

3

Filtro de Kalman

3.1

Os Três Enfoques: Predição, Filtragem e Suavização

Uma representação em espaço de estado de uma série temporal é motivada por dois objetivos principais, quais sejam:

- dado um instante t , estimar e/ou prever α_t com base na σ -álgebra $\zeta_j \equiv \sigma(Y_1, Y_2, \dots, Y_j)$, sendo $j \in \{1, 2, \dots, n\}$, e depois interpretar os resultados obtidos;
- uma vez previsto α_t , prever Y_t .

Dessa forma, teríamos que primeiramente estimar e/ou prever α_t para que pudéssemos atingir os dois objetivos. A depender da σ -álgebra ζ_j abordada, temos então três questões a serem consideradas:

- se $t > j$, temos então uma operação de *predição* ou *previsão* de α_t ;
- se $t = j$, temos então uma operação de *filtragem* ou *atualização* de α_t ;
- se $t < j$, temos então uma operação de *suavização* ou *interpolação* de α_t .

Nas próximas seções, serão apresentadas as equações recursivas que determinam o filtro de Kalman para os três tipos de operação enunciados acima. Para tanto, faz-se necessário antecipadamente estabelecermos a seguinte notação dada por:

- $a_{t/j} \equiv E(\alpha_t / \zeta_j)$;
- $P_{t/j} \equiv E[(\alpha_t - a_{t/j})(\alpha_t - a_{t/j})']$;
- $a_t \equiv a_{t/t-1}$;
- $P_t \equiv P_{t/t-1}$;
- $v_t \equiv Y_t - E(Y_t / \zeta_{t-1}) = Y_t - Z_t a_t - d_t$;
- $F_t \equiv V(v_t / \zeta_{t-1}) = Z_t P_t Z_t' + H_t$;

nas quais:

- $a_{t/j}$ é um vetor aleatório definido como a esperança condicional dada uma σ -álgebra ζ_j no contexto de normalidade ou, em caso de uma representação em espaço de estado mais amplo, como o estimador linear ótimo de α_t com base em Y_1, Y_2, \dots, Y_j ou a projeção ortogonal do vetor de estado α_t sobre o conjunto de todas as combinações lineares das coordenadas de $(1, Y_1', Y_2', \dots, Y_j')$. Com o intuito de facilitar a notação, será admitida a notação $E(./.)$ em ambos os casos, sem perda de interpretação, uma vez que os dois tratamentos são muito similares em termos algébricos;

- $P_{t/j}$ representa a matriz de erros médios quadráticos incondicionais associados às projeções de $a_{t/j}$. Especificamente no contexto de normalidade, $P_{t/j}$ representa também a matriz variância-covariância condicional de α_t dada uma σ -álgebra ζ_j .

- a_t e P_t representam, respectivamente, notações simplificadas para $a_{t/t-1}$ e $P_{t/t-1}$, definidos acima onde $j = t - 1$ em ambos os casos.

- v_t é denominado de **vetor de inovação**, que é o conteúdo informacional de Y_t que não está presente em ζ_{t-1} . v_t possui média incondicional nula e variância condicional dada por F_t . Além disso, temos que v_1, v_2, \dots, v_n são não

correlacionados. Num contexto específico de normalidade, v_t possui também variância incondicional dada por F_t , uma vez que v_t e $(Y_1', Y_2', \dots, Y_j)'$ são não correlacionados para todo $j \in \{1, 2, \dots, t-1\}$ ou, como nesse caso específico, independentes. Ainda num ambiente Gaussiano, temos que v_1, v_2, \dots, v_n são independentes.

3.1.1

Filtro de Kalman: Previsão e Atualização

Uma vez que o modelo de interesse está representado em um formato de espaço de estado linear, podemos aplicar o filtro de Kalman, um conjunto de equações recursivas utilizado, em especial, para estabelecer algoritmos de previsão, atualização e suavização.⁵

Os seguintes conjuntos de equações constituem o filtro de Kalman ou as recursões de Kalman para previsão (ou predição) e atualização (ou filtragem) para um dado $t \in \{1, 2, \dots, n\}$, seguindo a notação já explicitada na seção anterior:

$$\left. \begin{aligned} a_{t+1} &= T_t a_{t/t} + c_t \\ P_{t+1} &= T_t P_{t/t} T_t' + R_t Q_t R_t' \end{aligned} \right\} \text{Filtro de Kalman : equações de predição} \quad (3.1)$$

$$\left. \begin{aligned} a_{t/t} &= a_t + P_t Z_t' F_t^- v_t \\ P_{t/t} &= P_t - P_t Z_t' F_t^- Z_t P_t \end{aligned} \right\} \text{Filtro de Kalman : equações de atualização} \quad (3.2)$$

onde F_t^- é qualquer inversa generalizada de F_t .

São pertinentes as seguintes observações acerca das equações do filtro de Kalman enunciadas acima:

⁵ Para maiores detalhes acerca da dedução do filtro de Kalman para previsão, atualização e suavização sob o contexto de normalidade, vide Harvey (1989, 1993) e Durbin e Koopman (2001). Sob o contexto de estimadores lineares ótimos, por outro lado, conferir Brockwell e Davis (2002) e Shumway e Stoffer (2006) para uma leitura mais aprofundada.

1. As equações do filtro de Kalman para previsão e atualização são válidas num contexto geral, seja num ambiente Gaussiano ou mais amplo. A diferença é que $a_{t/j}$, onde $j = t - 1, t$, consiste de esperança condicional dada uma σ -álgebra ζ_j quando no ambiente de normalidade pleno e, em caso contrário, é visto como o estimador linear ótimo de α_t com base em Y_1, Y_2, \dots, Y_j , como mencionado na seção anterior.
2. As recursões de Kalman produzem estimadores ótimos, no sentido de que são não tendenciosos e minimizam a matriz de erros médios quadráticos incondicional associada à estimação de α_t . Quando o contexto Gaussiano se aplica, temos que as recursões de Kalman se traduzem em esperanças condicionais e, portanto, em *estimadores ótimos globais*. Em contrapartida, as mesmas recursões constituem-se em *estimadores lineares ótimos*, isto é, uma vez que restringimos nossa análise aos estimadores que são combinações lineares das observações, temos então que as recursões de Kalman nos fornecem os estimadores lineares que minimizam a matriz de erros médios quadráticos incondicionais.
3. As equações de predição e atualização se intercomunicam, no sentido de que os resultados adquiridos na equação de predição são necessários para a obtenção dos resultados dados pela equação de atualização e vice-versa. Nesse sentido, é possível gerar um filtro de Kalman mais sintético, através do processo de substituição das equações de atualização nas de previsão. Esse processo origina o *filtro de Kalman "2 em 1"*, dado pelas seguintes equações:

$$\left. \begin{aligned} a_{t+1} &= T_t a_t + c_t + K_t v_t \\ P_{t+1} &= T_t P_t L_t' + R_t Q_t R_t' \end{aligned} \right\} \text{Filtro de Kalman "2 em 1"} \quad (3.3)$$

nas quais $K_t = T_t P_t Z_t' F_t^{-1}$ e $L_t = T_t - K_t Z_t'$. Importante mencionar que a nomenclatura "2 em 1" não é necessariamente adotada formalmente na literatura.

4. A necessidade de se utilizar inversas generalizadas para F_t surge nos casos em que F_t não é invertível, nos casos em que ao menos uma das coordenadas univariadas de Y_t é combinação linear perfeita das outras coordenadas univariadas de Y_t e/ou é combinação linear perfeita das coordenadas de Y_1, Y_2, \dots, Y_{t-1} . Embora possamos evitar esses casos singulares ao eliminar tais observações redundantes, o uso de inversas generalizadas, quaisquer que elas sejam, não viola a propriedade de unicidade das recursões de Kalman.

3.1.2

Filtro de Kalman: Suavização

Nesta seção, são apresentadas as recursões de Kalman no contexto de suavização, nas quais são consideradas todas as informações disponíveis para a estimação do vetor de estado, inclusive as informações baseadas em instantes de tempo posteriores ao de análise.

A operação de suavização, usualmente, pode ser dividida em três enfoques, a saber: ponto fixo, defasagem fixa e intervalo fixo. O estudo dos dois primeiros, em virtude de não corresponder ao escopo desta dissertação, não serão aqui examinados [vide Harvey (1989) e Anderson e Moore (1979) para um estudo mais aprofundado desses dois procedimentos].

O enfoque de interesse para esta dissertação – o *enfoque do intervalo fixo* – consiste de equações recursivas dentro de um contexto *backward*, no sentido de que as recursões “andam de trás para frente”, percorrendo desde observações posteriores ao instante de tempo de análise até o instante de tempo mais remoto.

Antes de apresentarmos as recursões correspondentes ao *suavizador de Kalman* sob o enfoque do intervalo fixo, é pertinente mencionar que há duas formulações para o suavizador de Kalman que possuem diferenças consideráveis quanto à eficácia computacional, embora sejam matematicamente equivalentes (com efeito, visto que tanto sob o contexto de esperança condicional quanto sob o contexto de estimadores lineares ótimos, a propriedade da unicidade do suavizador de Kalman se mantém válida).

Além disso, devemos inserir uma notação adicional à estabelecida na seção 3.1. Para o que segue, $\hat{\alpha}_t \equiv a_{t/n}$ e $V_t \equiv P_{t/n}$, onde $a_{t/n}$ e $P_{t/n}$ são definidos como na seção 3.1., com $j = n$.

Posto isso, seguem abaixo as equações recursivas que definem as duas formulações para o suavizador de Kalman [cf. Harvey (1989, 1993) e Durbin e Koopman (2001)] sob o enfoque do intervalo fixo, o *suavizador de Kalman tradicional (SKT)* e o *suavizador de Kalman de de Jong (SKJ)*, em alusão ao artigo atribuído a de Jong (1989), onde essa formulação é discutida de forma mais abrangente e aprofundada:

$$\left. \begin{aligned} \hat{\alpha}_t &= a_{t/t} + A_t(\hat{\alpha}_{t+1} - a_{t+1}) \\ V_t &= P_{t/t} - A_t(P_{t+1} - V_{t+1})A_t' \\ A_t &= P_{t/t}T_t'P_{t+1}^- \\ t &= n-1, n-2, \dots, 2, 1 \end{aligned} \right\} \text{Suavizador de Kalman Tradicional} \quad (3.4)$$

$$\left. \begin{aligned} \hat{\alpha}_t &= a_t + P_t r_{t-1} \\ r_{t-1} &= Z_t' F_t^- v_t + L_t' r_t \\ V_t &= P_t - P_t N_{t-1} P_t \\ N_{t-1} &= Z_t' F_t^- Z_t + L_t' N_t L_t \\ t &= n, n-1, \dots, 2, 1 \end{aligned} \right\} \text{Suavizador de Kalman de de Jong} \quad (3.5)$$

com as condições iniciais dadas por $r_n = 0$ e $N_n = 0$. Ademais, F_t^- e P_{t+1}^- correspondem a quaisquer inversas generalizadas de F_t e de P_{t+1} , respectivamente, e L_t é definido igualmente como no filtro de Kalman “2 em 1” [vide equação (3.3)].

São pertinentes as seguintes observações acerca das equações que constituem os suavizadores de Kalman enunciados acima:

1. A nomenclatura utilizada para classificar as duas formulações para o suavizador de Kalman não corresponde necessariamente à adotada formalmente na literatura. Entretanto, esta nomenclatura será eventualmente utilizada ao longo desta dissertação.

2. As observações de números 1, 2 e 4 pertinentes às equações do filtro de Kalman para previsão e atualização permanecem válidas sob o contexto de suavização.
3. Fica patente que as equações do filtro de Kalman sob os contextos de previsão e atualização, incluindo o filtro de Kalman “2 em 1”, são fundamentais para o cálculo das equações de suavização, em especial a_t , P_t , v_t e F_t^- para o cálculo do SKJ e a_t , P_t , $a_{t/t}$ e $P_{t/t}$ para o cálculo do SKT.
4. O SKJ apresenta vantagens consideráveis em relação ao SKT, tanto do ponto de vista teórico quanto ao que tange à eficiência computacional. Uma vantagem teórica diz respeito ao fato de já haver uma teoria bem desenvolvida para a inicialização do SKJ, um ponto delicado na teoria acerca do filtro de Kalman que será visto em seções posteriores desta dissertação. Por outro lado, SKJ apresenta vantagens competitivas em relação ao SKT no que diz respeito à implementação computacional, quais sejam:
 - o SKT exige um número maior de multiplicações e de operações de inversão matricial (com efeito, são necessárias $n - 1$ inversões matriciais de P_{t+1} , além das inversões matriciais utilizadas no cálculo do filtro de Kalman). Dependendo do tamanho da matriz P_{t+1} , o ganho computacional ao implementar o SKJ pode se mostrar considerável, em especial em linguagens “mais altas”;
 - Como mencionado na observação 3, o SKT exige o armazenamento de a_t , P_t , $a_{t/t}$ e $P_{t/t}$, onde usualmente $a_{t/t}$ e $P_{t/t}$ possuem dimensões maiores do que v_t e F_t^- , o que confere maior necessidade de memória para o armazenamento das primeiras quantidades;
 - As saídas r_t , N_t , K_t e L_t , obtidas via a implementação do SKJ, podem desempenhar papel importante na análise de diagnósticos do modelo de estudo em espaço de estado, além de constituírem-se em quantidades necessárias ao cálculo de $\hat{\alpha}_t$ e V_t .

5. No que concerne à minimização de erros médios quadráticos, o estimador $\hat{\alpha}_t$ é mais preciso que $a_{t/t}$ o qual, por sua vez, é mais preciso que a_t , uma vez que $V_t \leq P_{t/t} \leq P_t$. Isso confirma a intuição de que mais precisa será a estimação quanto maior for o conjunto de informações utilizado ou, em termos mais técnicos, mais ampla for a σ -álgebra ζ_j a qual o estimador está submetido. Para uma demonstração dessa propriedade, ver Pizzinga (2004), seção 3.5, página 28.

3.2

Inicialização do Filtro de Kalman

Para que as recursões de Kalman apresentadas nas seções anteriores possam ser implementadas, é necessário que o vetor de estado inicial α_1 seja definido de uma forma adequada no que diz respeito aos seus momentos de 1ª e 2ª ordens.

Entretanto, a_1 e P_1 não são usualmente conhecidos, uma vez que, na prática, é pouco provável que todas as coordenadas do vetor de estado sejam estacionárias. Sendo assim, torna-se necessária uma abordagem teórica para lidar com o problema de *inicialização do filtro de Kalman*, isto é, a definição de a_1 e P_1 .

Em Durbin e Koopman (2001), um *modelo geral* é proposto para o vetor de estado α_1 , dado por:

$$\alpha_1 = a + A\delta + R_0\eta_0 \quad (3.6)$$

$$\delta \sim (0, \mathbf{\Sigma}_\delta) \quad (3.7)$$

$$\eta_0 \sim (0, Q_0) \quad (3.8)$$

no qual a é vetor $m \times 1$ de constantes, δ é vetor $q \times 1$ não-observável cuja distribuição está definida em (3.7), bem como $Cov(\delta, \eta_0) = E(\delta\eta_0') = 0$. A e R_0 são matrizes tais que $A = [col_{i_1} I_m \dots col_{i_q} I_m]$ e $R_0 = [col_{j_1} I_m \dots col_{j_{m-q}} I_m]$ sendo que $i_k \neq j_l, \forall k \neq l$, o que resulta em $A'R_0 = 0$, na qual 0 representa um vetor nulo de dimensão $q \times (m-q)$. Finalmente, supõe-se que Q_0 é uma matriz positiva definida

e as coordenadas do vetor a , correspondentes às coordenadas não-nulas de $A\delta$, são nulas.

Dessa forma, podemos desmembrar α_t em dois blocos diferentes: um correspondente às *coordenadas não-estacionárias* de α_t presentes em $A\delta$ e outro correspondente às *coordenadas estacionárias* de α_t , presentes em $R_0\eta_0$. Cumpre notar que é no contexto de *não-estacionariedade* que encontramos dificuldades para estabelecer médias e variâncias incondicionais. Quando α_t consiste de um processo estocástico estacionário, podemos facilmente obter a_1 e P_1 via a equação do estado, tendo apenas que, eventualmente, estimar algum parâmetro desconhecido.

Dispondo do modelo geral como estabelecido acima, podemos então manusear κ de forma que a variância de δ assuma um *valor arbitrariamente grande*, a fim de refletir a total ignorância a respeito das condições iniciais referentes às coordenadas não-estacionárias de α_t .

Desse modo, temos então um problema de *inicialização difusa do filtro de Kalman*, uma vez que ao menos uma das coordenadas do vetor de estado inicial é tratada de forma difusa, ou seja, com a variância assumindo um valor arbitrariamente grande.

Tendo em vista a solução deste problema, podemos adotar dois procedimentos de diferentes complexidades tanto do ponto de vista teórico quanto do computacional, quais sejam:

- estabelecer um valor arbitrariamente grande, mas, obviamente, não infinito para κ , como algo entre 10^3 e 10^7 , por exemplo. Nesse caso, temos um problema de *inicialização difusa aproximada*.
- manusear κ de forma que ele tenda a infinito. Nesse caso, temos o problema de *inicialização difusa exata*.

3.2.1

Inicialização Difusa Aproximada

Como já mencionado anteriormente, trata-se de uma solução aproximada para o problema de inicialização do filtro de Kalman, onde se arbitra um valor

suficientemente alto para κ [vide Harvey e Phillips (1979) e Harvey (1989) para maiores detalhes].

Entre as vantagens que esse procedimento possui em comparação ao de inicialização de forma exata, citam-se: trata-se de uma solução prática e fácil de ser implementada; e não está estruturado dentro de um contexto teórico, o que demanda poucos esforços em termos de pesquisa e estudo.

A opção por adotar este tipo de procedimento possui, entretanto, alguns inconvenientes, dentre eles:

- A dificuldade em se estabelecer o instante de tempo d , a partir do qual as recursões de Kalman se estabilizam, ou seja, adquirem um caráter não-difuso. Nesses casos, são adotados usualmente procedimentos *ad hoc*, que não possuem sustentação teórica quanto à validade dessa escolha. Em modelos univariados, por exemplo, usualmente d corresponde ao número de coordenadas não estacionárias de α_t , conforme Harvey (1989);
- A possibilidade de que as recursões implementadas não gerem resultados confiáveis, devido à propagação de erros de aproximação ao longo das recursões em decorrência da própria instabilidade desse tipo de processo.

3.2.2

Inicialização Difusa Exata

Diferentemente do enfoque prático da inicialização aproximada do filtro de Kalman, esta metodologia está bem estruturada dentro de um contexto teórico e mais acadêmico, fornecendo um procedimento mais convincente quando comparado ao anterior. A inicialização difusa exata aposta em uma nova formulação de recursões para o filtro de Kalman, que se dividem em duas partes: uma anterior ao instante d , incluindo-o, e outra que tem início em $d + 1$ e prossegue até o final da amostra.

O primeiro grupo se caracteriza por um conjunto de recursões que levam em conta o caráter difuso do processo de inicialização, o que o torna bem diferente do conjunto de recursões que definem o filtro de Kalman usual, como podemos ver

na seqüência, em que são apresentadas as equações que definem o filtro de Kalman sob o contexto de inicialização exata. As recursões são dadas por⁶:

Filtro de Kalman: Inicialização Exata

$$\begin{aligned}
 F_{\infty,t} &= Z_t P_{\infty,t} Z_t' & F_{*,t} &= Z_t P_{*,t} Z_t' + H_t \\
 M_{\infty,t} &= P_{\infty,t} Z_t' & M_{*,t} &= P_{*,t} Z_t' \\
 F_t^{(1)} &= F_{\infty,t}^{-1} & F_t^{(2)} &= -F_{\infty,t}^{-1} F_{*,t} F_{\infty,t}^{-1} \\
 K_t^{(0)} &= T_t M_{\infty,t} F_t^{(1)} & K_t^{(1)} &= T_t M_{\infty,t} F_t^{(2)} + T_t M_{*,t} F_t^{(1)} \\
 L_t^{(0)} &= T_t - K_t^{(0)} Z_t & L_t^{(1)} &= -K_t^{(1)} Z_t \\
 v_t^{(0)} &= Y_t - Z_t a_t^{(0)} - d_t & a_{t+1}^{(0)} &= T_t a_t^{(0)} + K_t^{(0)} v_t^{(0)} + c_t \\
 P_{\infty,t+1} &= T_t P_{\infty,t} L_t^{(0)'} & P_{*,t+1} &= T_t P_{\infty,t} L_t^{(1)'} + T_t P_{*,t} L_t^{(0)'} + R_t Q_t R_t'
 \end{aligned} \tag{3.9}$$

nas quais

$$\begin{aligned}
 a_1^{(0)} &= a \\
 P_{\infty,1} &= AA' \\
 P_{*,1} &= R_0 Q_0 R_0'
 \end{aligned} \tag{3.10}$$

Segundo Koopman (1997), $P_{\infty,t}$ se anula num determinado instante implicando assim a anulação de $P_{\infty,t+j}$ e $F_{\infty,t+j}$ para $j = 1, 2, \dots$. O momento imediatamente anterior ao que $P_{\infty,t}$ se anula é conhecido na literatura como d . Assim, a partir de $d + 1$, o filtro de Kalman inicial exato converte-se num conjunto de recursões coincidentes com o filtro de Kalman “2 em 1” já conhecido, no qual $P_t = P_{*,t}$ e $a_t = a_t^{(0)}$ para os instantes posteriores a d . Dessa forma, pode-se extrair o instante de estabilização d a partir da implementação do filtro de Kalman, o que confere uma vantagem significativa em comparação à inicialização por aproximação.

⁶ Aqui, mencionamos as recursões quando $F_{\infty,t} > 0$ nos instantes difusos. Nos casos específicos em que $F_{\infty,t} = 0$ em pelo menos um destes instantes, as equações dadas em (3.9) sofrem algumas modificações. Como não há nesta dissertação nenhuma modelagem que se enquadre neste caso específico, optou-se por não mencioná-las. O mesmo argumento subsiste no contexto de suavização. Detalhes sobre este caso podem ser encontrados em Koopman e Durbin (2003).

De forma análoga, o caráter eventual de difusibilidade do vetor de estado também provoca alterações significativas nas recursões do suavizador de Kalman de *de Jong* (SKJ). Abaixo, são apresentadas as recursões no contexto de inicialização exata, dadas por:

Suavizador de Kalman: Inicialização Exata

$$\begin{aligned}
r_{t-1}^{(0)} &= L_t^{(0)'} r_t^{(0)} \\
r_{t-1}^{(1)} &= Z_t' F_t^{(1)} v_t^{(0)} + L_t^{(0)'} r_t^{(1)} + L_t^{(1)'} r_t^{(0)} \\
N_{t-1}^{(0)} &= L_t^{(0)'} N_t^{(0)} L_t^{(0)} \\
N_{t-1}^{(1)} &= Z_t' F_t^{(1)} Z_t + L_t^{(0)'} N_t^{(1)} L_t^{(0)} + L_t^{(1)'} N_t^{(0)} L_t^{(0)} + L_t^{(0)'} N_t^{(0)} L_t^{(1)} \\
N_{t-1}^{(2)} &= Z_t' F_t^{(2)} Z_t + L_t^{(0)'} N_t^{(2)} L_t^{(0)} + L_t^{(0)'} N_t^{(1)} L_t^{(1)} + L_t^{(1)'} N_t^{(0)} L_t^{(1)} + L_t^{(1)'} N_t^{(1)} L_t^{(0)} \\
\hat{\alpha}_t &= a_t^{(0)} + P_{*,t} r_{t-1}^{(0)} + P_{\infty,t} r_{t-1}^{(1)} \\
V_t &= P_{*,t} - P_{*,t} N_{t-1}^{(0)} P_{*,t} - P_{*,t} N_{t-1}^{(1)} P_{\infty,t} - P_{\infty,t} N_{t-1}^{(1)} P_{*,t} - P_{\infty,t} N_{t-1}^{(2)} P_{\infty,t}
\end{aligned} \tag{3.11}$$

onde

$$\begin{aligned}
r_d^{(0)} &= r_d \\
r_d^{(1)} &= 0 \\
N_d^{(0)} &= N_d \\
N_d^{(1)} &= N_d^{(2)} = 0
\end{aligned} \tag{3.12}$$

Segundo a abordagem teórica do suavizador de Kalman, torna-se necessário implementar estas recursões somente no período compreendido entre o instante d e o instante 1. Para os instantes de tempo remanescentes, as recursões do suavizador no contexto de inicialização exata coincidem plenamente com os apresentados para o SKJ anteriormente em (3.5). Para um estudo mais aprofundado acerca do FKIE e do SKIE, conferir Koopman (1997), Durbin e Koopman (2001), capítulo 5 e Koopman e Durbin (2003).

Dentre outros atrativos proporcionados na adoção de inicialização exata para o filtro de Kalman, destacam-se:

- Por se tratar de uma solução exata, baseada em uma metodologia bem estruturada, não proporciona erro de aproximação e, em especial, a sua propagação ao longo da implementação;

- Promove a constituição apropriada de recursões para o suavizador de Kalman, bem a construção de funções de verossimilhança *exatas* para a estimação de parâmetros desconhecidos no contexto de normalidade, inclusive no período anterior ao instante d , incluindo-o. Isso será mostrado na seção 3.4.

3.3

Imposição de Restrições ao Filtro de Kalman

Em determinadas situações, justifica-se a execução do filtro de Kalman com o vetor de estado obedecendo a determinadas restrições. Um exemplo pode ser encontrado na implementação do filtro de Kalman aplicada à análise de estilo sob os contextos semi-forte e forte, os quais apresentam restrições de carteira e/ou de assumir posição vendida em qualquer das classes de ativos (vide capítulo 4 para maiores detalhes).

Uma vez que o enfoque desta dissertação *não* está centrado na aplicação de análise de estilo forte, o que pressupõe o estabelecimento de restrições *não-lineares*, abordaremos tão somente duas metodologias indicadas para imposição de restrições *lineares de igualdade* ao filtro de Kalman: o *filtro de Kalman restrito aumentado* e o *filtro de Kalman restrito reduzido*. Essas metodologias se aplicam mais à análise de estilo semi-forte, que será largamente empregada nesta dissertação.

Para o que segue, devemos considerar a hipótese de que o vetor de estado α_t satisfaz restrições lineares dadas por

$$A_t \alpha_t = q_t \quad (3.13)$$

nas quais A_t é matriz $k \times m$ determinística e conhecida e q_t é vetor $k \times 1$ observável e possivelmente aleatório.

Além disso, as seguintes observações mostram-se pertinentes:

- Restrições afins do tipo $A_t \alpha_t + b_t = q_t$, onde b_t assume um caráter exógeno, podem também ser implementadas, bastando que seja redefinido o vetor q_t para q_t^* , onde $q_t^* = q_t - b_t$.
- O número de restrições podem ser variantes no tempo. Nesse caso específico, sugere-se, para um dado instante de tempo t , zerar apropriadamente as linhas de A_t e de q_t a fim de evidenciar corretamente as reais restrições do modelo. Uma outra forma de lidar adequadamente com esta questão seria tratar A_t e q_t com dimensões k_t variantes no tempo.
- Nos casos em que $k > m$ (i.e. o número de restrições é *maior* que o número de coordenadas do vetor de estado) ou as linhas de A_t são *linearmente dependentes*, evidencia-se a presença de restrições redundantes, as quais podem ser eliminadas para que se dê a correta implementação do filtro de Kalman sob restrições.

3.3.1

Filtro de Kalman Restrito Aumentado

A metodologia do *filtro de Kalman restrito aumentado* (daqui por diante, *filtro aumentado*) caracteriza-se por expandir a equação das observações para que sejam incorporadas as restrições dadas na forma $A_t \alpha_t = q_t$, apresentada em (3.13).

Nesse sentido, basta que substituamos na equação das observações original apresentada em (2.1) as seguintes variáveis: Y_t por $Y_t^* = (Y_t', q_t)'$, Z_t por $Z_t^* = (Z_t', A_t)'$, d_t por $d_t^* = (d_t', 0)'$ e ε_t por $\varepsilon_t^* = (\varepsilon_t', 0)'$, o que resulta, por sua vez, na substituição de H_t por $H_t^* = \text{diag}(H_t', 0)'$, onde as *matrizes nulas* correspondentes à d_t^* , ε_t^* e H_t^* possuem, respectivamente, dimensões equivalentes a $1 \times k$, $1 \times k$ e $k \times k$. Mantendo a equação do estado original inalterada, como em (2.2), aplicamos então o filtro de Kalman usual sob os enfoques de atualização e suavização.

Como o filtro aumentado apresenta notáveis desvantagens em comparação à metodologia do filtro de Kalman restrito reduzido, a serem discutidas na próxima seção, optamos por *não* utilizá-la nesta dissertação, justificando assim a pouca ênfase dada à mesma nesta seção. Para maiores detalhes acerca do filtro aumentado, incluindo algumas provas técnicas, vide Pizzinga (2008).

3.3.2

Filtro de Kalman Restrito Reduzido

Uma alternativa mais atraente à incorporação de restrições ao filtro de Kalman se encontra na metodologia do *filtro de Kalman restrito reduzido* (daqui por diante, *filtro reduzido*), o qual consiste basicamente em “reescrever algumas coordenadas do vetor de estado em função de outras” (cf. Marques, 2009), resultando na *redução* do tamanho do vetor de estado e, por conseguinte, na *redução* da dimensão da equação do estado, daí o seu nome. O resultado assim obtido é então substituído na equação das observações original que, após algumas manipulações algébricas e redefinições de matrizes, adquire uma nova expressão, mais sintética, a qual será utilizada na aplicação do filtro de Kalman usual para a obtenção das estimações desejadas das coordenadas do vetor de estado.

O algoritmo a seguir exhibe o passo-a-passo que se requer para construirmos o modelo reduzido. Antes, porém, devemos considerar a hipótese adicional e crucial de que q_t é ζ_t -mensurável, ou seja, q_t é função afim de $(Y_1', Y_2', \dots, Y_t)'$.

Cumpramos mencionar que, neste algoritmo, foram suprimidos os vetores d_t e c_t , tendo em vista facilitar a notação. Estes termos, entretanto, podem ser perfeitamente incluídos, uma vez que a modelagem necessite dessa intervenção. Ademais, o algoritmo é essencialmente o mesmo que foi apresentado em Marques (2009) e Pizzinga (2008), embora com uma notação diferente.

Posto isso, segue abaixo o algoritmo, estruturado nas seguintes etapas:

Seja t um instante de tempo arbitrário.

1. Reescreva a restrição linear dada em (3.13), de forma a incorporar o vetor de estado particionado, sendo a matriz $A_{1,t}$ de dimensão $k \times k$ e invertível:

$$A_{1,t}\alpha_{1,t} + A_{2,t}\alpha_{2,t} = q_t \quad (3.14)$$

2. Resolva a equação (3.14) para $\alpha_{1,t}$, obtendo a seguinte expressão:

$$\alpha_{1,t} = A_{1,t}^{-1}q_t - A_{1,t}^{-1}A_{2,t}\alpha_{2,t} \quad (3.15)$$

3. Substitua, na equação das observações original dada em (2.1), a expressão de $\alpha_{1,t}$ deduzida na equação anterior:

$$\begin{aligned} Y_t &= Z_{1,t}\alpha_{1,t} + Z_{2,t}\alpha_{2,t} + \varepsilon_t \\ Y_t &= Z_{1,t}(A_{1,t}^{-1}q_t - A_{1,t}^{-1}A_{2,t}\alpha_{2,t}) + Z_{2,t}\alpha_{2,t} + \varepsilon_t \\ Y_t &= Z_{1,t}A_{1,t}^{-1}q_t - Z_{1,t}A_{1,t}^{-1}A_{2,t}\alpha_{2,t} + Z_{2,t}\alpha_{2,t} + \varepsilon_t \\ Y_t^* &\equiv Y_t - Z_{1,t}A_{1,t}^{-1}q_t = (Z_{2,t} - Z_{1,t}A_{1,t}^{-1}A_{2,t})\alpha_{2,t} + \varepsilon_t \\ Y_t^* &\equiv Z_{2,t}^*\alpha_{2,t} + \varepsilon_t \end{aligned} \quad (3.16)$$

4. Proponha uma equação do estado para $\alpha_{2,t}$ de tal forma a preservar um modelo em espaço de estado linear e obtenha o seguinte modelo em espaço de estado linear **reduzido**:

$$\begin{aligned} Y_t^* &= Z_{2,t}^*\alpha_{2,t} + \varepsilon_t, \quad \varepsilon_t \sim (0, H_t) \\ \alpha_{2,t+1} &= T_{2,t}\alpha_{2,t} + R_{2,t}\eta_{2,t}, \quad \eta_{2,t} \sim (0, Q_{2,t}) \\ \alpha_{2,1} &\sim (a_{2,1}, P_{2,1}) \end{aligned} \quad (3.17)$$

5. Para o modelo definido em (3.17), aplique o filtro de Kalman usual sob os contextos de atualização e suavização (i.e. nos casos em que $t \leq j$), obtendo assim $a_{2,t/j}$.

6. De posse do resultado obtido no passo 5, calcule $a_{1,t/j}$ para todo $t \leq j$:

$$a_{1,t/j} = A_{1,t}^{-1}q_t - A_{1,t}^{-1}A_{2,t}a_{2,t/j} \quad (3.18)$$

$$P_{1,t/j} = (A_{1,t}^{-1}A_{2,t})P_{2,t/j}(A_{1,t}^{-1}A_{2,t})' \quad (3.19)$$

O filtro reduzido possui algumas vantagens consideráveis em relação ao filtro aumentado, dentre as quais citamos:

- Pelo fato de a especificação da equação do estado só se dar no 4º passo do algoritmo, após o processo de redução, evita-se incorrer em quaisquer riscos de inconsistência matemática que poderiam ocorrer em caso contrário. Nesse sentido, denota-se uma evidente vantagem metodológica do filtro reduzido em confronto com o filtro aumentado.
- Ao longo do processo de modelagem, podem-se confrontar diversos modelos concorrentes mediante os critérios de informação, como AIC e BIC, entre outros. Temos então, uma vez mais, uma vantagem metodológica do filtro reduzido.
- Em termos operacionais, tem-se uma notória vantagem do filtro reduzido, no sentido de que este produz uma ***redução da dimensão da equação do estado***, enquanto que o filtro aumentado proporciona uma ***ampliação da dimensão da equação das medidas***.

Para maiores informações sobre o processo de imposição de restrições lineares em modelos em espaço de estado lineares, vide Pizzinga (2008).

3.4

Estimação de Parâmetros por (Quasi) Máxima Verossimilhança

As matrizes do sistema, a saber Z_t , d_t , T_t , c_t , R_t , H_t e Q_t , constituem papel essencial na implementação do filtro de Kalman. Essas matrizes seguem em geral um processo determinístico e conhecido ao longo do tempo, eventualmente (ou quase sempre) possuindo parâmetros desconhecidos a serem estimados, usualmente via ***(quasi) máxima verossimilhança***.

Por outro lado, para que estes parâmetros possam ser estimados por este método, torna-se imprescindível a construção da função de ***(quasi) verossimilhança***, a qual leva em conta as recursões do filtro de Kalman na sua

formulação. Nesse sentido, depreende-se que a estimação do vetor de estado, bem como dos parâmetros desconhecidos, se dão *simultaneamente*.

A construção da função de (*quasi*) verossimilhança requer a informação antecipada de qual metodologia de inicialização do filtro pretende-se empregar (inicialização não-difusa, inicialização difusa aproximada ou inicialização difusa exata), para os quais corresponde uma função de (*quasi*) verossimilhança diferente, apropriada a cada caso, o que será visto posteriormente.

Para as próximas seções, referentes à estimação por (*quasi*) máxima verossimilhança, será admitida a hipótese dos modelos em espaço de estado serem Gaussianos. Denote por ψ o vetor de todos os parâmetros desconhecidos, distribuídos ao longo de todas as matrizes do sistema, e seu correspondente espaço paramétrico por Θ .

A partir de uma série temporal observada y_1, y_2, \dots, y_n e avaliada em ψ , temos que a função de verossimilhança, em termos gerais, é dada por:

$$L(\psi) \equiv p(y_1, y_2, \dots, y_n) = p(y_1) \prod_{t=2}^n p(y_t / \zeta_{t-1}) = \prod_{t=1}^n p(y_t / \zeta_{t-1}) \quad (3.20)$$

na qual $\zeta_0 = \{\emptyset, \Omega\}$.

O *estimador de máxima verossimilhança*, por sua vez, consiste em avaliar o $\hat{\psi}$ que maximize essa função. Assim, definimos $\hat{\psi}$ da seguinte forma:

$$\hat{\psi} = \arg \max_{\psi \in \Theta} L(\psi) = \arg \max_{\psi \in \Theta} \log L(\psi) \quad (3.21)$$

Como usualmente as funções de log-verossimilhança são pouco tratáveis e não-lineares, métodos numéricos, como os do tipo *quasi Newton*, são imprescindíveis no processo de maximização. Nesse sentido, apresentamos o seguinte algoritmo, que pode ser de grande valia no processo de otimização:

1. Estabeleça um valor inicial arbitrário, mas apropriado para ψ , digamos $\psi^{(0)}$.

2. A partir das matrizes do sistema construídas com base em $\psi^{(0)}$, implemente o filtro de Kalman apropriado para a metodologia de estimação por máxima verossimilhança empregada (i.e. filtro de Kalman inicial exato para função de verossimilhança difusa exata, etc).
3. Com as “saídas” obtidas da implementação do filtro de Kalman, avalie a função de log-verossimilhança em $\psi^{(0)}$.
4. Com o otimizador de escolha, realize uma iteração com $\log L(\psi^{(0)})$ e, desse modo, obtenha $\psi^{(1)}$.
5. Utilize **critérios de convergência** para avaliar a proximidade entre $\psi^{(0)}$ e $\psi^{(1)}$. Caso não seja atingida a proximidade desejada, repetir todo o algoritmo, sendo $\psi^{(1)}$ o novo valor inicial para desencadear o algoritmo.

Para modelos lineares não-Gaussianos, pode-se também utilizar as funções de verossimilhança que aqui serão apresentadas em detrimento das exatas, as quais se caracterizam por serem muito difíceis de serem obtidas.

Neste contexto específico, as funções de verossimilhança são usadas na qualidade de **funções de quasi verossimilhança** e, de forma similar, os estimadores de máxima verossimilhança são denominados **estimadores de quasi máxima verossimilhança**. No entanto, deve-se atentar para os cuidados adicionais requeridos no tocante à inferência estatística de ψ .

3.4.1

Estimação por Máxima Verossimilhança Não-Difusa

Nos casos específicos em que **não** se verifica a presença de elementos difusos em α_1 , a função de verossimilhança, conhecida neste contexto como **verossimilhança decomposta pelo erro de predição**, é dada em logaritmo por:

$$\log L(\psi) = -\frac{np}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n [\log |F_t| + v_t' F_t^{-1} v_t], \quad \forall \psi \in \Theta \quad (3.22)$$

na qual $v_t = v_t(\psi)$ e $F_t = F_t(\psi)$.

Uma vez estabelecida a função de log-verossimilhança no contexto de não-difusibilidade, pode-se implementar o algoritmo anunciado na seção anterior, empregando as recursões do filtro de Kalman usuais, neste caso específico.

Maiores detalhes acerca desta abordagem teórica de estimação, podem ser obtidas em Harvey (1989) e em Durbin e Koopman (2001).

3.4.2

Estimação por Máxima Verossimilhança Difusa

Nos casos em que ao menos uma das coordenadas de α_1 é tratada de forma difusa, o filtro de Kalman demanda um procedimento de inicialização, que pode se dar de forma *aproximada* ou *exata*, conforme discutido na seção 3.2 referente à inicialização do filtro de Kalman.

De forma similar, devem-se construir funções de log-verossimilhança apropriadas para cada procedimento de inicialização. Em se tratando de inicialização aproximada do filtro de Kalman, isto é, quando κ é pré-fixado em um valor arbitrariamente grande, adota-se uma versão *condicional* de (3.22), onde são suprimidas as d informações iniciais que caracterizam o período difuso do modelo.

A escolha deste d , no entanto, não possui qualquer embasamento teórico, se constituindo simplesmente numa escolha arbitrária. A função log-verossimilhança no contexto de difusibilidade aproximada está estruturada da seguinte forma:

$$\log L(\psi)^{cond.} = -\frac{(n-d)p}{2} \log(2\pi) - \frac{1}{2} \sum_{t=d+1}^n [\log|F_t| + v_t' F_t^{-1} v_t], \quad \forall \psi \in \Theta \quad (3.23)$$

onde $v_t = v_t(\psi)$ e $F_t = F_t(\psi)$.

Por outro lado, sob o enfoque de inicialização difusa exata, isto é, quando manipulamos κ de forma que ele tenda a infinito, temos que a função de log-

verossimilhança dada em (3.22) sofre algumas alterações nos instantes iniciais correspondentes ao período difuso, se apresentando da seguinte forma⁷:

$$\log L(\psi) = -\frac{np}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^d \log |F_{\infty,t}| - \frac{1}{2} \sum_{t=d+1}^n [\log |F_t| + v_t' F_t^{-1} v_t], \quad \forall \psi \in \Theta \quad (3.24)$$

onde $v_t = v_t(\psi)$, $F_{\infty,t} = F_{\infty,t}(\psi)$ e $F_t = F_t(\psi)$.

Similarmente ao que se dá no contexto de inicialização do filtro de Kalman, os mesmos itens que configuram uma notória desvantagem ao filtro de Kalman sob inicialização aproximada frente ao filtro de Kalman inicial exato reaparecem no contexto de otimização da função de log-verossimilhança.

Assim sendo, sob a abordagem “big κ ”, a escolha arbitrária de κ pode produzir aumento de viés em amostras finitas de $\hat{\psi}$ (com efeito: d pode não ter sido escolhido adequadamente, de forma que a_{d+1} e P_{d+1} podem não se constituir em boas condições iniciais).

Um agravante adicional à utilização da log-verossimilhança difusa aproximada diz respeito à eliminação das d informações iniciais que caracterizam o período difuso. Agindo dessa forma, pode-se incorrer em sérias dificuldades na otimização de ψ , em especial quando se trata de séries pequenas e/ou quando há muitos parâmetros a serem estimados em ψ .

3.5

Diagnósticos e Seleção de Modelos

Uma análise de diagnósticos consiste basicamente na análise gráfica e/ou inferencial dos resíduos e pelo poder preditivo do modelo ajustado, de forma que os dados estimados reproduzam relativamente bem os dados.

Quanto à análise de resíduos, deve-se verificar se as inovações padronizadas, dadas por $v_t^p \equiv (F_t^{1/2})^{-1} v_t$, nas quais $F_t = (F_t^{1/2})(F_t^{1/2})$ com $t =$

⁷ Nos casos específicos onde $F_{\infty,t} = 0$ nos instantes difusos, a função dada em (3.24) apresenta algumas modificações. Como não há nesta dissertação nenhuma modelagem que se enquadre neste caso específico, optou-se por não mencioná-la. Detalhes sobre este caso podem ser encontrados em Durbin e Koopman (2001), capítulo 7.

$1, \dots, n$, constituem vetores aleatórios e identicamente distribuídos, com média constante e aproximadamente nula; homocedasticidade incondicional e de variância em torno de 1; ausência de correlação serial e, possivelmente, com distribuição normal, embora isso não seja imprescindível.

Nesse sentido, as ferramentas usuais de averiguação dos resíduos consistem em procedimentos gráficos e descritivos, como *plot* das inovações padronizadas no tempo, função de autocorrelação (FAC) e função de autocorrelação parcial (FACP), Q-Q plot com base na distribuição normal, histograma, média e variância amostrais, etc. Além disso, com o intuito de complementar a análise gráfica e descritiva, usualmente são implementados testes de hipóteses de especificações, de forma que sejam checadas os pressupostos em questão. Para este caso, ferramentas usuais consistem em testes de autocorrelação serial de Ljung-Box aplicado sobre as inovações padronizadas (e também sobre o quadrado das inovações padronizadas a fim de averiguar o comportamento heterocedástico dos resíduos), teste de Anderson-Darling e/ou Jarque-Bera para averiguar normalidade dos resíduos, entre outros.

Cumprе salientar que mesmo não havendo evidências de normalidade nos resíduos, as funções de verossimilhança ainda assim são utilizadas na qualidade de funções de *quasi* verossimilhança e, de forma similar, os estimadores de máxima verossimilhança são utilizados como estimadores de *quasi* máxima verossimilhança. No entanto, deve-se atentar para os cuidados adicionais requeridos no tocante à inferência estatística.

Quanto à análise do poder preditivo do modelo ajustado, é recomendada a averiguação através de cálculo de medidas descritivas, como o *pseudo-R*² e o Erro Quadrático Médio (EQM) apresentado na seqüência:

$$\begin{aligned}
 Pseudo - R^2 &= [corr(Y_t, \hat{Y}_{t/t-1})]^2 \\
 EQM &= \frac{1}{n-d} \sum_{t=d+1}^n (Y_t - \hat{Y}_{t/t-1})^2
 \end{aligned} \tag{3.25}$$

No tocante à seleção de modelos, o uso de critérios de informação como AIC e BIC são usualmente utilizados, bem como os valores observados para a log-verossimilhança calculada. Conforme Durbin e Koopman (2001), os critérios

de informação sob o contexto de espaço de estado podem ser expressos da seguinte forma:

$$\begin{aligned} AIC &= \frac{1}{n}[-2 \log L(\hat{\psi}) + 2(q + w)] \\ BIC &= \frac{1}{n}[-2 \log L(\hat{\psi}) + (q + w) \log n] \end{aligned} \tag{3.26}$$

nos quais $w = \dim(\psi)$ e $q = \dim(\delta)$, vetores enunciados nas seções anteriores.