

## *Estimação dos parâmetros de modelos em espaço de estado com coeficientes variáveis*

**Marco Costa**

*Universidade de Aveiro - Escola Superior de Tecnologia e Gestão de Águeda*

**Teresa Alpuim**

*Universidade de Lisboa - Departamento de Estatística e Investigação Operacional da Faculdade de Ciências da Universidade de Lisboa*

**Resumo:** A estimação dos parâmetros é um processo importante em qualquer processo de modelação, em particular nos modelos em espaço de estados (MEE) com coeficientes variáveis, em tempo discreto. Este trabalho pretende contribuir para a especificação de MEE univariados propondo estimadores que não dependem da distribuição dos erros (estimadores *distribution-free*) e comparando-os, via estudo de Monte Carlo, com a estimação pela máxima verosimilhança.

**Palavras-chave:** Modelos em espaço de estados, estimação *distribution-free*, filtro de Kalman

**Abstract:** The parameters estimation is an important step in any modeling process, in particular, in the state space models at discrete time. This work intends to contribute for the specification of univariate state space models, suggesting distribution-free estimators and comparing them to likelihood estimation by a Monte Carlo study.

**Keywords:** State space models, distribution-free estimation, Kalman filter

## 1 Introdução

A apresentação do filtro de Kalman (FK) pelo engenheiro húngaro Rudolf E. Kalman no seu artigo de 1960, [9], associado aos modelos em espaço de estados, até então aplicados em problemas de controlo, originou uma rápida aplicação destes a diversas áreas. O filtro de Kalman tem como principal objectivo a estimação de variáveis não-observáveis (*estados*), sendo estas componentes de um sistema dinâmico que admite uma representação em estado de estados. Em geral, uma representação em espaço de estados, em tempo discreto, é caracterizada pelas equações,

$$\mathbf{Y}_t = \mathbf{H}_t \boldsymbol{\beta}_t + \mathbf{e}_t \quad (1)$$

$$\boldsymbol{\beta}_t = \boldsymbol{\mu}_t + \boldsymbol{\Phi} (\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t. \quad (2)$$

A equação de observação (1) relaciona o vector  $\mathbf{Y}_t$ , um vector  $(n \times 1)$  de variáveis observáveis no instante  $t$ , com o vector  $\boldsymbol{\beta}_t$ ,  $(m \times 1)$ , de variáveis não-observáveis

ou *estados*, através da matriz<sup>1</sup>  $\mathbf{H}_t$  de constantes conhecidas e um ruído branco<sup>2</sup>  $\mathbf{e}_t$  com matriz de covariâncias  $E[\mathbf{e}_t \mathbf{e}_t'] = \Sigma_{\mathbf{e}}$ . O vector de estados  $\beta_t$  é actualizado pela *equação de transição* (2), onde  $\mu$  é um vector ( $m \times 1$ ) de médias<sup>3</sup>,  $\Phi$  é uma matriz autoregressiva ( $m \times m$ ) e  $\varepsilon_t$  é um ruído branco com matriz de covariâncias  $E[\varepsilon_t \varepsilon_t'] = \Sigma_{\varepsilon}$ . Assume-se que os erros  $\mathbf{e}_t$  e  $\varepsilon_t$  são não correlacionados, isto é,  $E[\mathbf{e}_t \varepsilon_s'] = \mathbf{0}$ , para todo o  $t$  e  $s$ .

Seja  $\beta_{t|t-1}$  o estimador de  $\beta_t$  com base na informação disponível até ao instante  $t-1$ ,  $\mathcal{Y}_{t-1} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_{t-1})$ , e  $\mathbf{P}_{t|t-1}$  o seu erro quadrático médio. Assim, a previsão de  $\mathbf{Y}_t$  é dada por

$$\mathbf{Y}_{t|t-1} = \mathbf{H}_t \beta_{t|t-1}.$$

Mas, quando o vector  $\mathbf{Y}_t$  é conhecido no instante  $t$ , a inovação  $\eta_t = \mathbf{Y}_t - \mathbf{Y}_{t|t-1}$  é incorporada para actualizar a estimação de  $\beta_t$  pela equação

$$\beta_{t|t} = \beta_{t|t-1} + \mathbf{K}_t \eta_t$$

com erro quadrático médio igual a  $\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t \mathbf{H}_t \mathbf{P}_{t|t-1}$  onde  $\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{H}_t' (\mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t' + \Sigma_{\mathbf{e}})^{-1}$  é conhecida pela matriz *ganho de Kalman*. No instante  $t$ , obtemos a previsão para  $\beta_{t+1}$  pela equação

$$\beta_{t+1} = \mu + \Phi (\beta_{t|t} - \mu)$$

com erro quadrático médio igual a  $\mathbf{P}_{t+1|t} = \Phi \mathbf{P}_{t|t} \Phi' + \Sigma_{\varepsilon}$ .

O sucesso do filtro de Kalman é devido ao facto de este ser um algoritmo recursivo que permite obter, de uma forma simples, estimativas e predições óptimas para o vector de estados. De facto, as equações do FK constituem um sistema que permite, em cada instante, obter as projecções lineares, conseguindo-se, desta forma, estimadores lineares com erro quadrático médio mínimo.

Na prática, a implementação do filtro de Kalman depende da identificação dos parâmetros  $\Theta = \{\mu, \Phi, \Sigma_{\mathbf{e}}, \Sigma_{\varepsilon}\}$ . Se o vector  $\Theta$  é conhecido o modelo fica especificado, caso contrário, é necessária a sua estimação, sendo que a abordagem mais comum é a estimação pela máxima verosimilhança. Neste caso, assume-se que os erros são normalmente distribuídos e, em termos práticos, é necessário um vector de valores iniciais  $\Theta_0$ .

Sob a normalidade dos erros, o vector  $\mathbf{Y}_t$  condicionado a  $\mathcal{Y}_{t-1}$  tem distribuição normal com valor médio e matriz de covariâncias obtidos pelo FK, isto é,  $\mathbf{Y}_t | \mathcal{Y}_{t-1} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_{t-1}) \sim N(\mathbf{Y}_{t|t-1}, \Omega_t)$ , com  $\Omega_t = \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t' + \Sigma_{\varepsilon}$ .

<sup>1</sup>Se  $\mathbf{H}_t = \mathbf{H}$  para todo o  $t$ , o modelo diz-se de coeficientes constantes, caso contrário diz-se de coeficientes variáveis.

<sup>2</sup>Neste contexto, considera-se que  $\{\mathbf{e}_t\}_{t \in \mathbb{N}}$  é uma sucessão de variáveis aleatórias de média nula, com matriz de covariâncias  $\Sigma_{\mathbf{e}}$ , e não correlacionadas, isto é,  $E(\mathbf{e}_t \mathbf{e}_s') = \mathbf{0}$ , para  $t \neq s$ .

<sup>3</sup>Nalgumas aplicações o vector de estados  $\beta_t$  pode não ser estacionário em média, e neste caso omite-se o vector das médias. No entanto, neste trabalho estamos interessados em modelos com vector de estados estacionários de 2ª ordem.

Assim, a log-verosimilhança de  $(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n)$  é obtida a partir das inovações e tem a forma

$$\ln L(\Theta; \mathcal{Y}_n) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^n \ln(|\Omega_t(\Theta)|) - \frac{1}{2} \sum_{t=1}^n \eta'_t(\Theta) \Omega_t^{-1}(\Theta) \eta_t(\Theta).$$

As estimativas de máxima verosimilhança são obtidas maximizando numericamente a log-verosimilhança em ordem aos parâmetros desconhecidos recorrendo ao método dos *scores* de Fisher, [8], ou ao algoritmo EM, [5].

Por um lado, em diversas situações práticas a normalidade dos erros não é a abordagem mais correcta, por exemplo, na estimação da precipitação média em área (ver [2]) ou no cálculo de provisões actuariais (ver [3]). Por outro lado, o processo de maximização da log-verosimilhança pode não convergir, ou a representação em espaço de estados pode ter problemas de identificação. Assim, a procura de métodos alternativos torna-se bastante pertinente e útil.

Os modelos em espaço de estado com erros da família exponencial não-gaussianos foram largamente estudados desde o trabalho pioneiro de [13] ou, mais tarde, no livro [12]. Surgiram posteriormente outros trabalhos no mesmo âmbito, por exemplo o trabalho [6], mas onde a estimação dos parâmetros foi abordada superficialmente. A estimação de parâmetros em modelos não-gaussianos foi tratada de uma forma mais incisiva em trabalhos posteriores onde o método da máxima verosimilhança associado à análise bayesiana prevalece. Existe uma vasta bibliografia sobre a análise bayesiana com recurso às técnicas de Monte Carlo via cadeias de Markov para o tratamento de verosimilhanças não tratáveis de outra forma, por exemplo [11].

Estimadores para as variâncias dos erros foram propostos em [1]. A autora propôs estimadores baseados no método dos momentos aplicado às variâncias de transformações dos valores amostrais. Estes estimadores podem ser aplicados em modelos uni e multivariados, bem como em modelos com coeficientes constantes ou variáveis.

## 2 Estimação *distribution-free* dos parâmetros em MEE univariados

Os estimadores que propomos neste trabalho incidem sobre um subconjunto dos modelos (1)-(2) com especial interesse prático. Consideremos o modelo em espaço de estados definido pelas equações

$$Y_t = h_t \beta_t + e_t \quad (3)$$

$$\Phi_p(B) (\beta_t - \mu) = \varepsilon_t \quad (4)$$

onde  $\Phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ , onde  $B$  é o usual operador atraso, isto é,  $B(\beta_t - \mu) = \beta_{t-1} - \mu$ ,  $h_t$  é uma constante não nula e onde admitimos as seguintes condições:

- C1 -  $\phi_i$ , com  $i = 1, 2, \dots, p$  são tais que as raízes do polinómio  $x^p - \phi_1 x^{p-1} - \dots - \phi_p$  estão dentro do círculo unitário;
- C2 -  $e_t$  e  $\varepsilon_t$  são não correlacionados, com variâncias  $\text{var}(e_t) = \sigma_e^2$  e  $\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$ ;
- C3 -  $e_t$  e  $\varepsilon_t$  têm momentos de 4ª ordem finitos, isto é,  $\text{var}(e_t^2) = \tau_e < \infty$  e  $\text{var}(\varepsilon_t^2) = \tau_\varepsilon < \infty$ .

O modelo em espaço de estado (3)-(4) é especificado pelo vector de parâmetros desconhecidos  $\Theta = \{\mu, \phi, \sigma_e^2, \sigma_\varepsilon^2\}$ , onde  $\phi' = [\phi_1, \phi_2, \dots, \phi_p]$ , e admite uma representação em espaço de estados equivalente ao modelo (1)-(2), (ver [7]).

## 2.1 Estimadores propostos

A estimação *distribution-free* apresentada nesta secção é baseada no métodos dos momentos aplicado a transformações do processo  $\{h_t^{-1}Y_t\}$ . Este processo não é estacionário em variância sendo, contudo, estacionário em média. Assim, dada a amostra  $(Y_1, Y_2, \dots, Y_n)$  e as constantes conhecidas  $h_1, h_2, \dots, h_n$ , o estimador natural para a média  $\mu$  é

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^n h_t^{-1} Y_t \quad (5)$$

O teorema 2.1 estabelece uma condição suficiente para a consistência de  $\hat{\mu}$  verificando-se, trivialmente, que é centrado.

**Teorema 2.1.** *Considere-se o modelo em espaço de estados definido pelas equações (3)-(4) onde se verificam as condições C1 e C2. Se existir  $c > 0$ , tal que  $|h_t| > c$  para todo o  $t \in \mathbb{N}$ , então  $\hat{\mu}$  é consistente para  $\mu$ .*

*Demonstração.* Uma vez que  $E(\hat{\mu}) = \mu$  basta provar que  $\text{var}(\hat{\mu}) \rightarrow 0$  quando  $n \rightarrow +\infty$ . Com pura manipulação algébrica, temos,

$$\begin{aligned} \text{var}(\hat{\mu}) &= \frac{1}{n^2} \sum_{t=1}^n \sum_{s=1}^n \text{cov}(h_t^{-1}Y_t, h_s^{-1}Y_s) \\ &= \frac{1}{n^2} \left[ \sigma_e^2 \sum_{t=1}^n h_t^{-2} + \sum_{t=1}^n \sum_{s=1}^n \gamma_\beta(t-s) \right] \\ &= \frac{\sigma_e^2}{n^2} \sum_{t=1}^n h_t^{-2} + \frac{\sigma_\beta^2}{n} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \rho_\beta(k). \end{aligned}$$

Sob hipótese,  $|h_t| > c$  para todo o  $t \in \mathbb{N}$ , pelo que  $\lim_{n \rightarrow +\infty} n^{-2} \sum_{t=1}^n h_t^{-2} = 0$ . Atendendo à estacionaridade devida à condição C1 temos que  $\rho_\beta(k) \rightarrow 0$ . Esta é uma condição suficiente para que  $\lim_{n \rightarrow +\infty} \sum_{k=-(n-1)}^{n-1} (1 - |k|n^{-1}) \rho_\beta(k) < \infty$ . Assim, concluímos que  $\text{var}(\hat{\mu}) \rightarrow 0$ , quando  $n \rightarrow +\infty$ .  $\square$

A estimação dos parâmetros autoregressivos  $\phi' = [\phi_1, \phi_2, \dots, \phi_p]$  é baseada na estrutura das covariâncias do processo  $\{h_t^{-1}Y_t\}_{t \in \mathbb{N}}$ . De facto, pela condição C2, para  $k \geq 1$ , temos

$$\begin{aligned} \gamma_{h^{-1}Y} &= \text{cov}(\beta_t + h_t^{-1}e_t, \beta_{t+k} + h_{t+k}^{-1}e_{t+k}) \\ &= \gamma_\beta(k). \end{aligned} \quad (6)$$

Por uma questão de simplificação da notação, adoptamos  $\gamma(k) = \gamma_\beta(k) = \gamma_{\frac{Y}{h}}(k)$ , para  $k \geq 1$ . Atendendo à condição de estacionaridade C1 e a (6), para  $k \geq p+1$ , temos

$$\gamma(k) = \phi_1\gamma(k-1) + \phi_2\gamma(k-2) + \dots + \phi_p\gamma(k-p)$$

pelo que, para  $p+1 \leq k \leq \ell$ , vem

$$\begin{aligned} \begin{bmatrix} \gamma(p+1) \\ \gamma(p+2) \\ \vdots \\ \gamma(\ell) \end{bmatrix} &= \begin{bmatrix} \gamma(p) & \gamma(p-1) & \cdots & \gamma(1) \\ \gamma(p+1) & \gamma(p) & \cdots & \gamma(2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(\ell-1) & \gamma(\ell-2) & \cdots & \gamma(\ell-p) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \\ \gamma(\ell) &= \mathbf{G}(\ell)\phi \end{aligned} \quad (7)$$

Em analogia com as equações de Yule-Walker, podemos obter um estimador para  $\phi$  resolvendo o sistema (7) em ordem aos parâmetros, com  $\ell = p+1$ . Contudo, como veremos mais adiante, é possível obter estimadores com melhores propriedades considerando  $\ell > p+1$  e aplicando a regressão linear múltipla. Assim, obtemos o estimador para  $\phi$  dado por

$$\hat{\phi} = \left( \hat{\mathbf{G}}'(\ell)\hat{\mathbf{G}}(\ell) \right)^{-1} \hat{\mathbf{G}}'(\ell)\hat{\gamma}(\ell) \quad (8)$$

com  $\ell \geq p+1$  e

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (h_t^{-1}Y_t - \hat{\mu}) (h_{t+k}^{-1}Y_{t+k} - \hat{\mu}). \quad (9)$$

Para provar a consistência de  $\phi$  basta provar que o estimador da função de autocovariância  $\hat{\gamma}(k)$  definido por (9) é consistente para  $\gamma(k)$ .

**Teorema 2.2.** *Considere-se o modelo em espaço de estados definido pelas equações (3)-(4) onde se verificam as condições C1, C2 e C3. Se existir  $c > 0$ , tal que  $|h_t| > c$  para todo o  $t \in \mathbb{N}$ , então  $\hat{\gamma}(k)$  é consistente para  $\gamma(k)$ .*

*Demonstração.* Atendendo às relações

$$\sum_{t=1}^{n-k} (h_t^{-1}Y_t - \hat{\mu}) \approx \sum_{t=1}^{n-k} (h_{t+k}^{-1}Y_{t+k} - \hat{\mu}) \approx \sum_{t=1}^n (h_t^{-1}Y_t - \hat{\mu})$$

e após alguma manipulação algébrica, temos

$$\sum_{t=1}^{n-k} (h_t^{-1}Y_t - \hat{\mu}) (h_{t+k}^{-1}Y_{t+k} - \hat{\mu}) \approx \sum_{t=1}^{n-k} (h_t^{-1}Y_t - \mu) (h_{t+k}^{-1}Y_{t+k} - \mu) - (n-k)(\hat{\mu} - \mu)^2$$

obtendo-se a igualdade quando  $n \rightarrow +\infty$ . Como  $|h_t| > c$  para todo o  $t \in \mathbb{N}$ , pelo teorema 2.1, concluímos que  $\hat{\mu}$  é consistente para  $\mu$ . Assim, basta provar que  $\hat{\gamma}^*(k)$  é consistente para  $\gamma(k)$ , onde

$$\hat{\gamma}^*(k) = \frac{1}{n} \sum_{t=1}^{n-k} (h_t^{-1}Y_t - \mu) (h_{t+k}^{-1}Y_{t+k} - \mu).$$

É fácil verificar que

$$\begin{aligned} \hat{\gamma}^*(k) &= \frac{1}{n} \left[ \sum_{t=1}^{n-k} (\beta_t - \mu)(\beta_{t+k} - \mu) + \sum_{t=1}^{n-k} (\beta_t - \mu) \frac{e_{t+k}}{h_{t+k}} + \right. \\ &\quad \left. + \sum_{t=1}^{n-k} (\beta_{t+k} - \mu) \frac{e_t}{h_t} + \sum_{t=1}^{n-k} \frac{e_t e_{t+k}}{h_t h_{t+k}} \right]. \end{aligned} \quad (10)$$

A primeira parcela representa o estimador da função de autocovariância,  $\hat{\gamma}(k)$ , de um processo autoregressivo  $AR(p)$  que, sob as condições C1 e C3, é consistente para  $\gamma(k)$  (ver [7], pp.192 e [4], pp. 321). As segunda e terceira parcelas têm o mesmo comportamento assintótico, pelo que basta analisar, por exemplo, a segunda. Temos

$$\text{var} \left( \frac{1}{n} \sum_{t=1}^{n-k} (\beta_t - \mu) \frac{e_{t+k}}{h_{t+k}} \right) = \frac{\sigma_e^2 \sigma_\beta^2}{n^2} \sum_{t=1}^{n-k} \frac{1}{h_{t+k}^2}.$$

Por hipótese,  $|h_t| > c$  para todo o  $t \in \mathbb{N}$ , então  $\text{var} \left( n^{-1} \sum_{t=1}^{n-k} (\beta_t - \mu) \frac{e_{t+k}}{h_{t+k}} \right) \rightarrow 0$ , quando  $n \rightarrow +\infty$ . Assim, concluímos que as segunda e terceira parcelas de (10) convergem em probabilidade para zero. A quarta parcela tem média nula e variância igual a

$$\text{var} \left( \frac{1}{n} \sum_{t=1}^{n-k} \frac{e_t e_{t+k}}{h_t h_{t+k}} \right) = \frac{\sigma_e^4}{n^2} \sum_{t=1}^{n-k} \frac{1}{h_t^2 h_{t+k}^2}.$$

Como  $|h_t| > c$ , para todo o  $t \in \mathbb{N}$ , temos que  $\text{var} \left( n^{-1} \sum_{t=1}^{n-k} \frac{e_t e_{t+k}}{h_t h_{t+k}} \right) \rightarrow 0$ , isto é, a quarta parcela converge em probabilidade para zero. Assim, concluímos que  $\hat{\gamma}(k)$  é consistente para  $\gamma(k)$ .  $\square$

Deste teorema decorre, trivialmente, o seguinte corolário.

**Corolário 2.1.** *Nas condições do teorema 2.2 o estimador  $\hat{\phi}$  dado por (8) é consistente para  $\phi$ .*

Na prática, a escolha de  $\ell$  para (8) é uma questão importante. Neste sentido procedemos a um estudo preliminar de simulação considerando o modelo (3)-(4) com  $p = 1$ , erros gaussianos,  $h_t = 1$  para todo o  $t$  e média  $\mu = 0$ . Para cada combinação  $(\phi, \sigma_\varepsilon^2, \sigma_\varepsilon^2)$  foram obtidas 2000 amostras de dimensões 50, 100, 200 e 500, cujas estimativas de  $\phi$ , para os vários valores  $\ell = \ell_1, \ell_2, \dots$ , se encontram dentro do intervalo  $] -1, 1[$ . Para comparar o desempenho dos estimadores, calculámos as raízes quadradas dos erros quadráticos médios amostrais das 2000 estimativas de  $\hat{\phi}$  relativas a cada  $\ell_i$ . A partir deste estudo de simulação concluímos que a escolha de  $\ell$  deve ser baseada na dimensão da amostra. Como regra geral, sugerimos que os melhores resultados são obtidos tomando  $\ell = 45$ ,  $\ell = 80$ ,  $\ell = 60$  e  $\ell = 50$  para  $n = 50$ ,  $n = 100$ ,  $n = 200$  e  $n = 500$ , respectivamente.

Sob a condição C1, temos

$$\gamma_\beta(0) = \frac{\sigma_\varepsilon^2}{1 - \phi_1\rho_1 - \dots - \phi_p\rho_p}$$

pelo que a função de autocovariância,  $\gamma(k)$ , pode ser escrita como função de  $\phi = (\phi_1, \phi_2, \dots, \phi_p)'$  e  $\sigma_\varepsilon^2$  na forma  $\gamma(k) = \sigma_\varepsilon^2 f_k(\phi)$ .

Se tomarmos  $k = 1, 2, \dots, \ell_\varepsilon$  temos  $\ell_\varepsilon$  equações a partir das quais podemos construir um estimador para  $\sigma_\varepsilon^2$ , através da regressão linear sem termo independente,

$$\hat{\sigma}_\varepsilon^2 = \left( \sum_{k=1}^{\ell_\varepsilon} \hat{\gamma}(k) f_k(\hat{\phi}) \right) \left( \sum_{k=1}^{\ell_\varepsilon} f_k^2(\hat{\phi}) \right)^{-1}. \quad (11)$$

A consistência do estimador  $\hat{\sigma}_\varepsilon^2$  fica estabelecida a partir da consistência do estimador da função de autocovariância  $\hat{\gamma}_k$  e do estimador dos parâmetros autoregressivos  $\hat{\phi}$ , uma vez que é apenas função destes. Assim, a consistência de  $\hat{\sigma}_\varepsilon^2$  vem naturalmente como corolário do teorema 2.2, como segue.

**Corolário 2.2.** *Nas condições do teorema 2.2 temos que  $\hat{\sigma}_\varepsilon^2$  dado por (11) é consistente para  $\sigma_\varepsilon^2$ .*

Para analisar o número de equações,  $\ell_\varepsilon$ , procedemos a um estudo de simulação similar ao estudo realizado para  $\ell$ , para estimar  $\phi$ . Concluímos que, nos cenários simulados, os melhores resultados são obtidos quando  $\ell_\varepsilon = 1$  para todas as dimensões das amostras.

Para estimar a variância do erro da equação de observação  $\sigma_\varepsilon^2$ , podemos recorrer à igualdade

$$E \left[ \frac{1}{n} \sum_{t=1}^n (h_t^{-1} Y_t - \mu)^2 \right] = \sigma_\beta^2 + \sigma_\varepsilon^2 \frac{1}{n} \sum_{t=1}^n h_t^{-2}.$$

Assim, um estimador natural para  $\sigma_e^2$  é

$$\hat{\sigma}_e^2 = \left[ \sum_{t=1}^n (h_t^{-1} Y_t - \hat{\mu})^2 - n\hat{\sigma}_\beta^2 \right] \left( \sum_{t=1}^n h_t^{-2} \right)^{-1} \quad (12)$$

Para analisar a consistência do estimador (12) vamos começar por considerar que  $\mu$  e  $\sigma_\beta^2$  são conhecidos,

$$\hat{\sigma}_e^{*2} = \left[ \sum_{t=1}^n (h_t^{-1} Y_t - \mu)^2 - n\sigma_\beta^2 \right] \left( \sum_{t=1}^n h_t^{-2} \right)^{-1}.$$

O teorema seguinte estabelece condições suficientes para a consistência de  $\hat{\sigma}_e^{*2}$ .

**Teorema 2.3.** *Considere-se o modelo em espaço de estados definido pelas equações (3)-(4) onde as condições C1, C2 e C3 são verificadas. Se existirem  $c_1 > 0$  e  $c_2 > 0$  tais que  $c_1 < |h_t| < c_2$  para todo o  $t \in \mathbb{N}$ , então  $\hat{\sigma}_e^{*2}$  é consistente para  $\sigma_e^2$ .*

*Demonstração.* Pela sua própria construção  $\hat{\sigma}_e^{*2}$  é centrado para  $\sigma_e^2$  e a sua variância é igual a

$$\text{var}(\hat{\sigma}_e^{*2}) = \left( \sum_{t=1}^n h_t^{-2} \right)^{-2} \text{var} \left[ \sum_{t=1}^n (h_t^{-1} Y_t - \mu)^2 \right]$$

mas

$$\begin{aligned} \text{var}(\hat{\sigma}_e^{*2}) &= \left( \sum_{t=1}^n h_t^{-2} \right)^{-2} \left\{ \text{var} [(\beta_t - \mu)^2] + \text{var} \left( \sum_{t=1}^n h_t^{-2} e_t^2 \right) + \right. \\ &\quad \left. + 4\text{var} \left[ \sum_{t=1}^n h_t^{-1} e_t (\beta_t - \mu) \right] \right\} \\ &= \left( \sum_{t=1}^n h_t^{-2} \right)^{-2} \left\{ \text{var} \left[ \sum_{t=1}^n (\beta_t - \mu)^2 \right] + \tau_e \sum_{t=1}^n h_t^{-4} + \sigma_e^2 \sigma_\beta^2 \sum_{t=1}^n h_t^{-2} \right\}. \end{aligned}$$

Se atendermos a que, sob a condição C3,  $\text{var} \left[ \sum_{t=1}^n (\beta_t - \mu)^2 \right]$  é constante, e que, por hipótese,  $c_1 < |h_t| < c_2$  concluímos que  $\text{var}(\hat{\sigma}_e^{*2}) \rightarrow 0$ , quando  $n \rightarrow +\infty$ .  $\square$

Como corolário deste teorema obtemos condições suficientes para a consistência de  $\hat{\sigma}_e^2$  como segue.

**Corolário 2.3.** *Sob as condições do teorema 2.3, se  $\mu$ ,  $\phi$  e  $\sigma_\varepsilon^2$  são substituídos por estimadores consistentes, por exemplo (5), (8) e (12), então  $\hat{\sigma}_e^2$  é consistente para  $\sigma_e^2$ .*

## 2.2 Estimadores propostos em Alpuim (1999)

Nesta secção expomos os estimadores para as variâncias dos erros propostos em [1]. Por simplicidade, referimos apenas os estimadores para o modelo (3)-(4) com  $p = 1$ . Estes estimadores são consistentes e assumem que os restantes parâmetros são conhecidos. Neste trabalho optamos por adoptar estes estimadores associados aos estimadores para a média e para os parâmetros autoregressivos propostos na secção anterior. Assim, obtemos os estimadores

$$\hat{\sigma}_\varepsilon^2 = \frac{\bar{D}(k)\Lambda(\ell) - \bar{D}(\ell)\Lambda(k)}{\Psi(k)\Lambda(\ell) - \Psi(\ell)\Lambda(k)} \text{ e } \hat{\sigma}_e^2 = \frac{\bar{D}(\ell)\Lambda(k) - \bar{D}(k)\Lambda(\ell)}{\Psi(k)\Lambda(\ell) - \Psi(\ell)\Lambda(k)} \quad (13)$$

onde

$$\bar{D}(k) = \frac{1}{n-k} \sum_{t=1}^{n-k} \left[ (h_{t+k}^{-1} Y_{t+k} - \hat{\mu}) - \hat{\phi}^k (h_t^{-1} Y_t - \hat{\mu}) \right],$$

$$\Psi(k) = \frac{1 - \hat{\phi}^{2k}}{1 - \hat{\phi}^2} \text{ e } \Lambda(k) = 1 + \hat{\phi}^{2k}.$$

Os valores  $\ell$  e  $k$  são arbitrários. Contudo, a autora sugere  $\ell = 1$  e  $k = 2$ . Para mais detalhes e suas aplicações ver [2] e [3].

## 3 Estudo comparativo

A consistência dos estimadores propostos neste trabalho é uma propriedade importante do ponto de vista estatístico mas que, em termos práticos, para uma dada aplicação torna-se menos interessante. Do ponto de vista prático, tem enorme relevância comparar os métodos *distribution-free* propostos com o método da máxima verosimilhança avaliando a eficiência relativa, não excluindo uma possível ponderação quanto à exequibilidade de cada método.

Neste sentido, torna-se útil o delineamento de um estudo de simulação onde se compare o desempenho de ambos os métodos de estimação. O desempenho será aferido por duas vias distintas, mas cuja complementaridade é notória. Numa primeira abordagem, são calculadas as raízes dos erros quadráticos médios amostrais (REQM) das estimativas dos parâmetros pelos vários métodos de modo a permitir uma análise sobre a eficiência relativa. Numa segunda perspectiva, são avaliadas as percentagens das amostras cujas estimativas, por cada método, se encontram no espaço de parâmetros.

O estudo de simulação via Monte Carlo tem por base modelos invariantes da forma (3)-(4) com  $p = 1$ ,  $h_t = 1$  para todo o  $t$ , e onde o estado tem média nula. São escolhidas várias combinações de parâmetros  $(\phi, \sigma_e^2, \sigma_\varepsilon^2)$  e cada amostra é obtida a partir de um valor inicial com distribuição  $\beta_1 \sim N(0, \sigma_\varepsilon^2/(1 - \phi^2))$ . Os erros são gaussianos com distribuições  $e_t \sim N(0, \sigma_e^2)$  e  $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ .

O estudo abrange amostras com dimensões 50, 100, 200 e 500 para as quais são estimados os parâmetros por três métodos: MV (máxima verosimilhança),

Tabela 1: Raízes quadradas dos erros quadráticos médios (REQM) dos estimadores da máxima verosimilhança (MV) e pelos métodos *distribution-free* M1 e M2 para amostras de dimensão 50.).

Parâmetros		REQM									
$\sigma_\varepsilon^2$	$\sigma_\varepsilon^2$	$\hat{\mu}$		$\hat{\phi}$		$\hat{\sigma}_\varepsilon^2$			$\hat{\sigma}_\varepsilon^2$		
		MV	M1	MV	M1	MV	M1	M2	MV	M1	M2
$\phi = ,50$											
,05	,10	,095	,095	,180	,182	,035	,032	,032	,045	,042	,042
,10	,50	,204	,204	,173	,202	,160	,186	,168	,216	,251	,226
,10	1	,298	,298	,173	,207	,312	,394	,342	,437	,532	,473
1	1	,319	,591	,234	,191	,712	,381	,436	,821	,398	,459
2	5	,654	1,261	,194	,187	1,757	1,705	1,559	2,267	2,101	1,893
2	10	,880	1,611	,181	,177	2,625	3,697	3,147	3,673	4,618	3,995
$\phi = ,75$											
,05	,10	,164	,165	,137	,108	,030	,030	,023	,043	,039	,040
,10	,50	,376	,383	,119	,099	,097	,151	,111	,167	,215	,177
,10	1	,537	,567	,104	,091	,208	,317	,237	,346	,455	,379
1	1	,545	,546	,163	,141	,492	,338	,400	,620	,375	,473
2	5	1,242	1,285	,139	,107	1,367	1,503	1,324	1,925	1,857	1,782
2	10	1,678	1,769	,121	,099	2,113	3,022	2,311	3,357	4,203	3,565
$\phi = ,90$											
,05	,10	,391	,391	,120	,096	,026	,028	,027	,040	,036	,040
,10	,50	,866	,869	,112	,088	,083	,123	,092	,158	,184	,171
,10	1	1,221	1,222	,107	,076	,152	,260	,168	,313	,387	,339
1	1	1,228	1,228	,145	,128	,455	,358	,450	,570	,421	,543
2	5	2,862	2,859	,123	,091	1,193	1,326	1,211	1,781	1,725	1,799
2	10	3,819	3,818	,111	,080	1,690	2,661	1,809	3,146	3,615	3,243

M1 (constituído pelos estimadores aqui propostos (5), (8), (11) e (12)) e M2 (combinação dos estimadores (5), (8) com os estimadores (13) propostos em Alpuim (1999)).

Na tabela 1 encontram-se as raízes quadradas dos erros quadráticos médios amostrais relativas a 1000 réplicas válidas <sup>4</sup>de amostras de dimensão 50, isto é,

$$\text{REQM}(\theta_i) = \sqrt{\frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_i - \theta_i)^2}.$$

Por uma questão de espaço optámos por apresentar apenas os resultados obtidos para amostras de dimensão 50, uma vez que há um especial interesse em comparar o desempenho dos estimadores em amostras de menor dimensão.

Os resultados obtidos relativamente à estimação do valor médio  $\mu$  pela máxima verosimilhança e pelo estimador *distribution-free* são muito semelhantes. De facto, embora não estejamos com amostras aleatórias simples de inovações, uma vez que  $\text{var}(Y_{t|t-1}) = P_{t|t-1} + \sigma_\varepsilon^2$ , os modelos estabelecidos são invariantes e com estado estacionário pelo que, após algumas iterações, faz sentido considerar que o filtro de Kalman se encontra em regime estacionário ( $P_{t|t-1} \rightarrow P$ , quando  $t \rightarrow +\infty$ ).

<sup>4</sup>Note-se que, neste contexto, interpretamos que  $\text{REQM}(\theta_i) \equiv \text{REQM}(\theta_i | \Theta \in \Omega(\Theta))$ .

Tabela 2: Percentagens de estimativas dos parâmetros  $(\phi, \sigma_e^2, \sigma_\varepsilon^2)$  que se encontram dentro do espaço de parâmetros para amostras de dimensão 50, 100, 200 e 500.

Parâmetros			%												
$\phi$	$\sigma_e^2$	$\sigma_\varepsilon^2$	$n = 50$			$n = 100$			$n = 200$			$n = 500$			
			MV	M1	M2	MV	M1	M2	MV	M1	M2	MV	M1	M2	
,50	,05	,10	27	89	79	56	96	93	69	98	96	86	99	98	
		,50	29	91	78	55	98	94	61	97	95	74	96	95	
	,10	1	30	90	76	44	98	91	56	97	93	65	94	91	
		1	1	19	86	76	37	93	90	71	96	96	89	98	98
		2	5	26	89	79	50	97	94	70	97	96	85	98	98
	2	10	32	90	78	47	98	93	63	97	94	78	97	95	
,75	,05	,10	54	96	87	81	100	98	96	100	100	100	100	100	
		,50	49	94	79	71	99	93	81	98	96	96	99	99	
	,10	1	46	92	71	58	99	85	67	97	89	82	95	93	
		1	1	45	96	88	85	100	98	99	100	100	100	100	
		2	5	53	96	85	82	100	97	93	100	99	99	100	100
	2	10	51	94	79	72	99	92	83	99	95	95	99	98	
,90	,05	,10	74	98	90	94	100	98	100	100	100	100	100	100	
		,50	64	91	76	79	99	88	92	97	96	100	98	100	
	,10	1	50	86	64	68	96	80	78	90	84	91	90	92	
		1	1	71	99	92	98	100	98	100	100	100	100	100	
		2	5	70	96	87	92	99	98	99	100	100	100	100	
	2	10	61	91	76	84	99	91	93	97	95	99	98	100	

Contrariamente ao que seria de esperar, o estimador *distribution-free* para  $\phi$  evidenciou um melhor desempenho em diversos cenários que o estimador da máxima verosimilhança. É de salientar que existe um padrão comum no desempenho do método M1 para todas as dimensões das amostras. Como o estimador *distribution-free* é baseado na estrutura de covariâncias de  $\{h_t^{-1}Y_t\}_{t \in \mathbb{N}}$  apresenta melhores resultados quanto mais próximo de 1 for  $\phi$ . Este comportamento também se verifica na estimação pela máxima verosimilhança, embora de uma forma menos acentuada. Esta característica mais vincada no método M1 não constitui uma desvantagem, uma vez que, se um conjunto de dados não apresentar uma correlação temporal significativa, os modelos em espaço de estados poderão não constituir a melhor opção.

O estudo de simulação não permitiu concluir que um dos métodos analisados produz melhores resultados do que os restantes no que diz respeito à estimação das variâncias dos erros  $\sigma_e^2$  e  $\sigma_\varepsilon^2$ . No entanto, os resultados obtidos indicam que os estimadores propostos em [1] apresentam, regra geral, um melhor desempenho que os do método M1, e não muito diferentes da estimação pela máxima verosimilhança.

Uma análise, possivelmente mais interessante, consiste na comparação das taxas de sucesso na estimação dos parâmetros (ver tabela 2). Se, por um lado, a estimação pela máxima verosimilhança apresenta taxas de sucesso muito baixas nas amostras de dimensão mais baixa, com os métodos *distribution-free* obtemos taxas bastante elevadas, mesmo para  $n = 50$ , em especial com o método M1. Este é, porventura, o resultado mais relevante deste trabalho. Associando-se o

bom desempenho dos estimadores *distribution-free*, em termos da REQM, aos bons resultados em termos das taxas de sucesso na estimação dos parâmetros, estes constituem uma relevante alternativa à complexidade computacional e aos possíveis problemas de identificação da estimação pela máxima verosimilhança.

## Agradecimentos

Este trabalho foi parcialmente suportado pelo PRODEP III - acção 5.3 .

## Referências

- [1] Alpuim, T. (1999). Noise variance estimators in state space models based on the method of moments. *Annales de l'I.S.U.P.*, Vol. XXXXIII, Fasc. 2-3, p. 3–23.
- [2] Alpuim, T., Barbosa, S. (1999). The Kalman filter in the estimation of area precipitation. *Environmetrics*, Vol. 10, p. 377–394.
- [3] Alpuim, T., Ribeiro, I. (2003). A state space model for run-off triangles. *Appl. Stochastics Models Bus. Ind.*, Vol. 19, p. 105–120.
- [4] Brockwell, J.P., Davis, R.A. (1991). *Time series: theory and methods*. 2ª Edição. Springer.
- [5] Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society*, Vol. 39 (série B), p. 1–38.
- [6] Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models. *J. Amer. Stat. Assoc.*, Vol. 87, p. 501–509.
- [7] Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.
- [8] Harvey, A.C. (1996). *Forecasting structural time series models and the Kalman filter*. Cambridge University Press.
- [9] Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions ASME*, Vol. 82 (série D), p. 35–45.
- [10] Kitagawa, G. (1987). Non-Gaussian State-Space Modeling of Nonstationary Time Series. *J. Amer. Stat. Assoc.*, Vol. 82, p. 1032–1063.
- [11] Shephard, N., Pitt, M. (1997). Likelihood analysis of non-Gaussian measurement times series. *Biometrika*. Vol. 84, p. 653–667.
- [12] West, M., Harrison, P.J. (1989). *Bayesian forecasting and dynamic models*. Springer-Verlag.
- [13] West, M., Harrison, P.J., Migon, H.S. (1985). Dynamic generalized linear models and Bayesian forecasting. *J. Amer. Stat. Assoc.* 87, 501–509 , Vol. 80, p. 73-97.