

3

Modelo na forma espaço-estado e o Filtro de Kalman

Este capítulo apresenta os modelos lineares gaussianos na forma espaço-estado sob a perspectiva clássica de estimação por máxima verossimilhança e sua aplicação a modelos de regressão com coeficientes variantes no tempo.

A primeira seção, 3.1, apresenta a forma geral de um modelo na forma espaço-estado. A seção 3.2 apresenta o filtro de Kalman. A seção 3.3 discute as questões relacionadas à inicialização do filtro. A seção 3.4 explica o processo de estimação dos hiperparâmetros do modelo geral através de máxima verossimilhança, apresentando a construção da função de verossimilhança e o processo de otimização envolvido. A seção 3.5 apresenta os modelos de regressão com coeficientes variáveis no tempo sob a abordagem da modelagem na forma espaço-estado, de especial interesse nesse trabalho para o estudo de modelos de fatores condicionais. Esta modelagem tem sido bastante utilizada na literatura sobre modelos condicionais envolvendo estimação utilizando filtro de Kalman (Adrian e Franzoni, 2009; Bentz, 2003; Faff, Hillier e Hillier, 2000; Mergner e Bulla, 2008; Mergner, 2009). Finalmente, a seção 3.6 apresenta informações sobre análise de diagnóstico dos modelos na forma espaço-estado. De forma geral, Durbin e Koopman (2001) é a principal referência para este capítulo.

3.1

Modelos na forma espaço-estado

A modelagem espaço-estado possibilita descrever um vasto conjunto de problemas na análise de séries temporais, incluindo modelos lineares e não lineares (Harvey, 1989; Durbin e Koopman, 2001). Modelos na forma espaço-estado são descritos por duas equações: a equação de observação e a equação de estado. Neste sentido, o desenvolvimento do sistema em estudo é descrito por uma série de vetores não observados, compostos pelas chamadas variáveis de estado, relacionados a uma série de variáveis observadas. A equação de estado descreve a dinâmica as variáveis de estado, enquanto a equação de observação associa as variáveis observadas ao vetor de estado.

Seja \mathbf{y}_t um vetor multivariado $p \times 1$ de observações de uma série temporal, cujo desenvolvimento no tempo pode ser caracterizado em termos de um vetor de

estado não observado $\boldsymbol{\mu}_t$ composto por m variáveis de estado, ou seja, de dimensão $m \times 1$, para cada instante de tempo t . Um modelo linear gaussiano na forma espaço-estado pode ser escrito como:

$$\mathbf{y}_t = \mathbf{S}_t \boldsymbol{\mu}_t + \mathbf{d}_t + \boldsymbol{\epsilon}_t \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{H}_t) \quad (3.1)$$

$$\boldsymbol{\mu}_{t+1} = \mathbf{T}_t \boldsymbol{\mu}_t + \mathbf{c}_t + \mathbf{U}_t \boldsymbol{\eta}_t \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}_t) \quad t = 1 \text{ a } N \quad (3.2)$$

onde $E[\boldsymbol{\epsilon}_t \boldsymbol{\eta}_s'] = 0$ para $t = 1 \text{ a } N$; $E[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_s'] = 0$ para todo $t \neq s$; $E[\boldsymbol{\eta}_t \boldsymbol{\eta}_s'] = 0$ para todo $t \neq s$; $\boldsymbol{\mu}_1 \sim N(\mathbf{a}_1, \mathbf{P}_1)$; $E[\boldsymbol{\eta}_t' \boldsymbol{\mu}_1] = E[\boldsymbol{\epsilon}_t' \boldsymbol{\mu}_1] = 0$ para $t = 1 \text{ a } N$

As equações (3.1) e (3.2) são as chamadas equação de observação e equação de estado, respectivamente. As matrizes \mathbf{S}_t , \mathbf{T}_t , \mathbf{d}_t , \mathbf{c}_t , \mathbf{U}_t , \mathbf{H}_t e \mathbf{Q}_t são chamadas matrizes do sistema e assume-se que são não estocásticas, ou seja, podem variar no tempo de forma conhecida. Há que se considerar que alguns elementos nestas matrizes dependem de um vetor de parâmetros desconhecidos, chamados de hiperparâmetros, que podem ser estimados por máxima verossimilhança como apresentado mais a frente na seção 3.4. Considera-se ainda que os termos de erro $\boldsymbol{\epsilon}_t$ e $\boldsymbol{\eta}_t$ são serialmente independentes e independentes um do outro em todo instante de tempo. Assume-se que o vetor de estado inicial $\boldsymbol{\mu}_1$ possui distribuição $N(\mathbf{a}_1, \mathbf{P}_1)$ e que independe dos termos de erro $\boldsymbol{\epsilon}_t$ e $\boldsymbol{\eta}_t$ para qualquer instante de tempo. De forma geral, as dimensões dos elementos envolvidos no sistema de equações (3.1)-(3.2) são:

Tabela 3.1 – Dimensão de vetores e matrizes do modelo das eqs. (3.1)-(3.2)

Vetores		Matrizes	
\mathbf{y}_t	$p \times 1$	\mathbf{S}_t	$p \times m$
$\boldsymbol{\mu}_t$	$m \times 1$	\mathbf{T}_t	$m \times m$
\mathbf{d}_t	$p \times 1$	\mathbf{H}_t	$p \times p$
\mathbf{c}_t	$m \times 1$	\mathbf{Q}_t	$r \times r$
$\boldsymbol{\epsilon}_t$	$p \times 1$	\mathbf{U}_t	$m \times r$
$\boldsymbol{\eta}_t$	$r \times 1$		
\mathbf{a}_1	$m \times 1$	\mathbf{P}_1	$m \times m$

Assumindo por ora que os elementos das matrizes do sistema são conhecidos, podem ser derivadas as equações para o filtro de Kalman.

3.2 Filtro de Kalman

Considerando os modelos estruturais na forma espaço-estado apresentados na seção 3.1, o filtro de Kalman, através de um algoritmo recursivo, permite a estimação da variável não observável, denominada variável de estado, a partir da série temporal da variável observável. No decorrer do desenvolvimento histórico da teoria dos modelos em espaço-estado, ficou convencionado que a estimação do vetor de estado de um determinado modelo pode ser caracterizada em três categorias, dependendo do tipo de informação disponível da variável observável que estará sendo utilizado (Pizzinga, 2004). Considerando a estimação da variável de estado $\boldsymbol{\mu}_t$ a partir de informações disponíveis em um dado instante de tempo j , define-se que: se $j < t$, tem-se um problema de previsão ou predição; se $j = t$, tem-se um problema de filtragem ou atualização; e se $j > t$, tem-se um problema de suavização ou interpolação.

3.2.1 Equações de previsão do filtro de Kalman

A partir do modelo escrito na forma espaço-estado, o filtro de Kalman é usado para computar as previsões ótimas para a média e a variância do vetor de estado $\boldsymbol{\mu}_{t+1}$, de forma recursiva, a cada nova observação \mathbf{y}_t . Considerando o sistema dados pelas equações (3.1)-(3.2), o filtro de Kalman pode ser derivado sob a premissa de que o vetor de estado inicial $\boldsymbol{\mu}_1 \sim N(\mathbf{a}_1, \mathbf{P}_1)$ é conhecido, ou seja, \mathbf{a}_1 e \mathbf{P}_1 conhecidos. O objetivo é a atualização do nosso conhecimento acerca do vetor de estado a cada nova observação disponível no tempo t . Assim, deseja-se obter a distribuição condicional do vetor de estado $\boldsymbol{\mu}_{t+1}$ para $t = 1$ a N , com base em \mathbf{Y}_t , o conjunto de observações até o tempo t , ou seja, $\mathbf{Y}_t = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t\}$.

Tendo em vista que todas as distribuições consideradas no sistema são normais, as distribuições condicionais de subconjuntos de variáveis dados outros subconjuntos de variáveis também são normais. Desta forma, a distribuição condicional de $\boldsymbol{\mu}_{t+1}$ pode ser determinada pela sua média condicional e sua variância condicional. Sejam:

$$\mathbf{a}_{t+1|t} = E[\boldsymbol{\mu}_{t+1} | \mathbf{Y}_t] \quad (3.3)$$

$$\mathbf{P}_{t+1|t} = \text{Var}[\boldsymbol{\mu}_{t+1} | \mathbf{Y}_t] \quad (3.4)$$

a média condicional e a variância condicional de $\boldsymbol{\mu}_{t+1}$ dado o conjunto de informação \mathbf{Y}_t . Para simplificar, usaremos a notação $\mathbf{a}_{t+1} = \mathbf{a}_{t+1|t}$ e $\mathbf{P}_{t+1} = \mathbf{P}_{t+1|t}$. Considerando que $\boldsymbol{\mu}_t$ dado o conjunto de informação \mathbf{Y}_{t-1} tem distribuição $N(\mathbf{a}_t, \mathbf{P}_t)$, pode-se mostrar que \mathbf{a}_{t+1} e \mathbf{P}_{t+1} podem ser calculados recursivamente através das equações do filtro de Kalman, combinando os passos de atualização e previsão (Durbin e Koopman, 2001):

$$\mathbf{a}_{t+1} = \mathbf{T}_t \mathbf{a}_t + \mathbf{c}_t + \mathbf{k}_t \mathbf{v}_t \quad (3.5)$$

$$\mathbf{P}_{t+1} = \mathbf{T}_t \mathbf{P}_t \mathbf{L}_t' + \mathbf{U}_t \mathbf{Q}_t \mathbf{U}_t' \quad (3.6)$$

onde

$$\mathbf{v}_t = \mathbf{y}_t - E[\mathbf{y}_t | \mathbf{Y}_{t-1}] = \mathbf{y}_t - \mathbf{S}_t \mathbf{a}_t$$

$$\mathbf{F}_t = \text{Var}[\mathbf{v}_t] = \mathbf{S}_t \mathbf{P}_t \mathbf{S}_t' + \mathbf{H}_t$$

$$\mathbf{k}_t = \mathbf{T}_t \mathbf{M}_t \mathbf{F}_t^{-1} \quad (3.7)$$

$$\mathbf{M}_t = \mathbf{P}_t \mathbf{S}_t'$$

$$\mathbf{L}_t = \mathbf{T}_t - \mathbf{k}_t \mathbf{S}_t$$

para $t = 1$ a N . O conjunto de equações (3.5)-(3.7) é chamado de filtro de Kalman do modelo dado pelas equações (3.1)-(3.2). A matriz \mathbf{k}_t é o ganho de Kalman e o vetor $\mathbf{v}_t = \mathbf{y}_t - E[\mathbf{y}_t | \mathbf{Y}_{t-1}]$ é o erro de previsão um-passo-a-frente de \mathbf{y}_t dado o conjunto de informação \mathbf{Y}_{t-1} , comumente chamado de inovação.

3.2.2

Equações de suavização do filtro de Kalman

O suavizador de estado permite basear a estimação do vetor de estado na amostra completa de observações de $t = 1$ a N . Seja o conjunto de informação $\mathbf{Y}_N = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$. Tendo em vista que todas as distribuições consideradas no sistema são normais, a distribuição condicional de $\boldsymbol{\mu}_t$ com base em \mathbf{Y}_N também será normal podendo ser determinada pela sua média condicional e sua variância condicional. Sejam:

$$\hat{\boldsymbol{\mu}}_t = E[\boldsymbol{\mu}_{t+1} | \mathbf{Y}_N] \quad (3.8)$$

$$\mathbf{V}_t = \text{Var}[\boldsymbol{\mu}_{t+1} | \mathbf{Y}_N] \quad (3.9)$$

o vetor de estado suavizado e a variância de estado suavizada. Considerando ainda que \mathbf{a}_1 e \mathbf{P}_1 são conhecidos, mostra-se que o vetor e a variância de estado suavizados podem ser obtidos através das seguintes equações recursivas *backwards*, ou seja, de $t = N$ a 1 :

$$\hat{\boldsymbol{\mu}}_t = \mathbf{a}_t + \mathbf{P}_t \mathbf{x}_{t-1} \quad (3.10) \quad \mathbf{V}_t = \mathbf{P}_t + \mathbf{P}_t \mathbf{W}_{t-1} \mathbf{P}_t \quad (3.11)$$

$$\mathbf{x}_{t-1} = \mathbf{S}'_t \mathbf{F}_t^{-1} \mathbf{v}_t + \mathbf{L}'_t \mathbf{x}_t \quad (3.12) \quad \mathbf{W}_{t-1} = \mathbf{S}'_t \mathbf{F}_t^{-1} \mathbf{S}_t + \mathbf{L}'_t \mathbf{W}_t \mathbf{L}_t \quad (3.13)$$

onde $\mathbf{x}_N = \mathbf{0}$ e $\mathbf{W}_N = \mathbf{0}$. As equações (3.10)-(3.13) são conhecidas como equações recursivas para estado suavizado.

3.3 Inicialização

Nas seções anteriores, os resultados partiram da premissa de que o vetor de estado inicial $\boldsymbol{\mu}_1 \sim N(\mathbf{a}_1, \mathbf{P}_1)$ era conhecido, ou seja, \mathbf{a}_1 e \mathbf{P}_1 conhecidos. Entretanto, na maior parte dos problemas práticos, ao menos alguns elementos de \mathbf{a}_1 e \mathbf{P}_1 não são conhecidos. Neste caso, há métodos para começar as séries tratando esta situação. Este procedimento é conhecido como inicialização e no caso em que há elementos não estacionários, trabalha-se com a chamada *inicialização difusa* do filtro. Considerando de forma abrangente o caso em que alguns elementos de $\boldsymbol{\mu}_1$ são difusos e outros não, um modelo geral para o vetor de estado inicial é dado por:

$$\boldsymbol{\mu}_1 = \mathbf{a} + \mathbf{A}\boldsymbol{\theta} + \mathbf{U}_0 \boldsymbol{\omega}_0 \quad \boldsymbol{\omega}_0 \sim N(\mathbf{0}, \mathbf{Q}_0) \quad (3.14)$$

onde \mathbf{a} é um vetor de dimensão $m \times 1$ conhecido, geralmente nulo; \mathbf{A} e \mathbf{U}_0 são matrizes de seleção de dimensão $m \times q$ e $m \times (m-q)$, respectivamente, com colunas correspondentes às da matriz identidade \mathbf{I}_m , de forma que \mathbf{A} seleciona as variáveis relacionadas às componentes não estacionárias do vetor de estado e \mathbf{U}_0 as componentes estacionárias; $\boldsymbol{\theta}$ é um vetor de dimensão $q \times 1$, de quantidades desconhecidas e estocásticas, ou de variáveis aleatórias normais com variância

infinita ($\theta \sim N(\mathbf{0}, \kappa \mathbf{I}_q)$ para $\kappa \rightarrow \infty$), chamado de *difuso*; ω_0 vetor aleatório tal que sua distribuição $N(\mathbf{n}_0, \mathbf{Q}_0)$, onde \mathbf{n}_0 e \mathbf{Q}_0 são a média e a variância incondicionais das variáveis estacionárias do vetor de estado. Inicializa-se então o filtro de Kalman com as condições iniciais:

$$\begin{aligned} \mathbf{a}_1 &= E[\boldsymbol{\mu}_1] = \mathbf{a} \\ \mathbf{P}_1 &= \text{Var}[\boldsymbol{\mu}_1] = \kappa \mathbf{P}_\infty + \mathbf{P}_* \\ \mathbf{P}_\infty &= \mathbf{A}\mathbf{A}' \quad \text{e} \quad \mathbf{P}_* = \mathbf{U}_0 \mathbf{Q}_0 \mathbf{U}_0' \end{aligned} \quad (3.15)$$

As componentes não-estacionárias do vetor de estado são chamadas de *difusas*. A *inicialização difusa* do filtro de Kalman pode envolver dois procedimentos. O primeiro é um procedimento aproximado (*inicialização difusa aproximada*), no qual o valor de κ é substituído por um número arbitrariamente muito grande de forma que são utilizadas as equações do filtro de Kalman padrão (equações (3.5)-(3.7)). Entretanto, esta abordagem apesar de útil para trabalhos exploratórios aproximados, não é recomendada para uso geral, uma vez que pode levar a grandes erros de arredondamento. A outra abordagem considera um tratamento exato do procedimento e é a chamada *inicialização difusa exata*. A técnica se baseia na expansão de produtos de matrizes com séries de potências em κ^{-1} , tomando apenas os dois ou três primeiros termos das séries e fazendo $\kappa \rightarrow \infty$ para obter o termo dominante (Durbin e Koopman, 2001).¹

3.4 Estimação por máxima verossimilhança

Para derivação do filtro de Kalman, assume-se como premissa que as matrizes do sistema são todas conhecidas. Como mencionado na seção 3.1, há que se considerar que alguns elementos nestas matrizes dependem de um vetor de parâmetros desconhecidos $\boldsymbol{\psi}$, chamados de hiperparâmetros, que podem ser estimados por máxima verossimilhança.

¹ O detalhamento das equações para o filtro de Kalman com inicialização exata pode ser obtido em Durbin e Koopman (2001, Capítulo 5).

3.4.1 Função de verossimilhança

Para que o modelo possa ser estimado por máxima verossimilhança, ele deve ser especificado de forma paramétrica pela função de densidade de probabilidade conjunta. Para o conjunto de N observações y_1, \dots, y_N , sob a premissa de que a distribuição do vetor de estado inicial $\boldsymbol{\mu}_1 \sim N(\mathbf{a}_1, \mathbf{P}_1)$ é conhecida, a função de verossimilhança é dada por:

$$L(\mathbf{y}, \boldsymbol{\psi}) = p(\mathbf{y}) = p(\mathbf{y}_1, \dots, \mathbf{y}_N) = \prod_{t=1}^N p(\mathbf{y}_t | \mathbf{Y}_{t-1}) \quad (3.16)$$

onde $p(\mathbf{y}_1 | \mathbf{y}_0) = p(\mathbf{y}_1)$ e $\mathbf{Y}_{t-1} = \{\mathbf{y}_1, \dots, \mathbf{y}_{t-1}\}$. Na prática, trabalha-se com a função logaritmo, de forma que a função de log-verossimilhança é dada por:

$$\log L(\mathbf{y}, \boldsymbol{\psi}) = \log L(\mathbf{y}, \boldsymbol{\psi}) = \sum_{t=1}^N \log p(\mathbf{y}_t | \mathbf{Y}_{t-1}) \quad (3.17)$$

Considerando o sistema dado pelas equações (3.1)-(3.2)², a distribuição condicional de \mathbf{y}_t é normal com média e variância dadas por

$$E[\mathbf{y}_t | \mathbf{Y}_{t-1}] = \mathbf{S}_t \mathbf{a}_t \quad (3.18)$$

$$\text{Var}[\mathbf{y}_t | \mathbf{Y}_{t-1}] = \mathbf{F}_t \quad (3.19)$$

onde \mathbf{F}_t é a variância do erro de previsão um passo a frente \mathbf{v}_t definida no conjunto de equações (3.7). Desta forma:

$$p(\mathbf{y}_t | \mathbf{Y}_{t-1}) \sim N(\mathbf{S}_t \mathbf{a}_t, \mathbf{F}_t) \quad (3.20)$$

A função densidade de probabilidade será dada por:

$$p(\mathbf{y}_t | \mathbf{Y}_{t-1}) = \frac{1}{(2\pi |\mathbf{F}_t|)^{1/2}} \exp \left[-\frac{1}{2} \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \right] \quad (3.21)$$

e, substituindo na equação (3.17), a função de log-verossimilhança será:

$$\log L(\mathbf{y}, \boldsymbol{\psi}) = -\frac{Np}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^N \log |\mathbf{F}_t| - \frac{1}{2} \sum_{t=1}^N \mathbf{v}_t' \mathbf{F}_t^{-1} \mathbf{v}_t \quad (3.22)$$

² Neste caso, considerando $\mathbf{d}_t = \mathbf{0}$, sem perda de generalidade.

Os hiperparâmetros do vetor $\boldsymbol{\psi}$ a serem estimados aparecem nas equações do filtro de Kalman para \mathbf{F}_t e \mathbf{v}_t . Na situação em que há componentes desconhecidas no vetor de estado inicial, pode-se derivar a função de log-verossimilhança para os casos de inicialização difusa aproximada e inicialização difusa exata, como detalhado em Durbin e Koopman (2001, Capítulo 5).

3.4.2 Otimização Numérica

Uma vez definida a função de verossimilhança, ela pode ser maximizada por métodos de otimização numérica. Na prática, estimam-se os hiperparâmetros $\hat{\boldsymbol{\psi}}$ do sistema que maximizam a função de log-verossimilhança.

Os algoritmos numéricos são utilizados de forma a comparar valores numéricos das funções de log-verossimilhança para diferentes conjuntos de valores de $\boldsymbol{\psi}$. Para calcular as estimativas dos valores da função de verossimilhança, o algoritmo parte de um determinado conjunto inicial de valores de $\boldsymbol{\psi}$, realiza uma série de passos, escolhendo em que direção seguir com a busca e o quanto mover nessa direção e calcula a cada iteração um novo valor para a função. Se um determinado conjunto de valores de $\boldsymbol{\psi}$ leva a valores próximos de máxima verossimilhança, o algoritmo para. Geralmente, os métodos de otimização diferem em relação à direção da busca, ao tamanho dos passos de iteração e à regra de parada (Mergner, 2009).

3.4.2.1 Método de Newton

Há uma grande diversidade de algoritmos numéricos de busca para maximização da log-verossimilhança, muitos deles baseados no método de Newton (Durbin e Koopman, 2001). No método de Newton, para um dado valor inicial de $\boldsymbol{\psi}$, a direção de busca é determinada pelo vetor gradiente $g(\boldsymbol{\psi})$ e o tamanho do passo pela matriz hessiana $H(\boldsymbol{\psi})$, de modo que o processo de busca pelo ponto ótimo é repetido até convergir ou até que se mude para um outro método de otimização. Na prática, o cálculo numérico do gradiente é geralmente factível, mas a hessiana é geralmente aproximada por diferentes métodos para evitar seu cálculo direto de forma analítica ou computacional. Um exemplo é o

método BFGS (Broyden-Fletcher-Goldfarb-Shannon), bastante utilizado em pacotes de programas computacionais para esta finalidade, através do qual a hessiana é obtida de forma recursiva. Detalhes sobre o método de Newton para otimização, em particular sobre o método BFGS, podem ser obtidos em Fletcher (1987).

3.4.2.2

Algoritmos genéticos

Uma alternativa para refinar o processo de otimização pode combinar o uso de um algoritmo de busca a partir do método de Newton com algoritmos genéticos. Os sistemas desenvolvidos a partir deste princípio são utilizados geralmente em problemas complexos ou com espaço de busca muito grande, por sua difícil modelagem e busca pela solução quando se aplicam métodos de otimização convencionais.

O uso da técnica de algoritmos genéticos consiste em um método de otimização inspirado nos conceitos da teoria de seleção natural, partindo de conceitos baseados nos processos genéticos para procurar soluções ótimas ou sub-ótimas. É utilizada uma analogia direta do fenômeno de evolução na natureza, onde cada indivíduo representa uma possível solução para um problema dado. Cada possível solução de um problema é codificada em uma estrutura chamada de "cromossomo", composta por uma cadeia de bits ou símbolos. Estes cromossomos representam indivíduos, que são evoluídos ao longo de várias gerações, de acordo com os princípios de seleção natural e sobrevivência. Os indivíduos são então submetidos a um processo evolucionário que envolve avaliação, seleção, recombinação, ou *crossover*, e mutação. A cada indivíduo atribui-se um valor de adaptação, que indica quanto a solução representada por este indivíduo é boa em relação às outras soluções da "população", ou seja, em relação ao conjunto de todas as soluções com as quais trabalha o sistema.

O processo de evolução começa com a criação aleatória dos indivíduos que formarão a população inicial. No caso prático utilizado neste trabalho, soluções iniciais também podem ser dadas por outros métodos de otimização, como o método de Newton, para que façam parte dessa população inicial. A partir de um processo de seleção baseado na aptidão de cada indivíduo, são escolhidos indivíduos para a fase de reprodução, que cria novas soluções utilizando-se para

isto um conjunto de operadores genéticos. Para determinar o final do processo, pode-se fixar o número de gerações ou de indivíduos criados; ou, ainda, condicionar à obtenção de alguma solução satisfatória, ao atingir um ponto ótimo. Detalhes sobre métodos de otimização baseados em algoritmos genéticos podem ser obtidos em Goldberg (1989), Koza (1992), Mitchell (1994) e Back (1996).

3.4.2.3 Restrições de valores dos parâmetros

Usualmente, os valores dos hiperparâmetros a serem estimados podem estar restritos a determinados intervalos. Por exemplo, parâmetros relativos a variâncias devem ser sempre positivos por definição. Entretanto, a introdução de restrições deste tipo em procedimentos numéricos pode ser inconveniente, sendo mais fácil realizar algumas transformações nos parâmetros de modo que as estimativas possam assumir qualquer valor no conjunto de números reais. Seja um parâmetro ψ na forma original em que aparece no modelo a ser estimado, restrito a determinados valores, e φ o valor correspondente a partir de uma transformação paramétrica de modo que $\varphi \in \mathbb{R}$. Dentre algumas restrições no espaço paramétrico mais utilizadas e suas transformações correspondentes, a Tabela 3.2 destaca algumas que são importantes no contexto deste trabalho.

Tabela 3.2 – Funções de reparametrização para otimização

Restrição	Transformação de ψ para φ	Transformação de φ para ψ
$\psi > 0$	$\varphi = \frac{1}{2} \ln \psi$	$\psi = e^{2\varphi}, \varphi \in \mathbb{R}$
$-1 < \psi < 1$	$\varphi = \frac{\psi}{\sqrt{1-\psi^2}}$	$\psi = \frac{\varphi}{\sqrt{1+\varphi^2}}, \varphi \in \mathbb{R}$
$0 < \psi < 1$	$\varphi = \ln\left(\frac{\psi}{1-\psi}\right)$	$\psi = \frac{1}{1+e^{-\varphi}}, \varphi \in \mathbb{R}$

3.5 Modelos de regressão com coeficientes variantes no tempo

De especial interesse neste trabalho são os modelos de regressão cujos coeficientes variam no tempo, abordados de forma resumida por Durbin e Koopman (2001). Seja o modelo univariado de regressão linear múltipla do tipo:

$$y_t = \mathbf{S}_t \boldsymbol{\mu} + \epsilon_t \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad t = 1 \text{ a } N \quad (3.23)$$

onde y_t é a série observada que se deseja explicar, ou seja, o regressor, \mathbf{S}_t é o vetor $1 \times k$ de variáveis explicativas a cada instante t , $\boldsymbol{\mu}$ é o vetor $k \times 1$ de coeficientes da regressão e ϵ_t é o termo de erro normalmente distribuído com variância σ_ϵ^2 . Considerando agora que se deseja atribuir uma dinâmica temporal ao coeficiente $\boldsymbol{\mu}$, fazendo $\boldsymbol{\mu} = \boldsymbol{\mu}_t$ e impondo uma equação de variação para ele, pode-se analisar este modelo como um caso especial do modelo geral dado pelas equações (3.1)-(3.2), de forma que o filtro de Kalman pode ser aplicado.

Considerando uma abordagem geral para modelos de regressão linear cujos coeficientes variam estocasticamente ao longo do tempo, um modelo univariado de regressão com coeficientes variantes no tempo pode ser escrito como:

$$y_t = \mathbf{S}_t \boldsymbol{\mu}_t + \epsilon_t \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad t = 1 \text{ a } N \quad (3.24)$$

$$\boldsymbol{\mu}_{t+1} = \mathbf{T} \boldsymbol{\mu}_t + \boldsymbol{\eta}_t \quad \boldsymbol{\eta}_t \sim N(0, \mathbf{Q}) \quad (3.25)$$

Se $\mathbf{Q} = \mathbf{0}$ e $\mathbf{T} = \mathbf{I}$, o modelo é reduzido ao modelo de regressão linear simples da equação (3.23). Nosso interesse aqui é a estimação de $\boldsymbol{\mu}_t$ ($t = 1 \dots N$), bem como dos hiperparâmetros envolvidos nas matrizes do sistema ($\sigma_\epsilon^2, \mathbf{T}, \mathbf{Q}$).

Da mesma forma que para o modelo geral, as estimativas de $\boldsymbol{\mu}_t$ podem ser obtidas a partir das equações de previsão do filtro de Kalman apresentadas na seção 3.2. Diferentes modelagens para a dinâmica temporal dos coeficientes $\boldsymbol{\mu}_t$ da regressão podem ser propostas, derivados a partir de diferentes premissas acerca da matriz \mathbf{T} .

3.5.1 Modelo de reversão à média

Mergner (2009) apresenta uma especificação alternativa do modelo representado pelo sistema de equações (3.24)-(3.25) dado por:

$$y_t = \mathbf{S}_t \boldsymbol{\mu}_t + \epsilon_t \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (3.26)$$

$$\boldsymbol{\mu}_{t+1} - \bar{\boldsymbol{\mu}} = \mathbf{T}(\boldsymbol{\mu}_t - \bar{\boldsymbol{\mu}}) + \boldsymbol{\eta}_t \quad \boldsymbol{\eta}_t \sim N(0, \mathbf{Q}) \quad (3.27)$$

onde as raízes características da matriz T tem valor absoluto menor do que um de forma que o vetor de coeficientes μ_t é estacionário. Caracterizado como processo de reversão à média, pode-se atribuir interpretação a $\bar{\mu}$ e T , sendo o primeiro a média de longo prazo do processo estocástico e o segundo relacionado à velocidade de reversão, ou à persistência com que os valores dos coeficientes μ_t reverterem à média. Esta especificação é bastante utilizada na literatura para caracterizar a evolução de betas em modelos de fatores condicionais, em trabalhos como os de Rosenberg (1973), Collins (1987) e, de especial interesse nesta tese, Mergner (2009) e Adrian e Franzoni (2009). Como explicitado por Mergner (2009), definindo $\mu_t^* = \mu_t - \bar{\mu}$, o modelo de reversão à média pode ser alternativamente reescrito como:

$$y_t = (\mathbf{S}_t \quad \mathbf{S}_t) \begin{pmatrix} \mu_t^* \\ \bar{\mu}_t \end{pmatrix} + \epsilon_t \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (3.28)$$

$$\begin{pmatrix} \mu_{t+1}^* \\ \bar{\mu}_{t+1} \end{pmatrix} = \begin{pmatrix} T & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix} \begin{pmatrix} \mu_t^* \\ \bar{\mu}_t \end{pmatrix} + \begin{pmatrix} \eta_t \\ \mathbf{0} \end{pmatrix} \quad \eta_t \sim N(0, Q) \quad (3.29)$$

Ou ainda, para manter os coeficientes μ_t diretamente no vetor de estado, poder-se-ia escrever:

$$y_t = (\mathbf{S}_t \quad \mathbf{0}) \begin{pmatrix} \mu_t \\ \bar{\mu}_t \end{pmatrix} + \epsilon_t \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (3.30)$$

$$\begin{pmatrix} \mu_{t+1} \\ \bar{\mu}_{t+1} \end{pmatrix} = \begin{pmatrix} T & I - T \\ \mathbf{0} & I \end{pmatrix} \begin{pmatrix} \mu_t \\ \bar{\mu}_t \end{pmatrix} + \begin{pmatrix} \eta_t \\ \mathbf{0} \end{pmatrix} \quad \eta_t \sim N(0, Q) \quad (3.31)$$

Há duas formas de se trabalhar com a estimação dos hiperparâmetros e do vetor de estado deste modelo no que se refere ao tratamento da média de longo prazo $\bar{\mu}$. Considerando as equações (3.26)-(3.27), $\bar{\mu}$ pode ser estimado como um hiperparâmetro do modelo por máxima verossimilhança. Por outro lado, se incluído no vetor de estado como no modelo dado pelas equações (3.30)-(3.31), $\bar{\mu}$ não precisa ser tratado como um hiperparâmetro, podendo ser estimado recursivamente a cada nova observação da série y_t a partir do filtro de Kalman. Trata-se de um procedimento equivalente à inclusão de um vetor de coeficientes no vetor de estado (Durbin e Koopman, 2001; seção 6.2.2). O primeiro tratamento é utilizado por Mergner e Bulla (2008) e Mergner (2009) e o segundo por Adrian e Franzoni (2009).

Outra característica deste modelo é que dependendo dos valores de \mathbf{T} , é possível derivar outros casos particulares para a dinâmica temporal dos coeficientes. Se $\mathbf{T} = \mathbf{I}$, o modelo se enquadra no caso em que os coeficientes seguem um processo de passeio aleatório. No caso em que $\mathbf{T} = \mathbf{0}$, o modelo passa a ser chamado de coeficientes aleatórios, de forma que os coeficientes flutuam aleatoriamente em torno da média de longo prazo.

3.5.2 Modelo de passeio aleatório

Considerando o caso em que $\mathbf{T} = \mathbf{I}$ no modelo descrito pelas equações (3.24)-(3.25), os coeficientes $\boldsymbol{\mu}_t$ apresentam dinâmica temporal dada por um processo de passeio aleatório, de forma que:

$$y_t = \mathbf{S}_t \boldsymbol{\mu}_t + \epsilon_t \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (3.32)$$

$$\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t + \boldsymbol{\eta}_t \quad \boldsymbol{\eta}_t \sim N(0, \mathbf{Q}) \quad (3.33)$$

Alguns autores propõem a análise de modelos de fatores nos quais os coeficientes são descritos por um processo de passeio aleatório. Zivot (2003) e Tsay (2010) e apresentam exemplos de estimação do CAPM modelando betas como processo estocásticos de passeio aleatório. De forma prática e com bons resultados, Mergner (2009) e Faff, Hillier e Hillier (2000) utilizam esta modelagem em modelos de fatores para descrever a evolução dos coeficientes e constatarem a boa performance quando comparado com modelos alternativos. Neste caso, apenas os hiperparâmetros das variâncias precisam ser estimados.

3.5.3 Valores iniciais

Para aplicar o algoritmo do filtro de Kalman, dois conjuntos de valores iniciais são necessários. O primeiro se refere a valores iniciais para os hiperparâmetros $\boldsymbol{\psi}$ a serem estimados e o segundo os valores de inicialização da média e variância do vetor de estado.

O conjunto inicial de valores para os hiperparâmetros é necessário para o processo de estimação a partir da maximização da função de log-verossimilhança.

De forma geral, considerando o modelo dado pelas equações (3.1)-(3.2), os hiperparâmetros podem estar presentes nas matrizes dos sistemas, quais sejam, \mathbf{S}_t , \mathbf{T}_t , \mathbf{d}_t , \mathbf{c}_t , \mathbf{U}_t , \mathbf{H}_t e \mathbf{Q}_t . Especificamente para os modelos abordados na seção anterior, o conjunto de hiperparâmetros inclui a variância σ_ϵ^2 , as informações da matriz de variância-covariância \mathbf{Q} e os dados da matriz \mathbf{T} . Os valores iniciais de média e variância do vetor de estados são necessários para a inicialização do filtro de Kalman. No Capítulo 4, serão estimados os modelos de interesse para as aplicações desta tese a partir de séries sintéticas, de forma que serão destacados os valores utilizados em cada caso. Nos Capítulos 5 e 6, serão estimados modelos em que os coeficientes seguem processos de passeio aleatório e reversão à média, sendo que neste último caso, o tratamento da média de longo prazo será realizado de duas formas diferentes, uma no vetor de estado e outra estimada como hiperparâmetro. Para os casos de passeio aleatório e reversão à média com a média de longo prazo no vetor de estado, a inicialização do filtro de Kalman será através da forma difusa exata. Para o caso de reversão à média em que a média de longo prazo é estimada como hiperparâmetro, a inicialização será padrão.

3.6 Ajuste e diagnóstico do modelo

Uma vez estimado o modelo, é necessário verificar o quão bem ele se ajusta aos dados e se os resíduos obtidos a partir dele confirmam as premissas adotadas. Considerando a estimação dos hiperparâmetros $\boldsymbol{\psi}$ a partir dos conceitos apresentados, é desejável medir o ajuste do modelo à série de dados. Quando se avaliam modelos alternativos, uma das formas de compará-los é através das medidas de AIC (*Akaike Information Criteria*) e BIC (*Bayesian Information Criteria*), que consideram uma comparação entre os valores assumidos pela função de verossimilhança de um determinado modelo já penalizando-a pelo número de parâmetros estimados, de forma que a comparação se torne justa no sentido de não beneficiar o modelo com mais parâmetros. Sendo $L(y|\hat{\boldsymbol{\psi}})$ o valor da função de verossimilhança, as medidas AIC e o BIC são dadas por (Durbin e Koopman, 2001):

$$AIC = \frac{1}{N} [-2 \log L(y|\hat{\boldsymbol{\psi}}) + 2w] \quad (3.34)$$

$$BIC = \frac{1}{N} [-2 \log L(y|\hat{\psi}) + w \log N] \quad (3.35)$$

onde N é o tamanho da série e w é o número de hiperparâmetros a serem estimados. No caso de inicialização difusa, usa-se o valor da função de verossimilhança difusa, considerando ainda o número de elementos difusos no vetor de estado, de forma que:

$$AIC = \frac{1}{N} [-2 \log L(y|\hat{\psi}) + 2(q + w)] \quad (3.36)$$

$$BIC = \frac{1}{N} [-2 \log L(y|\hat{\psi}) + (q + w) \log N] \quad (3.37)$$

onde q é o número de elementos difusos no vetor de estado.

Uma análise de diagnósticos é também necessária. A premissa do modelo é que os distúrbios ϵ_t e η_t são normalmente distribuídos e serialmente independentes com variâncias constantes. Considerando modelos univariados como apresentados na seção 3.5, os erros de previsão um-passo-a-frente padronizados são dados por:

$$e_t = \frac{v_t}{\sqrt{F_t}} \quad t = 1 \text{ a } N \quad (3.38)$$

(ou para $t = q$ a N no caso de inicialização difusa) são também normalmente distribuídos e serialmente independentes com variância unitária. Estas propriedades podem ser verificadas através de testes de diagnóstico relativos à normalidade, autocorrelação e heterocedasticidade (Durbin e Koopman, 2001).

Para testar a normalidade dos resíduos, será utilizado nesta tese o teste de Jarque-Bera, que combina os valores observados de assimetria e curtose da série temporal de forma a verificar se são consistentes com as premissas de normalidade. As hipóteses nula e alternativa do teste de Jarque-Bera são, respectivamente:

H_0 : Série é normalmente distribuída

H_a : Série não é normalmente distribuída

Nesse caso, se os resíduos padronizados são assintoticamente normalmente distribuídos, $S \sim N\left(0, \frac{6}{N}\right)$ e $K \sim N\left(3, \frac{24}{N}\right)$, onde S é a assimetria amostral e K a curtose amostral. A estatística de teste e sua distribuição sob hipótese nula são dadas por

$$JB = N \left[\frac{S^2}{6} + \frac{(K-3)^2}{24} \right] \sim \chi^2_2 \quad (3.39)$$

Para testar a existência de autocorrelação dos resíduos, será utilizado nesta tese o teste de Ljung-Box, que avalia se a autocorrelação presente na série é insignificante até determinado *lag* m . As hipóteses nula e alternativa do teste de Ljung-Box são, respectivamente:

H_0 : FAC da série até a ordem m são iguais a zero

H_a : Pelo menos uma das FAC é diferente de zero

A estatística de teste e sua distribuição sob hipótese nula é dada por:

$$Q(m) = N(N+2) \sum_{h=1}^m \frac{\hat{\rho}_h^2}{N-h} \sim \chi^2_m \quad (3.40)$$

onde $\hat{\rho}_h$ é a autocorrelação de ordem h da série, no caso dos resíduos padronizados.

Para testar a existência de heterocedasticidade condicional dos resíduos, característica essa equivalente à autocorrelação no seu quadrado, nesta tese será usado o teste ARCH de Engle. O teste mede a significância dos efeitos ARCH. Considerando os resíduos padronizados, se supusermos efeitos ARCH até o lag m , pode-se escrever:

$$e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \dots + \alpha_m e_{t-m}^2 + u_t \quad (3.41)$$

Assim, a hipótese nula e a hipótese alternativa do teste ARCH são, respectivamente:

H_0 : Não há heterocedasticidade, ou seja, $\alpha_0 = \alpha_1 = \dots = \alpha_m = 0$

H_a : Há heterocedasticidade

A estatística de teste e sua distribuição sob hipótese nula é dada por

$$LM(m) = NR^2 \sim \chi^2_m \quad (3.42)$$

onde R^2 é o coeficiente de determinação do ajuste do modelo ARCH(m) através de regressão.