

3

Modelo Não-Gaussiano: Abordagem Modificada de Durbin–Koopman

3.1

Introdução

Como mencionado anteriormente, existem situações em que um modelo linear gaussiano não proporciona uma representação apropriada para a série temporal investigada. Isto pode acontecer porque os dados provêm de uma distribuição diferente da normal, ou porque um modelo em espaço de estado não-linear é mais apropriado para descrever as observações em estudo.

A seguir, apresentam-se, brevemente, uma discussão destes modelos e, com maior riqueza de detalhes, uma metodologia de estimação dos mesmos baseada em uma modificação daquela proposta em [8]. Como será visto, este novo processo de estimação possui maior eficiência do que o original. Além disso, expressões de derivadas analíticas associadas às implementações computacionais serão apresentadas, assim como uma pertinente discussão sobre identificação paramétrica.

3.2

O Modelo Não-Gaussiano

Seja $\theta_t = Z_t\alpha_t$, chamado sinal do modelo. A distribuição das observações, y_t , condicional ao sinal θ_t , é dada por:

$$p(y_t|\alpha_1, \dots, \alpha_t, y_1, \dots, y_{t-1}) = p(y_t|\theta_t). \quad (3-1)$$

O vetor de estado obedece à seguinte equação recursiva:

$$\alpha_{t+1} = T_t\alpha_t + R_t\eta_t, \quad \eta_t \sim p(\eta_t) \quad (3-2)$$

para $t = 1, \dots, n$; onde os η_t 's são independentes e $p(y_t|\theta_t)$ e/ou $p(\eta_t)$ podem ser não-gaussianos.

A especificação das densidades em termos de θ_t , em vez do vetor de estado α_t , proporciona um ganho computacional significativo devido à redução da dimensionalidade, pois geralmente a dimensão de θ_t é menor que a dimensão de α_t .

DK consideram os dois casos seguintes:

1. Observações da família exponencial com densidades da forma

$$p(y_t|\theta_t) = \exp[y_t'\theta_t - b_t(\theta_t) + c_t(y_t)], \quad -\infty < \theta_t < \infty, \quad (3-3)$$

$b_t(\theta_t)$ é duas vezes diferenciável e $c_t(y_t)$ é função só de y_t .

2. Observações geradas pela relação

$$y_t = \theta_t + \epsilon_t, \quad \epsilon_t \sim p(\epsilon_t), \quad (3-4)$$

$p(\epsilon_t)$ não-gaussiana.

O modelo (3-3) com a equação de estado gaussiana, foi originalmente proposto por West, Harrison e Migon[?] com o nome de Modelo Linear Dinâmico Generalizado.

A família exponencial considera, entre outras, as distribuições Poisson, Bernoulli, Binomial, Binomial Negativa e Multinomial. No caso (2), é possível considerar perturbações com caudas mais pesadas que a distribuição normal, por exemplo, a distribuição t .

A abordagem modificada de DK será introduzida neste capítulo considerando-se somente a situação em que o estado mantém sua forma gaussiana tradicional definida por (2-2). Em sua metodologia original, DK consideram, em primeiro lugar, a estimação do vetor de hiperparâmetros do modelo, e em seguida a estimação do vetor de estado ou alguma função dele condicional às observações disponíveis. A nossa metodologia proposta seguirá a mesma perspectiva. As modificações em relação à proposta original ficarão claras no decorrer do texto. Enumeram-se, primeiramente, os objetivos principais:

1. Estimação do vetor de hiperparâmetros ψ , sendo que, em primeiro lugar, deve-se obter a verossimilhança do modelo e, em seguida utilizar, um método adequado de otimização;

2. Alisamento ou suavização (*Smoothing*), isto é, estimação do vetor de estado ou alguma função dele, condicional ao conhecimento de todas as observações.

Na seção seguinte, é estudado o problema da estimação dos hiperparâmetros do modelo. Na última seção é considerado o problema da estimação de funções do vetor de estado condicional as observações.

3.3 Estimação dos Hiperparâmetros

3.3.1 Amostragem por Importância

Para resolver o problema da estimação dos hiperparâmetros do modelo via abordagem de DK, deve-se estimar o logaritmo da função de verossimilhança por simulação de Monte Carlo, posto que, geralmente, não é possível resolver as integrais analiticamente no caso não-gaussiano.

Uma vez obtida a função de verossimilhança, esta deve ser maximizada numericamente com respeito a ψ , onde ψ é o vetor de hiperparâmetros, o qual contém os parâmetros fixos do modelo, tipicamente as variâncias dos choques aleatórios.

Sejam $y = (y'_1, \dots, y'_n)'$, $\alpha = (\alpha'_1, \dots, \alpha'_n)'$ e $\theta = (\theta'_1, \dots, \theta'_n)'$.

A função de verossimilhança do modelo não-gaussiano em estudo, denotada por $L_p(\psi)$, pode ser escrita como:

$$L_p(\psi) = p(y|\psi) = \int p(y, \theta|\psi) d\theta = \int p(y|\theta, \psi) p(\theta|\psi) d\theta. \quad (3-5)$$

Em princípio $L_p(\psi)$ poderia ser estimada por simulação, escolhendo N amostras independentes: $\theta^{(1)}, \dots, \theta^{(N)}$ da distribuição com densidade $p(\theta|\psi)$ e a estimação de $L_p(\psi)$ seria feita pela média dos valores resultantes de $p(y|\theta^{(i)}, \psi)$, para $i = 1, \dots, N$. Todavia em geral, não é possível utilizar a equação (3-5) para o cálculo de $L_p(\psi)$, uma vez que não é possível gerar valores desde $p(\theta|\psi)$. Para resolver esta dificuldade utiliza-se a técnica denominada *amostragem por importância*, que consiste em calcular (3-5) utilizando-se amostras de outra densidade, chamada *densidade de importância*.

Uma discussão de amostragem por importância é apresentada em Ripley [29].

Com o intuito de manter uma notação simples, e como todas as densidades consideradas dependem de ψ , esta dependência não será mais indicada explicitamente.

Para fazer uma ilustração simples da idéia de amostragem por importância, considere o problema da estimação da média condicional:

$$\bar{x} = E[x(\alpha)|y] = \int x(\alpha)p(\alpha|y)d\alpha \quad (3-6)$$

de uma função arbitrária $x(\alpha)$ de α dado o vetor de observações. Esta formulação inclui, por exemplo, a estimação de quantidades como a média $E(\alpha_t|y)$ do vetor de estado α_t dado y (e ψ), em que $x(\alpha) = \alpha_t$.

Poderíamos obter uma amostra da distribuição $p(\alpha|y)$, e estimar \bar{x} pela média amostral dos valores correspondentes de $x(\alpha)$.

Na prática, posto que para os modelos estudados não estão disponíveis expressões explícitas para $p(\alpha|y)$, esta idéia não é factível. Para resolver este problema deve-se procurar por uma densidade, o mais próxima possível de $p(\alpha|y)$, a partir da qual possamos obter uma amostra aleatória, e amostramos desta distribuição, fazendo-se os ajustes apropriados da integral (3-6). Esta técnica é chamada de *amostragem por importância* e a densidade da qual é feita a amostragem é referida como *densidade de importância*, DK utilizam densidades de importância gaussianas, as quais também serão adotadas no nosso trabalho.

A idéia da amostragem por importância esta baseada no seguinte raciocínio: Dado ψ , seja $g(\alpha|y)$ uma densidade de importância gaussiana escolhida como a mais próxima possível de $p(\alpha|y)$. De (3-6) temos,

$$\bar{x} = \int x(\alpha) \frac{p(\alpha|y)}{g(\alpha|y)} g(\alpha|y) d\alpha = E_g \left[x(\alpha) \frac{p(\alpha|y)}{g(\alpha|y)} \right], \quad (3-7)$$

onde E_g denota esperança com respeito à densidade de importância $g(\alpha|y)$.

Deste modo deve-se amostrar de $g(\alpha|y)$, ao invés de $p(\alpha|y)$, para avaliarmos (3-6).

3.3.2 Cálculo da Verossimilhança

Na metodologia de DK, a obtenção da verossimilhança do modelo de interesse (não-gaussiana), $L_p(\psi)$, é feita calculando-se em primeiro lugar a verossimilhança de um modelo linear gaussiano chamado de *Modelo Linear Gaussiano Aproximado*, para em seguida fazer um ajuste na verossimilhança deste modelo gaussiano, denotada por $L_g(\psi)$, e então obter a verossimil-

hança do modelo não-gaussiano em estudo, isto é, $L_p(\psi)$. O ajuste é feito utilizando-se a idéia de amostragem por importância desenvolvida anteriormente.

Como foi mostrado, a verossimilhança do modelo não-gaussiano é definida por:

$$L_p(\psi) = p(y) = \int p(y, \theta) d\theta = \int p(y|\theta)p(\theta) d\theta \quad (3-8)$$

A verossimilhança do modelo gaussiano aproximado é definida por:

$$L_g(\psi) = g(y) = \frac{g(y, \theta)}{g(\theta|y)} = \frac{g(y|\theta)p(\theta)}{g(\theta|y)} \quad (3-9)$$

Substituindo $p(\theta)$, obtido a partir de (3-9) em (3-8) temos

$$L_p(\psi) = L_g(\psi) \int \frac{p(y|\theta)}{g(y|\theta)} g(\theta|y) d\theta = L_g(\psi) E_g \left\{ \frac{p(y|\theta)}{g(y|\theta)} \right\}, \quad (3-10)$$

onde E_g denota esperança com respeito à densidade gaussiana $g(\theta|y)$.

Uma outra forma equivalente de obter $L_p(\psi)$ a partir de $L_g(\psi)$ pode ser deduzida considerando que:

$$L_p(\psi) = \int p(\alpha, y) d\alpha$$

Dividindo e multiplicando pela densidade de importância $g(\alpha|y)$ obtém-se:

$$\begin{aligned} L_p(\psi) &= \int \frac{p(\alpha, y)}{g(\alpha|y)} g(\alpha|y) d\alpha \\ &= g(y) \int \frac{p(\alpha, y)}{g(\alpha, y)} g(\alpha|y) d\alpha \\ &= L_g(\psi) E_g \left[\frac{p(\alpha, y)}{g(\alpha, y)} \right] \end{aligned}$$

Se o vetor de estado é gaussiano a última equação é equivalente à equação (3-10).

Portanto, para estimar a verossimilhança do modelo não-gaussiano, utilizamos simulação para estimar o ajuste necessário na verossimilhança gaussiana, dado pelo termo $E_g \left\{ \frac{p(y|\theta)}{g(y|\theta)} \right\}$.

Assim sendo, o problema da obtenção da verossimilhança do modelo de interesse é resolvido da forma seguinte:

1. Achar uma densidade de importância gaussiana o mais próxima possível da densidade em estudo.

2. Calcular $L_g(\psi)$ usando o filtro de Kalman.
3. Obter $L_p(\psi)$, gerando amostras desde $g(\theta|y)$ mediante um método apropriado e utilizando a equação (3-10).

O problema passa a ser então encontrar a densidade de importância, e obter amostras desta densidade condicional nas observações.

A forma de se achar a densidade de importância referente à metodologia será descrita na subseção seguinte. Cabe observar que originalmente, no artigo de Durbin & Koopman [8], é sugerido o algoritmo de simulação suavizada *Simulation Smoothing* de De Jong & Shephard [21] para gerar as amostras de $g(\theta|y)$; entretanto em artigo mais recente Durbin & Koopman propõem um novo algoritmo muito mais simples e mais fácil de implementar para simulação suavizada, o qual será utilizado na implementação do modelo proposto na tese.

3.3.3 Seleção da Densidade de Importância

Nesta seção, iremos descrever como se obtém a densidade de importância gaussiana necessária para simulação, de forma a manter a lógica proposta por DK, mas que se incorra a uma maior eficiência computacional.

A densidade de importância pode ser obtida assumindo que as observações y_t são geradas pelo modelo gaussiano linear, (2-1) e (2-2), mas permitindo que os choques ϵ_t tenham maior flexibilidade, isto é, supondo que $\epsilon_t \sim N(c_t, H_t)$. Assim sendo, o problema de achar a densidade de importância reduz-se a achar c_t e H_t , para $t = 1, \dots, n$.

Seja $\hat{\theta} = E_g(\theta)$, o valor de θ que é obtido pelo filtro de Kalman suavizado, sob o pressuposto de que o modelo aproximado é o correto. No processo de simulação suavizada os valores simulados de θ são distribuídos normalmente numa vizinhança da média $\hat{\theta}_t$. Durbin e Koopman [8] escolhem os c_t e H_t de modo que $p(y|\theta)$ e $g(y|\theta)$ sejam os mais próximos possíveis na vizinhança de $\hat{\theta}$, fazendo $p(y|\theta) \approx g(y|\theta)$, o que leva a:

$$\frac{\partial \log p(y_t|\theta_t)}{\partial \theta_t} = \frac{\partial \log g(y_t|\theta_t)}{\partial \theta_t}$$

$$\frac{\partial^2 \log p(y_t|\theta_t)}{\partial \theta_t \partial \theta_t'} = \frac{\partial^2 \log g(y_t|\theta_t)}{\partial \theta_t \partial \theta_t'}$$

Deste modo as restrições consideradas são:

$$\left. \frac{\partial \log p(y_t|\theta_t)}{\partial \theta_t} \right|_{\theta_t=\hat{\theta}_t} = \left. \frac{\partial \log g(y_t|\theta_t)}{\partial \theta_t} \right|_{\theta_t=\hat{\theta}_t}$$

e

$$\left. \frac{\partial^2 \log p(y_t|\theta_t)}{\partial \theta_t \partial \theta_t'} \right|_{\theta_t=\hat{\theta}_t} = \left. \frac{\partial^2 \log g(y_t|\theta_t)}{\partial \theta_t \partial \theta_t'} \right|_{\theta_t=\hat{\theta}_t}$$

onde:

$$\log g(y_t|\theta_t) = \text{constante} - \frac{1}{2}(y_t - \theta_t - c_t)' H_t^{-1} (y_t - \theta_t - c_t).$$

Derivando $g(y_t|\theta_t)$ e avaliando em $\theta = \hat{\theta}$, chega-se a:

$$\left. \frac{\partial \log p(y_t|\theta_t)}{\partial \theta_t} \right|_{\theta_t=\hat{\theta}_t} - H_t^{-1}(y_t - \hat{\theta}_t - c_t) = 0, \quad (3-11)$$

e

$$\left. \frac{\partial^2 \log p(y_t|\theta_t)}{\partial \theta_t \partial \theta_t'} \right|_{\theta_t=\hat{\theta}_t} + H_t^{-1} = 0. \quad (3-12)$$

A solução é dada por:

$$H_t = \left(- \left. \frac{\partial^2 \log p(y_t|\theta_t)}{\partial \theta_t \partial \theta_t'} \right|_{\theta_t=\hat{\theta}_t} \right)^{-1}, \quad (3-13)$$

e

$$c_t = y_t - \hat{\theta}_t - H_t \left. \frac{\partial \log p(y_t|\theta_t)}{\partial \theta_t} \right|_{\theta_t=\hat{\theta}_t}. \quad (3-14)$$

Na prática, se usa o vetor

$$y_t^* = y_t - c_t = \hat{\theta}_t + H_t \left. \frac{\partial \log p(y_t|\theta_t)}{\partial \theta_t} \right|_{\theta_t=\hat{\theta}_t}.$$

Na metodologia original de DK, estas equações são resolvidas iterativamente até convergirem, considerando valores iniciais arbitrários. Para cada par de ensaio $\{c_t, H_t\}$ é obtido um novo valor de $\hat{\theta}_t$ pelo filtro de Kalman suavizado. Este novo valor é usado em (3-13) e (3-14) para obter novos valores para c_t e H_t .

É justamente aqui que se configura a mudança da nossa metodologia: *ao invés de trabalharmos com matrizes de covariâncias H_t “cheias”, visaremos a obtenção de matrizes diagonais.* Logo, usaremos a seguinte fórmula iterativa:

$$H_t = \left[\text{diag} \left(- \frac{\partial^2 \log p(y_t | \theta_t)}{\partial \theta_t \partial \theta_t'} \Big|_{\theta_t = \bar{\theta}_t} \right) \right]^{-1}, \quad (3-15)$$

Esta simplificação, aparentemente inócua, implica em três conseqüências no que diz respeito à eficiência computacional, quais sejam:

1. As inversões descritas em (3-13), agora, passam a ser feitas com matrizes diagonais;
2. Com um Modelo Gaussiano Aproximador abrangendo matrizes H_t diagonais, torna-se viável a uso do tratamento univariado para séries multivariadas descrito em [7], pgs: 128-134, sem a necessidade de que sejam praticadas transformações em y_t^* como as baseadas na decomposição de Choleski lá descritas;
3. Como decorrência do item anterior, torna-se possível também a prática da inicialização exata descrita em [24] com mais desenvoltura.

Observe-se que a nossa metodologia coincide com a proposta originalmente por DK quando o modelo em espaço de estado não Gaussiano em uso for *univariado*.

3.3.4 Verossimilhança por Monte Carlo

De acordo com a seção [3.3.2], para se achar a verossimilhança do modelo não-gaussiano, uma vez encontrado o modelo gaussiano aproximado, basta ajustar a verossimilhança do modelo gaussiano aproximado via simulação..

Uma vez achado o modelo gaussiano aproximado, seja $\theta^{(i)}$, $i = 1, \dots, M$, uma amostra obtida da densidade $g(\theta|y)$, pelo algoritmo de simulação suavizada e seja $w_i = w(\theta^{(i)})$, onde :

$$w(\theta) = \frac{p(y|\theta)}{g(y|\theta)}. \quad (3-16)$$

O estimador por Monte Carlo de $L_p(\psi)$ é

$$\hat{L}_{p,1}(\psi) = L_g(\psi)\bar{w}, \quad (3-17)$$

onde $\bar{w} = M^{-1} \sum_{i=1}^M w_i$.

É mais conveniente trabalhar com $\log L_p(\psi)$ que com $L_p(\psi)$, já que os números considerados são de menor dimensão. Tomando log em (3-17), obtém-se

$$\log \hat{L}_{p,1}(\psi) = \log L_g(\psi) + \log \bar{w}. \quad (3-18)$$

Entretanto $E(\log \bar{w}) \neq \log \mu_w$, onde $\mu_w = E_g(w_i)$, de modo que (3-18) é um estimador viesado de $\log L_p(\psi)$.

DK mostram que :

$$\log \hat{L}_{p,2}(\psi) = \log L_g(\psi) + \log \bar{w} + \frac{s_w^2}{2N\bar{w}^2}. \quad (3-19)$$

é um estimador aproximadamente não viesado de $\log L_p(\psi)$.

3.3.5 Variáveis Antitéticas

A utilização de variáveis antitéticas permite aumentar a eficiência na estimação em dois sentidos: obtém-se mais de uma amostra para uma rodada do processo de simulação suavizada, e reduz-se a variabilidade do estimador.

DK consideram dois tipos de variáveis antitéticas: antitéticas por locação e antitéticas por escala.

Suponha que θ é uma amostra da densidade $g(\theta|y)$, obtida usando simulação suavizada. Seja $\tilde{\theta}_t = 2\hat{\theta}_t - \theta_t$, para $t = 1, \dots, n$. Isto proporciona uma amostra equiprovável, posto que, θ_t é distribuído normalmente na vizinhança de $\hat{\theta}_t$ e $\tilde{\theta}_t - \hat{\theta}_t = -(\theta_t - \hat{\theta}_t)$.

As amostras $\theta^{(i)}$ e $\tilde{\theta}^{(i)}$ são ditas balanceadas por locação. A correlação negativa entre $\tilde{\theta}$ e θ , induz uma correlação negativa entre \tilde{w} e w , onde \tilde{w} é obtido desde (3-16) com $\tilde{\theta}$ no lugar de θ , o que provoca uma redução na variância do estimador \bar{w} .

A segunda variável antitética é definida sob o presuposto que y_t é univariado, e portanto, θ_t é univariado. No procedimento de simulação suavizada utilizado para obter a amostra $\theta^{(i)}$ da densidade $g(\theta|y)$, θ é uma transformação linear $\theta = \hat{\theta}_t + Au$, com $u \sim N(0, 1)$.

Assim $V(\theta) = AA'$. Define-se :

$$c = (\theta - \hat{\theta})'V^{-1}(\theta - \hat{\theta}) = u'u \sim \chi_n^2.$$

Para um valor dado de c , seja $q = F(c)$, onde F é a função de distribuição da χ_n^2 e seja $\bar{c} = F^{-1}(1 - q)$. Então $F(\bar{c}) = 1 - F(c)$, isto é c e \bar{c} são os percentis da qui-quadrado a q e $1 - q$ por cento, respectivamente.

Define-se a variável :

$$\bar{\theta} = \hat{\theta} + (\bar{c}/c)^{\frac{1}{2}}(\theta - \hat{\theta}),$$

a qual tem a mesma distribuição que θ . Quando $|\theta - \hat{\theta}|$ é grande, $|\bar{\theta} - \hat{\theta}|$ é pequeno e vice-versa. As amostras consistentes dos pares $\bar{\theta}$ e θ são ditas de escalas balanceadas.

Seja \bar{w} o valor obtido desde (3-16) com $\bar{\theta}$ no lugar de θ .

DK utilizam ambas as técnicas simultaneamente, obtendo quatro valores de $w(\theta)$ para cada rodada da simulação suavizada.

Seja $\check{\theta}^{(i)}$ a variável balanceada de escala para $\check{\theta}^{(i)}$ e $\check{w}_{(i)}$ o valor de w_i obtido dela.

Fazendo

$$w_i^* = \frac{1}{4}(w_i + \check{w}_i + \bar{w}_i + \check{w}), \quad (3-20)$$

para $i = 1, \dots, M$.

Estima-se $\log L_p(\psi)$ por $\hat{L}_{p,3}(\psi)$, dado por (3-19) com w_i^* no lugar de w_i .

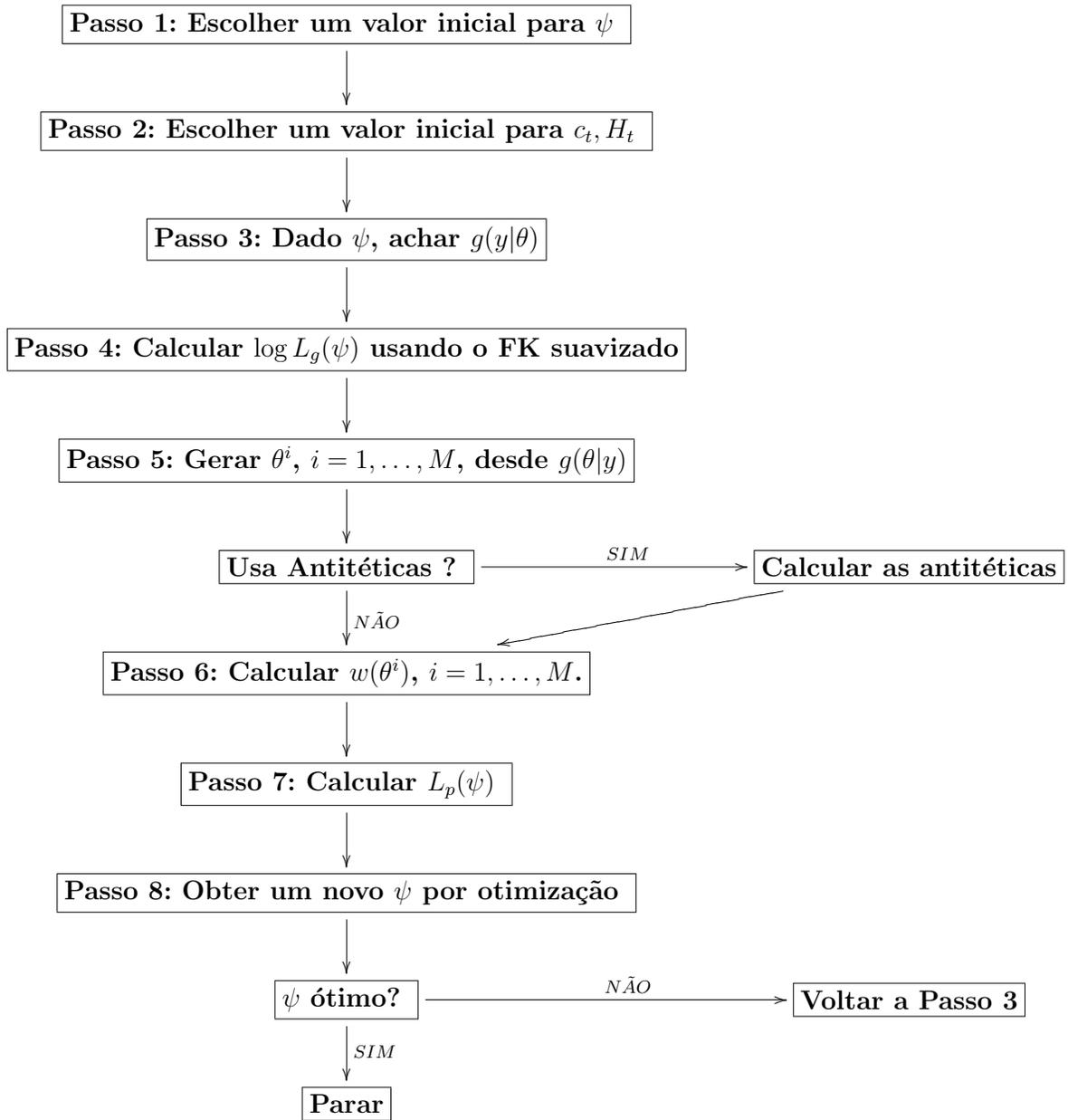
3.3.6 Estimação dos Hiperparâmetros Passo-a-Passo

As etapas necessárias para a estimação dos hiperparâmetros do modelo podem ser resumidas da maneira seguinte:

1. Escolher um valor inicial para o vetor de hiperparâmetros ψ , isto pode ser feito, por exemplo, considerando os hiperparâmetros obtidos ao supor um modelo gaussiano para o nosso problema.
2. Escolher um valor inicial para c_t e H_t , sendo esta última uma matriz diagonal.
3. Dado ψ , identificar uma densidade gaussiana $g(y|\theta)$ mais próxima possível de $p(y|\theta)$, isto é, achar os c_t e H_t através das fórmulas (3-15) e (3-13).
4. Calcular $\log L_g(\psi)$ usando o filtro de Kalman.
5. Considerar $g(\theta|y)$ como a densidade de importância; obter amostras $\theta^{(i)}$ desde $g(\theta|y)$ usando simulação suavizada, para $i = 1, \dots, M$.
6. Para cada i , calcular w_i^* por (3-20).

7. Calcular $\hat{L}_{p,2}(\psi)$ por (3-19), com w_i^* no lugar de w_i .
8. Obter um novo valor do vetor de hiperparâmetros ψ , através de um algoritmo de maximização de $\log \hat{L}_p(\psi)$, com respeito a ψ , utilizando um método numérico efetivo.
9. Verificar se o novo valor de ψ convergiu e parar em caso afirmativo, e voltar ao passo 2 caso contrário.

O diagrama seguinte apresenta um breve resumo dos passos necessários para a estimação de ψ :



3.4 Alisamento

Uma vez estimado ψ por maximização da função de verossimilhança do modelo, pode-se estimar a média condicional

$$\bar{x} = E[x(\alpha)|y] = \int x(\alpha)p(\alpha|y)d\alpha,$$

de uma função arbitrária $x(\alpha)$ de α , condicional ao vetor de observações.

Seja $\hat{\psi}$ a estimação de ψ , obtida da forma descrita na seção anterior e seja $g(\alpha|y)$ a densidade de importância gaussiana utilizada na estimação, da equação (3-7) tem-se:

$$\bar{x} = E_g \left[x(\alpha) \frac{p(\alpha|y)}{g(\alpha|y)} \right], \quad (3-21)$$

onde E_g denota a esperança com respeito a densidade de importância $g(\alpha|y)$. Geralmente $p(\alpha|y)$ e $g(\alpha|y)$ são complicados algebricamente, enquanto que $p(\alpha, y)$ e $g(\alpha, y)$ são mais simples de trabalhar.

Percebendo que

$$p(\alpha|y) = p(\alpha, y)/p(y)$$

e

$$g(\alpha|y) = g(\alpha, y)/g(y),$$

a equação (3-21) pode ser escrita como :

$$\bar{x} = \frac{g(y)}{p(y)} E_g \left[x(\alpha) \frac{p(\alpha, y)}{g(\alpha, y)} \right]. \quad (3-22)$$

Fazendo $x(\alpha) = 1$, temos

$$1 = \frac{g(y)}{p(y)} E_g \left[\frac{p(\alpha, y)}{g(\alpha, y)} \right]. \quad (3-23)$$

Considerando a razão (3-22)/(3-23) chega-se a

$$\bar{x} = \frac{E_g[x(\alpha)w(\alpha, y)]}{E_g[w(\alpha, y)]}, \quad \text{onde } w(\alpha, y) = \frac{p(\alpha, y)}{g(\alpha, y)}, \quad (3-24)$$

Esta fórmula é fundamental na metodologia de DK, posto que permite estimar a média de qualquer função do vetor de estado condicional às observações.

Poderíamos obter um estimador Monte Carlo \hat{x} de \bar{x} , escolhendo um conjunto de amostras independentes $\alpha^{(1)}, \dots, \alpha^{(M)}$, da distribuição com

densidade $g(\alpha|y)$ e considerando

$$\hat{x} = \frac{\sum_{t=1}^M x_t w_t}{\sum_{t=1}^M w_t}, \text{ onde } x_t = x(\alpha^{(t)}) \text{ e } w_t = w(\alpha^{(t)}, y). \quad (3-25)$$

Como as amostras são independentes e presupostos que se satisfazem geralmente na prática, \hat{x} converge para \bar{x} probabilisticamente, quando $M \rightarrow \infty$.

Este estimador é numericamente ineficiente e pode-se melhorá-lo utilizando-se variáveis antitéticas.

Note-se que quando o vetor de estado é linear e gaussiano, $p(\alpha) = g(\alpha)$, de modo que:

$$\frac{p(\alpha, y)}{g(\alpha, y)} = \frac{p(\alpha)p(y|\alpha)}{g(\alpha)g(y|\alpha)} = \frac{p(y|\alpha)}{g(y|\alpha)} = \frac{p(y|\theta)}{g(y|\theta)},$$

assim a equação (3-24) toma a forma mais simples:

$$\bar{x} = \frac{E_g[x(\alpha)w^*(\theta, y)]}{E_g[w^*(\theta, y)]}, \quad w^*(\theta, y) = \frac{p(y|\theta)}{g(y|\theta)},$$

\hat{x} é dado por uma formulação análoga à equação (3-25).

Finalmente, no caso que o vetor de estado é linear e gaussiano w^* , também pode ser escrita como:

$$w^*(\theta, y) = \prod_{t=1}^n \frac{p(y|\theta_t)}{g(\epsilon_t)}. \quad (3-26)$$