observações e as matrizes \mathbf{T}_t , \mathbf{c}_t , \mathbf{R}_t e \mathbf{Q}_t da equação de transição são chamadas de *matrizes do sistema*. Em geral, essas matrizes dependem de um conjunto de parâmetros desconhecidos. Estes parâmetros são representados pelo vetor $\boldsymbol{\Psi}$, de dimensão $n \times 1$ e serão chamados de *hiperparâmetros*, de forma a diferenciá-los dos parâmetros que por ventura possam entrar no modelo via \mathbf{c}_t ou \mathbf{d}_t . Segundo Harvey[13], os hiperparâmetros determinam as propriedades estocásticas do modelo, enquanto que os parâmetros que aparecem em \mathbf{c}_t e \mathbf{d}_t somente afetam o valor esperado do estado ou das observações de forma determinística. No modelo estudado (que será apresentado no capítulo 4), $\mathbf{c}_t = \mathbf{d}_t = 0$.

3.3

O filtro de Kalman

O filtro de Kalman consiste em um procedimento recursivo com objetivo de computar o estimador ótimo do vetor de estado no instante t , baseado na informação disponível até este instante. Tal informação consiste nas observações até \mathbf{y}_t inclusive. Presume-se que as matrizes de sistema, \mathbf{a}_0 e \mathbf{P}_0 sejam conhecidas.

No caso onde os distúrbios e o vetor de estado inicial são normalmente distribuídos, o filtro de Kalman possibilita calcular a função de máxima verossimilhança via a *decomposição preditiva do erro*. Assim, pode-se estimar qualquer parâmetro desconhecido do modelo. Também, segundo Harvey[13], a verossimilhança provê a base para testes estatísticos e testes de especificações do modelo.

3.3.1

Forma Geral do filtro de Kalman

Considerando o modelo de espaço de estado definido por (3-1) e (3-2), seja \mathbf{a}_{t-1} o estimador ótimo de $\boldsymbol{\alpha}_{t-1}$ baseado nas observações até \mathbf{y}_{t-1} inclusive. Seja \mathbf{P}_{t-1} a matriz covariância $m \times m$ de seu erro estimado, isto é:

$$\mathbf{P}_{t-1} = E [(\boldsymbol{\alpha}_{t-1} - \mathbf{a}_{t-1})(\boldsymbol{\alpha}_{t-1} - \mathbf{a}_{t-1})'] \quad (3-5)$$

Dado \mathbf{a}_{t-1} e \mathbf{P}_{t-1} , o estimador ótimo de $\boldsymbol{\alpha}_t$ é dado por

$$\mathbf{a}_{t|t-1} = \mathbf{T}_t \mathbf{a}_{t-1} + \mathbf{c}_t \quad (3-6)$$

enquanto que a matriz covariância do erro estimado é

$$\mathbf{P}_{t|t-1} = \mathbf{T}_t \mathbf{P}_{t-1} \mathbf{T}'_t + \mathbf{R}_t \mathbf{Q}_t \mathbf{R}'_t, \quad t = 1, \dots, T \quad (3-7)$$

Essas duas equações – (3-6) e (3-7) – são conhecidas como *equações preditivas*.

Uma vez que a nova observação, \mathbf{y}_t , torna-se disponível, o estimador de $\boldsymbol{\alpha}_t$ ($\mathbf{a}_{t|t-1}$), deve ser atualizado. As *equações de atualização* são

$$\mathbf{a}_t = \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{Z}'_t \mathbf{F}_t^{-1} (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_{t|t-1} - \mathbf{d}_t) \quad (3-8)$$

e

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{P}_{t|t-1} \quad (3-9)$$

onde

$$\mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_{t|t-1} \mathbf{Z}'_t + \mathbf{H}_t, \quad t = 1, \dots, T \quad (3-10)$$

As equações (3-6), (3-7), (3-8), (3-9) e (3-10) compõem o filtro de Kalman.

Os valores iniciais do filtro de Kalman devem ser especificados em função de \mathbf{a}_0 e \mathbf{P}_0 , ou $\mathbf{a}_{1|0}$ e $\mathbf{P}_{1|0}$. Dadas essas condições iniciais, o filtro de Kalman calcula o estimador ótimo do vetor de estado à medida que novas observações tornam-se disponíveis. Quando todas as T observações forem processadas, o filtro terá retornado o estimador ótimo do vetor de estado atual. Esse estimador possui todas as informações necessárias para a realização de predições ótimas dos valores futuros do estado e das observações.

3.3.2

Estimação por máxima verossimilhança e a decomposição preditiva dos erros

Como afirma Harvey[13], a teoria clássica de máxima verossimilhança assume que as T observações, $\mathbf{y}_1, \dots, \mathbf{y}_T$ são independentes e identicamente distribuídas.

Em um modelo de série temporal, as observações geralmente não são independentes. Assim, para a obtenção da função de máxima verossimilhança é preciso obter a função densidade de probabilidade conjunta do

modelo, a qual é dada por

$$L(\mathbf{y}; \Psi) = \prod_{t=1}^T p(\mathbf{y}_t | \mathbf{Y}_{t-1}) \quad (3-11)$$

onde $p(\mathbf{y}_t | \mathbf{Y}_{t-1})$ representa a distribuição de \mathbf{y}_t condicionada a toda a informação disponível até o instante $t-1$, isto é $\mathbf{Y}_{t-1} = \{\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_1\}$.

Se os ruídos e o vetor de estado inicial do modelo (3-1) possuírem distribuições normais multivariadas, a distribuição de \mathbf{y}_t condicional em \mathbf{Y}_{t-1} será também normal. Além disso, a média e a matriz covariância desta distribuição condicional serão obtidas diretamente através do filtro de Kalman. Das equações do filtro de Kalman mostradas na seção 3.3.1, podemos perceber que α_t condicionado em \mathbf{Y}_{t-1} é normalmente distribuído com média $\mathbf{a}_{t|t-1}$ e matriz covariância $\mathbf{P}_{t|t-1}$.

Para o caso Gaussiano podemos escrever a função de verossimilhança (3-11) como

$$\log L = -\frac{NT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T \log |\mathbf{F}_t| - \frac{1}{2} \sum_{t=1}^T \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \quad (3-12)$$

onde $\mathbf{v}_t = \mathbf{y}_t - \tilde{\mathbf{y}}_{t|t-1}$, $t = 1, \dots, T$. A equação (3-12) é chamada de *decomposição preditiva dos erros* da verossimilhança.

A função verossimilhança deve ser maximizada em relação aos parâmetros desconhecidos Ψ . Para isso é necessário a utilização de métodos numéricos, como será visto no capítulo 5. É importante a exploração de linearidades na função de forma a reduzir a complexidade da maximização.

3.3.3

Concentração de parâmetros da verossimilhança

Como mostrado em Durbin & Koopman[4], é vantajoso reparametrizar o modelo antes da maximização da verossimilhança de forma a reduzir a dimensão da busca numérica. Por exemplo, para um *modelo de nível local* podemos estabelecer uma razão sinal/ruído $q = \sigma_\eta^2 / \sigma_\varepsilon^2$, obtendo assim

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2), \\ \alpha_t &= \alpha_{t-1} + \eta_t, & \eta_t &\sim N(0, q\sigma_\varepsilon^2), \end{aligned}$$

e pode-se estimar o par de parâmetros σ_ε^2 e q ao invés de σ_ε^2 e σ_η^2 . Ainda segundo Durbin & Koopman[4] e Fernandes[7], é possível reparametrizar as

matrizes \mathbf{P}_t e \mathbf{F}_t como

$$\mathbf{P}_t^* = \frac{\mathbf{P}_t}{\sigma_\varepsilon^2},$$

$$\mathbf{F}_t^* = \frac{\mathbf{F}_t}{\sigma_\varepsilon^2}$$

Assim, para um modelo *univariado* pode-se reescrever a equação (3-10) do filtro de Kalman como

$$\mathbf{F}_t = \mathbf{Z}_t \mathbf{P}_{t|t-1} \mathbf{Z}'_t + \sigma_\varepsilon^2 \quad (3-13)$$

$$\sigma_\varepsilon^2 \mathbf{F}_t^* = \mathbf{Z}_t \sigma_\varepsilon^2 \mathbf{P}_{t|t-1}^* \mathbf{Z}'_t + \sigma_\varepsilon^2 \quad (3-14)$$

$$\mathbf{F}_t^* = \mathbf{Z}_t \mathbf{P}_{t|t-1}^* \mathbf{Z}'_t + 1 \quad (3-15)$$

Em seguida obtém-se a verossimilhança para a nova parametrização

$$\log L(q, \sigma_\varepsilon^2) = -\frac{T}{2} \log 2\pi - \frac{T}{2} \log \sigma_\varepsilon^2 - \frac{1}{2} \sum_{t=1}^T \log |\mathbf{F}_t^*| - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T \mathbf{v}'_t \mathbf{F}_t^{*-1} \mathbf{v}_t \quad (3-16)$$

Para se estimar o parâmetro σ_ε^2 , precisamos diferenciar (3-16) em relação à σ_ε^2 :

$$\frac{\partial \log L}{\partial \sigma_\varepsilon^2} = -\frac{T}{2\sigma_\varepsilon^2} + \frac{1}{2(\sigma_\varepsilon^2)^2} \sum_{t=1}^T \mathbf{v}'_t \mathbf{F}_t^{*-1} \mathbf{v}_t \quad (3-17)$$

Assim, chega-se a

$$\tilde{\sigma}_\varepsilon^2(q) = \frac{1}{T} \sum_{t=1}^T \mathbf{v}'_t \mathbf{F}_t^{*-1} \mathbf{v}_t \quad (3-18)$$

Finalmente, substituindo esta expressão na equação de verossimilhança reparametrizada, obtém-se a verossimilhança concentrada

$$\log L(q) = -\frac{T}{2} \log(2\pi + 1) - \frac{1}{2} \sum_{t=1}^T \log |\mathbf{F}_t^*| - \frac{T}{2} \log \tilde{\sigma}_\varepsilon^2(q) \quad (3-19)$$

3.4

Testes e Diagnósticos

Em um modelo bem-especificado, os resíduos devem ser aproximadamente aleatórios. Este fato pode ser investigado através de procedimentos gráficos e diversos testes estatísticos. Nesta seção serão abordados alguns testes comumente utilizados em modelos de espaço de estado, que serão empregados na modelagem a ser realizada no capítulo 5.

3.4.1

Resíduos e procedimentos gráficos

Diagnósticos realizados no domínio do tempo são baseados nas *inovações* obtidas através do filtro de Kalman inicializado com uma *priori* difusa. Para um dado conjunto de hiperparâmetros Ψ , os *resíduos* são definidos como as inovações padronizadas. Apesar de o modelo apresentado no capítulo 4 ser multivariado, os testes serão realizados em cada componente do resíduo ε_t separadamente. Assim, a expressão dos resíduos será

$$\tilde{v}_t = \frac{v_t}{f_t^{\frac{1}{2}}}, \quad t = 2, \dots, T \quad (3-20)$$

Se o modelo estiver corretamente especificado, e os hiperparâmetros Ψ são conhecidos, os resíduos terão a propriedade

$$\tilde{v}_t \sim \text{NID}(0, 1) \quad (3-21)$$

Segundo Harvey[13], no caso mais usual de se ter que estimar os hiperparâmetros em Ψ , a expressão (3-21) será aproximada.

3.4.2

Testes de especificação baseados nos resíduos

Correlação Serial

As autocorrelações dos resíduos são definidas como

$$r_v(\tau) = \frac{\sum_{t=2+\tau}^T (\tilde{v}_t - \bar{\tilde{v}}) (\tilde{v}_{t-\tau} - \bar{\tilde{v}})}{\sum_{t=2}^T (\tilde{v}_t - \bar{\tilde{v}})^2}, \quad \tau = 1, 2, \dots \quad (3-22)$$

e o correlograma resultante permite identificar a presença de correlação serial nos resíduos.

Normalidade

Para se obter a estatística de Jarque-Bera, é necessário calcular o terceiro e quarto momentos dos resíduos (inovações padronizadas). O terceiro momento (assimetria) e o quarto momento (curtose) de uma amostra são respectivamente

$$S = \frac{\text{E}(v_t - \mu)^3}{\sigma^3} \quad K = \frac{\text{E}(v_t - \mu)^4}{\sigma^4} \quad (3-23)$$

onde $\mu = E(v_t)$. No caso de normalidade, $S = 0$ e $K = 3$.

A estatística de Jarque-Bera é dada pela seguinte relação

$$JB = \frac{n}{6}(\hat{S} - 0)^2 + \frac{n}{24}(\hat{K} - 3)^2 \sim \chi^2 \quad (3-24)$$

onde \hat{S} e \hat{K} são as medidas amostrais de S e K , dadas por:

$$\hat{S} = \frac{\sum_{i=1}^n (\tilde{v}_i - \bar{\tilde{v}})^3/n}{(\sum_{i=1}^n (\tilde{v}_i - \bar{\tilde{v}})^2/n)^{3/2}} \quad \hat{K} = \frac{\sum_{i=1}^n (\tilde{v}_i - \bar{\tilde{v}})^4/n}{(\sum_{i=1}^n (\tilde{v}_i - \bar{\tilde{v}})^2/n)^2} \quad (3-25)$$