

Séries temporais para modelos com parâmetros aleatórios: uma abordagem bayesiana

Leonilce Mena¹ e Marinho Gomes de Andrade Filho²

¹Departamento de Estatística, Universidade Estadual de Maringá, Av. Colombo, 5790, 87020-900, Maringá - Paraná, Brasil.

²Instituto de Computação de Matemática Computacional, Universidade de São Paulo, C.P. 668. CEP 13560-970, São Carlos - São Paulo, Brasil. *Autor para correspondência. e-mail: lmena@uem.br

RESUMO. Este trabalho apresenta uma abordagem bayesiana para fazer inferência sobre os parâmetros de modelos auto-regressivos. Nesse contexto, quando os parâmetros variam de forma aleatória e independente, adotou-se um modelo hierárquico para descrever a densidade *a posteriori*. Uma segunda abordagem supõe que os parâmetros variam de acordo com um modelo auto-regressivo de primeira ordem; nesse caso a abordagem é vista como uma extensão do filtro de Kalman, no qual as variâncias dos ruídos são conhecidas. Os modelos foram analisados usando técnicas de simulação de Monte Carlo e a geração de amostras das densidades *a posteriori* permitiram fazer previsões de séries por intermédio das densidades preditivas. Para exemplificar a aplicação dos modelos, consideramos uma série real correspondente ao preço de ações.

Palavras-chave: processo auto-regressivo, inferência bayesiana, modelo hierárquico, modelo dinâmico, filtro de Kalman, Gibbs-Sampling, Metropolis-Hasting.

ABSTRACT. Temporal series for random parameter models: a bayesian approach.

This paper presents a Bayesian approach to make inference about the parameters of autoregressive models. In this context, when the parameters of models are independent and vary at random we consider a hierarchical model to describe the *a posteriori* density of parameters. A second approach assumes that the parameters of model vary according to a first order autoregressive model. In this case, the proposed approach is seen as an extension of Kalman filter where the variances of noises are known. The models were analysed using Monte Carlo simulation techniques and the resulting samples of *a posteriori* densities allowed to foresee a data series through predictable densities. Illustrations with actual data of a financial series are showed and the two models are evaluated by the quality of the prediction obtained, emphasizing the best model which represents the data.

Key words: autoregressive process, Bayesian inference, dynamic model, hierarchical model, Kalman filter, Gibbs-Sampling, Metropolis-Hasting.

Introdução

Neste trabalho, o objetivo é formular um procedimento bayesiano, utilizando simulação de Monte Carlo via Cadeias de Markov para análise e previsão de séries temporais modeladas por processos auto-regressivos com coeficientes aleatórios denominados processos auto-regressivos generalizados.

Muitos métodos podem ser utilizados para estimar os parâmetros de um processo auto-regressivo na análise de uma série temporal, tais como: método dos mínimos quadrados, já amplamente estudados (Box *et al.*, 1994).

Um método alternativo de análise e inferência nos modelos de séries temporais são os métodos

bayesianos aplicados para séries com coeficientes constantes (Broemeling e Cook, 1993).

Neste trabalho, é aplicado o método Bayesiano para séries com coeficientes aleatórios, descrevendo a utilização e aplicação de um modelo hierárquico para modelar uma série temporal na qual as componentes não observáveis têm uma densidade de probabilidade normal cuja média também tem uma distribuição normal, assegurando-se, assim, que os coeficientes sejam variáveis aleatórias, mas não independentes. Descreve-se, também, a aplicação de um modelo dinâmico em problemas de previsões, na qual a série é modelada por uma média que varia no tempo, superposta a um ruído aditivo. Essa média é, por hipótese, uma combinação linear de funções conhecidas cujos coeficientes são desconhecidos.

Ilustramos a utilidade de nossa técnica, aplicando o processo auto-regressivo de primeira ordem com coeficiente aleatório para modelagem de uma série de dados reais na área financeira correspondente ao preço de ações da Siderúrgica Vale do Rio Doce. Em nossa análise, mostramos o quanto esses modelos produzem proveitosas informações na previsão de observações futuras.

Descrição do modelo

Os modelos que estamos trabalhando consideram um processo estocástico X_t ; $t=1, 2, \dots$ com distribuição log-normal denotado por:

$$X_t = X_{t-1}^{\theta_t}; \quad t=1, 2, \dots \quad (1)$$

em que θ_t é um coeficiente cujos valores descrevem um crescimento ou decrescimento no processo. Para produzir uma generalização no modelo, introduzimos um erro multiplicativo δ_t , assim temos:

$$X_t = X_{t-1}^{\theta_t} \delta_t; \quad t=1, 2, \dots \quad (2)$$

Aplicando o logaritmo natural, em ambos os lados, em (2) e seja $\varepsilon_t = \log(\delta_t)$ nosso modelo torna-se:

$$y_t = \theta_t y_{t-1} + \varepsilon_t; \quad t=1, 2, \dots \quad (3)$$

em que, $y_t = \log(X_t)$ com y_t e ε_t normalmente distribuídos, ε_t tem média zero e variância τ^2 .

Em alguns problemas, é conveniente supormos que o parâmetro θ_t tem característica aleatória. (Diaz, 1990; Liu, 1995).

Assumindo os θ_t 's com densidade de probabilidade normal de média λ e variância γ^2 , a estratégia para assegurar que os θ_t 's sejam dependentes é assumir-se, também, uma distribuição de probabilidade para λ , isto é, que λ tem distribuição normal com média m e variância S^2 conhecidos.

Os modelos hierárquicos bayesianos têm sido utilizados em aplicações nas mais diversas áreas do conhecimento.

A análise desses modelos torna-se, geralmente, intratável quando não existe conjugação entre distribuição *a priori* e a função de verossimilhança. Nesses casos, métodos de aproximação numérica (Soyer, 1985) ou de simulação devem ser utilizados para a obtenção das distribuições *a posteriori* de interesse.

Aqui, através de simulação de Monte Carlo via Cadeia de Markov (MCMC), utilizando o algoritmo Gibbs Sampling (Casella e George, 1992), obtiveram-se as distribuições *a posteriori* das variáveis de interesse e fazemos inferência sobre os parâmetros e a previsão de observações futuras para um processo auto-regressivo de primeira ordem, com parâmetros aleatórios e variantes no tempo ajustado por um modelo hierárquico, considerando a variância do ruído conhecida. Desenvolvemos a extensão do modelo considerado por:

$$\begin{aligned} Y_t &= \theta_t y_{t-1} + \varepsilon_t; \quad \varepsilon_t \sim N(0, \tau^2); \quad (4) \\ &\quad \tau^2 \text{ conhecido} \\ \theta_t &= \lambda + \omega_t; \quad \omega_t \sim N(0, \gamma^2) \\ &\quad \gamma^2 \text{ conhecido} \\ \lambda &\sim N(m, S^2); \quad m, S^2 \text{ conhecidos} \end{aligned}$$

em que, ε_t e ω_t são independentes.

Seja a série $y^{(t)} = (y_1, y_2, \dots, y_t)$ e dado $y^{(t)}$, o objetivo é fazermos inferência sobre θ_t , λ e previsão de θ_{t+1} e da observação futura y_{t+1} .

Denotando a função densidade de θ_t , dado $y^{(t)}$ por $p(\theta_t | y^{(t)})$, calculou-se essa densidade considerando a função densidade de probabilidade:

$$f(y_t | y_{t-1}, \theta_t, \tau) \sim N(\theta_t y_{t-1}; \tau^2) \quad (5)$$

A densidade *a priori* para λ :

$$\Pi(\lambda) \sim N(m; S^2) \quad (6)$$

Obtemos os resultados:

A função de verossimilhança:

$$p(y^{(t)} | \lambda) = \prod_{i=2}^t \exp \left\{ -\frac{1}{2} \frac{(y_i - \lambda y_{i-1})^2}{\tau_i} \right\} \quad (7)$$

$$\tau_i = \gamma^2 y_{i-1}^2 + \tau^2$$

A densidade *a posteriori* de λ :

$$\lambda | y^{(t)} \sim N(m_t, S_t) \quad (8)$$

$$m_t = \frac{S_{t-1} y_t y_{t-1} + m_{t-1} \gamma_t}{S_{t-1} y_{t-1}^2 + \tau_t}$$

$$S_t = \frac{S_{t-1} \gamma_t}{S_{t-1} y_{t-1}^2 + \tau_t}$$

A densidade *a posteriori* de θ_t :

$$\begin{aligned} \theta_t | y^{(t)} &\sim N(\theta_t^*, \sum_t^*) \\ \theta_t^* &= \frac{\tau^2 m_t + \gamma^2 y_t y_{t-1}}{\tau_t} \\ \sum_t^* &= \frac{\tau^4}{\tau_t^2} S_t + \frac{\tau^2 \gamma^2}{\tau_t} \end{aligned} \quad (9)$$

A densidade preditiva, considerando k passos, para θ_{t+k} , $k \geq 1$, é dada por:

$$\theta_{t+k} \sim N(m_t, S_t + \gamma^2) \quad (10)$$

A média preditiva de y_{t+k} , dado $y^{(t)}$ pode ser escrita como:

$$\hat{y}_{t+k} = E(\lambda^k | y^t) y^t, \text{ ou seja;}$$

$$\hat{y}_{t+k} = m_t \hat{y}_{t+k-1} + (k-1) S_t \hat{y}_{t+k-2}; \quad k \geq 1 \quad (11)$$

Para o processo auto-regressivo de primeira ordem, com a variância do ruído (τ^2) desconhecida, descreve-se a incerteza sobre essa variância considerando uma distribuição *a priori* para τ :

$$\begin{aligned} y_t &= \theta_t y_{t-1} + \varepsilon_t; & \varepsilon_t &\sim N(0, \tau^2); \\ & & \tau^2 &\text{desconhecido} \\ \theta_t &= \lambda + \omega_t; & \omega_t &\sim N(\lambda, \gamma^2) \\ & & \gamma^2 &\text{conhecido} \\ \lambda &\sim N(m, S^2); & m \text{ e } S^2 &\text{conhecidos.} \end{aligned} \quad (12)$$

Seja a série $y^{(t)} = (y_1, y_2, \dots, y_t)$ e a sequência $\theta^{(t)} = (\theta_1, \dots, \theta_t)$, e dado $y^{(t)}$, nosso objetivo é fazer inferência sobre θ_t, τ, λ e previsão para θ_{t+1} e y_{t+1} .

Aplicando o raciocínio utilizado no caso da variância conhecida, com a função densidade de probabilidade:

$$f(y_t | \theta_t, \tau) \sim N(\theta_t y_{t-1}; \tau^2) \quad (13)$$

Considerando as densidades *a priori* para λ para τ :

$$\begin{aligned} \prod_1(\tau) &\sim IG(\alpha; \beta) \\ \prod_2(\lambda) &\sim N(m; S^2) \end{aligned} \quad (14)$$

Obteve-se:

A função de verossimilhança de (λ, τ) :

$$\begin{aligned} p(y^{(t)} | \lambda, \tau) &= \prod_{i=2}^t \frac{1}{(\tau_i(\tau))^{\frac{1}{2}}} \exp \\ &\left\{ -\frac{1}{2} \frac{(y_i - \lambda y_{i-1})^2}{\tau_i(\tau)} \right\} \end{aligned} \quad (15)$$

em que

$$\tau_i(\tau) = \gamma^2 y_{i-1}^2 + \tau^2$$

A distribuição conjunta *a posteriori* de λ e τ

$$\begin{aligned} p(\lambda, \tau | y^{(t)}) &\propto \frac{\tau^{-(\alpha+1)}}{(\tau_i(\tau))^{\frac{1}{2}}} \exp\left\{-\frac{\beta}{\tau}\right\} \times \\ &\exp\left\{-\frac{1}{2S_t(\tau)}(\lambda - m_t(\tau))^2\right\} \times \\ &\exp\left\{-\frac{1}{2} \sum_{i=2}^t \frac{y_{i-1}^2}{\tau_i(\tau)} \left(\frac{y_i}{y_{i-1}} - m\right)^2\right\} \end{aligned} \quad (16)$$

Distribuição marginal para τ :

$$p(\tau | y^{(t)}) \propto \frac{\left(\frac{1}{S^2} + \sum_{i=2}^t \frac{y_{i-1}^2}{\tau_i(\tau)} \right)^{-1, -(\alpha+1)}}{(\tau_i(\tau))^{\frac{1}{2}}} \times \quad (17)$$

$$\exp\left\{-\frac{\beta}{\tau}\right\} \exp\left\{-\frac{1}{2} \sum_{i=2}^t \frac{y_{i-1}^2}{\tau_i(\tau)} \left(\frac{y_i}{y_{i-1}} - m\right)^2\right\}$$

Distribuição condicional para λ :

$$p(\lambda | y^{(t)}, \tau) \propto \exp\left\{-\frac{1}{2S_t(\tau)}(\lambda - m_t(\tau))^2\right\} \quad (18)$$

$$S_t(\tau) = \left(\frac{1}{S^2} + \sum_{i=2}^t \frac{y_{i-1}^2}{\tau_i(\tau)} \right) - 1;$$

$$m_t(\tau) = \left(\frac{m}{S^2} + \sum_{i=2}^t \frac{y_i y_{i-1}}{\tau_i(\tau)} \right) S_t(\tau)$$

$$\tau_i(\tau) = \gamma^2 y_{i-1}^2 + \tau^2$$

Distribuição condicional para τ :

$$p(\tau | y^{(t)}, \lambda) \propto \frac{\tau^{-(\alpha+1)}}{(\tau_i(\tau))^{\frac{1}{2}}} \exp\left\{-\frac{\beta}{\tau}\right\} \times$$

$$\exp\left\{-\frac{1}{2S_t(\tau)}(\lambda - m_t(\tau))^2\right\} \times$$

$$\exp\left\{-\frac{1}{2} \sum_{i=2}^t \frac{y_{i-1}^2}{\tau_i(\tau)} \left(\frac{y_i}{y_{i-1}} - m\right)^2\right\} \quad (19)$$

Distribuição condicional para θ_t :

$$p(\theta_t | y_t, y_{t-1}, \lambda, \tau) \propto \exp\left\{-\frac{1}{2\sum_t} (\theta_t - \hat{\theta}_t)^2\right\} \quad (20)$$

em que

$$\hat{\theta}_t = E(\theta_t | y^{(t)}) = E\left[\frac{\tau^2 \lambda + \gamma^2 y_t y_{t-1}}{\gamma^2 y_{t-1}^2 + \tau^2}\right]$$

$$\hat{\Sigma}_t = V(\theta_t | y^{(t)}) = E\left[\frac{\tau^2 \gamma^2}{\gamma^2 y_{t-1}^2 + \tau^2}\right]$$

$$V\left[\frac{\tau^2 \lambda + \gamma^2 y_t y_{t-1}}{\gamma^2 y_{t-1}^2 + \tau^2}\right] = E\left[\frac{\tau^2 \lambda + \gamma^2 y_t y_{t-1}}{\gamma^2 y_{t-1}^2 + \tau^2}\right] -$$

$$+ [E(\theta_t | y^{(t)})]^2$$

para a amostra gerada $(\tau^{(j)}, \lambda^{(j)}, j = 1, 2, \dots, M)$.

Os valores de $E(\theta_t | y^{(t)})$ e $E(\lambda | y^{(t)})$ são calculados através de simulação de Monte Carlo via cadeia de Markov (MCMC).

Os valores esperados da preditiva para θ_{t+k} e y_{t+k} , $k \geq 1$ são dados por:

$$E(\theta_{t+k} | y^{(t)}) = E(\lambda | y^{(t)}) \quad (21)$$

$$V(\theta_{t+k} | y^{(t)}) = V(\lambda | y^{(t)}) + \gamma^2$$

$$E(y_{t+k} | y^{(t)}) = E(\lambda^k) y^t \quad (22)$$

em que

$$E(\lambda^k) = \frac{1}{M} \sum_{j=1}^M \lambda_j^k$$

$$\hat{S}_\lambda^2 = \frac{1}{M-1} \sum_{j=1}^M \left[(\lambda^{(j)})^2 - \hat{\lambda} \right]^2$$

$$\hat{\tau} = \frac{1}{M} \sum_{j=1}^M \tau^{(j)}$$

$$\hat{S}_\tau^2 = \frac{1}{M-1} \sum_{j=1}^M \left[(\tau^{(j)})^2 - \hat{\tau} \right]^2$$

em que M é o tamanho da amostra gerada para

$\lambda^{(j)}$ e $\tau^{(j)}$ através de simulação de Monte Carlo via Cadeias de Markov.

Como outra alternativa, apresentou-se uma ramificação do processo auto-regressivo de primeira ordem com coeficiente aleatório assumindo uma estrutura de dependência dos coeficientes do processo estocástico (θ_t^i, s) tão forte quanto sua permutabilidade. Descrevemos o modelo para uma série temporal em função de seus componentes não

observáveis θ_t' quando θ_t é expresso em função dos valores desses mesmos componentes no instante anterior.

Concentrou-se no procedimento pelo qual obtemos estimadores atualizados dos componentes não observáveis a todo instante de tempo a partir da informação dada pelo componente observável do sistema, ou seja, y_t . Devido, a natureza linear dinâmica do modelo, essa atualização seqüencial pode ser obtida representando o modelo sob a forma de espaço de estados e, então, utilizar o modelo filtro de Kalman (Kalman, 1960), cujo procedimento é recursivo.

O filtro de Kalman, segundo (Meinhold e Singpurwalla, 1990), proporciona a elaboração de um algoritmo de estimação recursivo que tem sido utilizado com sucesso por engenheiros na teoria de controle, e, mais recentemente, por estatísticos e econometristas. Descreveu-se a utilização do filtro de Kalman em problemas de previsão, nos quais a série temporal é modelada por uma média que varia no tempo, superposta a um ruído aditivo. Essa média é, por hipótese, uma combinação linear de funções conhecidas cujos coeficientes são desconhecidos. Desse modo, a série pode ser representada por um sistema linear cujo vetor de estados é formado pelos coeficientes desconhecidos e pelo valor da média do processo no instante t . Nessas circunstâncias, o filtro de Kalman pode ser utilizado para obter estimativas ótimas do vetor de estados, com a vantagem de permitir a variação dos coeficientes através do tempo.

Seja $y^{(t)} = (y_t, y_{t-1}, \dots, y_1)$, denotando os valores observados de uma variável de interesse para diferentes tempos, assumindo que y_t depende de uma quantidade não observável θ_t (y_t e θ_t podem ser escalares ou vetores). Aqui, consideramos θ_t como um vetor de parâmetros, com o objetivo de fazer inferência sobre θ_t .

Considerando que a relação entre y_t e θ_t é linear e especificada pela equação das observações e a característica dinâmica da variável não observável é θ_t , um modelo é caracterizado por (West e Harrison, 1997):

$$y_t = F_t' \theta_t + v_t \quad (23)$$

$$\theta_t = G_t \theta_{t-1} + w_t \quad (24)$$

em que:

y_t : equação observada;

F_t : uma quantidade conhecida, denominada, na literatura, como matriz de medida;

G_t : quantidade conhecida, denominada matriz de transição;

v_t : erro observado assumido, normalmente distribuído, com média zero e variância V_t conhecida;

w_t : erro do sistema com distribuição normal de média zero e variância W_t conhecida.

Por hipótese, v_t e w_t são considerados independentes entre si e entre estágios subseqüentes.

O objetivo, aqui, é analisar o comportamento do sistema descrito pelas equações (20) e (21) para avaliar a mudança de comportamento do vetor de parâmetros θ_t com o tempo.

Para atingir esse objetivo, empregou-se a técnica filtro de Kalman segundo Meinhold e Singpurwalla (1983), a qual permite atualizar passo-a-passo o valor das médias e matriz de covariância de θ_t a medida que é feita uma nova observação y_t . Esse procedimento de inferência de $\hat{\theta}_t$ (média de θ_t condicionada às observações y_t) e matriz de covariância Σ_t permite fazer previsões para a observação futura y_{t+1} .

O procedimento pode ser explicado se for considerado o tempo $(t-1)$ e observados os dados até esse tempo $y^{(t-1)} = (y_{t-1}, \dots, y_1)$.

No tempo $(t-1)$, o conhecimento sobre θ_{t-1} é dado pela distribuição de probabilidade de θ_{t-1} , ou seja:

$$(\theta_{t-1} | y^{(t-1)}) \sim N\left(\hat{\theta}_{t-1}, \sum_{t-1}\right) \quad (25)$$

que representa a distribuição *a posteriori* para θ_{t-1} .

Considerou-se o desenvolvimento para o tempo t em dois estágios:

1. Antes da observação y_t :

$$(\theta_t | y^{(t-1)}) \sim N\left(G_t \hat{\theta}_{t-1}, \sum_t\right) \quad (26)$$

sendo:

$$\hat{\theta}_t = G_t \hat{\theta}_{t-1}; \sum_t = G_t' \sum_{t-1} G_t + W_t$$

Após a observação y_t

A função de verossimilhança:

$$(e_t | \theta_t, y^{(t-1)}) \sim N[F_t(\theta_t - G_t \hat{\theta}_{t-1}), V_t] \quad (27)$$

A distribuição conjunta de θ_t e e_t , dado $y^{(t-1)}$ pode ser escrita como:

$$(\theta_t, e_t | y^{(t-1)}) \sim N \left[\begin{pmatrix} G_t \hat{\theta}_{t-1} \\ F_t \sum_t \end{pmatrix}, \begin{pmatrix} \sum_t F_t' & \sum_t F_t' \\ V_t + F_t \sum_t F_t' & \sum_t F_t' \end{pmatrix} \right] \quad (28)$$

A distribuição *a posteriori* de θ_t :

$$(\theta_t | e_t, y^{(t-1)}) \sim N(G_t \hat{\theta}_{t-1} + \sum_t F_t' (V_t + F_t \sum_t F_t')^{-1} e_t; \sum_t - \sum_t F_t' (V_t + F_t \sum_t F_t')^{-1} F_t \sum_t) \quad (29)$$

No modelo filtro de Kalman (Kalman, 1960), as quantidades $(F_t, G_t, V_t$ e $W_t)$ são consideradas conhecidas. Elas podem variar no tempo somente se a natureza da variação também for considerada conhecida. Em muitas aplicações, é difícil especificarmos algum ou todos esses parâmetros devido às informações inadequadas sobre o processo ou porque os parâmetros esperados são permutáveis, mas a natureza exata dessa permuta é desconhecida. Nesses casos, o filtro de Kalman é aplicado para estimar os parâmetros desconhecidos dos dados em cada estágio, cujos estimadores dos parâmetros podem ser obtidos na forma do estimador bayesiano.

Considerando todas as quantidades no modelo dado pela equações 23 e 24, como escalares e fazendo

$G_t = \lambda$, $F_t = y_{t-1}$, $V_t = \varepsilon_t$ e $W_t = \omega_t$ podemos descrever o modelo como:

$$\begin{aligned} y_t &= \theta_t y_{t-1} + \varepsilon_t \\ \theta_t &= \lambda \theta_{t-1} + \omega_t \end{aligned} \quad (30)$$

em que

$$\begin{pmatrix} \varepsilon_t \\ \omega_t \end{pmatrix} \sim N \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma^2 & 0 \\ 0 & \gamma^2 \end{pmatrix},$$

σ^2 e γ^2 conhecidos.

Por hipótese, ε_t e ω_t são independentes entre si e entre estágios e λ desconhecido é independente de ω_t .

Desde que λ é desconhecido, assumiu-se uma distribuição *a priori* para o mesmo. Dado os dados $y^{(t)} = (y_t, y_{t-1}, \dots, y_1)$, o objetivo é fazer inferência sobre θ_t e a previsão futura de y_{t+1} .

Pode-se escrever a função verossimilhança de λ como:

$$L(\lambda | y^{(t-1)}) = \prod_{i=1}^t (y_{i-1}^2 \tau_i + \sigma^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2(y_{i-1}^2 \tau_i + \sigma^2)} (y_i - \lambda \hat{\theta}_{i-1} y_{i-1})^2 \right\} \quad (31)$$

A distribuição conjunta de θ_t e e_t pode ser escrita como:

$$(e_t, \theta_t | y^{(t-1)}, \lambda) \sim N(\lambda \hat{\theta}_{t-1} + \tau_t y_{t-1} (\sigma^2 + y_{t-1}^2 \tau_t)^{-1} e^t; \tau_t - y_{t-1}^2 \tau_t (\sigma^2 + y_{t-1}^2 \tau_t)^{-1}) \quad (32)$$

A distribuição *a posteriori* de λ dado $y^{(t)}$:

$$\begin{aligned} p(\lambda | y^{(t)}) &\propto \prod_{i=1}^t (y_{i-1}^2 \tau_i + \sigma^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2(y_{i-1}^2 \tau_i + \sigma^2)} (y_i - \lambda \hat{\theta}_{i-1} y_{i-1})^2 \right\} \times \\ &S^{-1} \exp \left\{ -\frac{1}{2S^2} (\lambda - m)^2 \right\} \end{aligned} \quad (33)$$

A distribuição *a posteriori* de θ_t dado $y^{(t)}$ e λ :

$$p(\theta_t | y^{(t)}, \lambda) \propto \sum_t^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2 \sum_t} (\theta_t - \hat{\theta}_t)^2 \right\} \quad (34)$$

$$\hat{\theta}_t = \frac{\lambda \hat{\theta}_{t-1} \sigma^2 + \tau_t y_t y_{t-1}}{\sigma^2 + y_{t-1}^2 \tau_t} \quad (35)$$

$$\sum_t = \frac{\tau_t \sigma^2}{\sigma^2 + y_{t-1}^2 \tau_t} \quad (36)$$

Os valores de $E(\theta_t | y^{(t)})$ e $E(\lambda | y^{(t)})$ são calculados através de simulação de Monte Carlo em cadeia de Markov (MCMC) com amostras geradas pelos algoritmos Gibbs-Sampling e Metropolis-Hasting (Chib e Greenberg, 1995).

A distribuição preditiva para θ_{t+1} é dada por:

$$\left[(\theta_{t+1} | y^{(t)}) \sim N \left(\lambda \hat{\theta}_t, \lambda^2 \hat{\Sigma}_t + \gamma^2 \right) \right] \quad (37)$$

em que

$$\hat{\theta}_t = \frac{\lambda \hat{\theta}_{t-1} \sigma^2 + \tau_t y_t y_{t-1}}{\sigma^2 + y_{t-1}^2 \tau_t}$$

$$\tau_t = \lambda^2 \sum_{t-1} + \gamma^2$$

$$\sum_t = \frac{\tau_t \sigma^2}{\sigma^2 + y_{t-1}^2 \tau_t}$$

A distribuição preditiva para y_{t+1}

$$(y_{t+1} | y^{(t)}, \lambda) \sim N \left[\lambda y_t \hat{\theta}_t; y_t^2 \left(\lambda^2 \sum_t + \gamma^2 \right) + \sigma^2 \right] \quad (38)$$

Para exemplificar a aplicação desses modelos, considerou-se uma série real correspondente ao preço de ações da Siderúrgica Vale do Rio Doce, em dólares, negociadas no período de 02 de janeiro de 1996 a 1 de fevereiro de 1999, pela Bovespa (Bolsa de Valores de São Paulo).

A partir das condicionais *a posteriori* dos parâmetros de cada modelo, aplicando os algoritmos Gibbs-Sampling e Metropolis-Hasting, geraram-se amostras da série considerando 5 cadeias com 3000 iterações cada uma, e, para garantir a independência entre os valores simulados de cada parâmetro, tomamos 50% das iterações, demos um salto a cada 5 valores, selecionando, assim, uma amostra com 300 valores em cada cadeia, totalizando uma amostra final de 1500 valores.

A convergência das amostras geradas foi monitorada, utilizando o método proposto por Gelman e Rubin (1992), o qual se baseia na análise de variância.

Utilizando essa amostra, aproximou-se a densidade preditiva dos parâmetros de interesse de cada modelo, através dos estimadores de Monte Carlo e verificou-se graficamente qual o modelo que melhor representa os dados, em função dos valores previstos obtidos.

Na Tabela 1, pode-se observar os hiperparâmetros do modelo hierárquico considerando a variância conhecida, baseados em

uma análise preliminar (análise descritiva) dos dados da série de preço das ações da Siderúrgica Vale do Rio Doce, negociadas em dólares, pela Bovespa, no período de 2 de janeiro de 1996 a 1 de fevereiro de 1999, considerando *a priori* informativa normal para λ e θ_t .

Tabela 1. Valores *a Priori* para ajustar a Série

m	S	γ	τ
1,0	0,5	0,65	0,33

A Figura 1 ilustra o comportamento da série de dados reais:

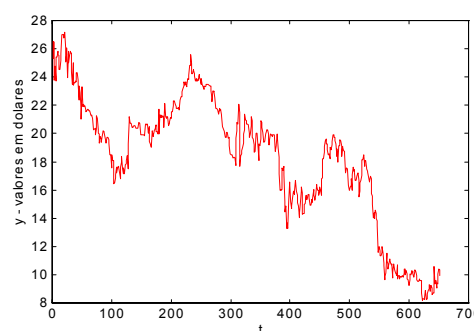


Figura 1. Evolução de preço diário de ações da Vale do Rio Doce.

Com o objetivo de compararmos a evolução de preços reais com os preços previstos, calculou-se a previsão desses preços para os últimos 30 dias do período e apresentamos os resultados obtidos nas Figuras 2 e 3.

Na Figura 2, é visualizado o comportamento dos valores previstos considerando um passo (dia) à frente, ou seja, $(y_{t+k} | y^{(t+k-1)}; k = 1, \dots, 30)$, enquanto que na Figura 3 visualiza-se tendência da previsão e dos valores reais para os 30 dias, considerando os dados observados até o instante t , isto é, $(y_t + k | y^{(t)})$, em que os 30 últimos dias foram deixados fora do conjunto de observações usadas no ajuste do modelo.

Observando as Figuras 2 e 3, pode-se constatar que os valores previstos, obtidos a partir da amostra gerada, possuem, basicamente, a mesma tendência que os valores reais; houve apenas uma diferença de locação nos valores previstos, que pode ser vista na Figura 2.

No cálculo dos valores previstos, considerando um passo (dia) à frente, obteve-se um erro médio de previsão (em percentual) $\bar{e}_t = 0,461\%$.

Erro esse inferior a 5%; assim, podemos concluir que o modelo apresentado é adequado para fazer

previsões de observações futuras de dados financeiros.

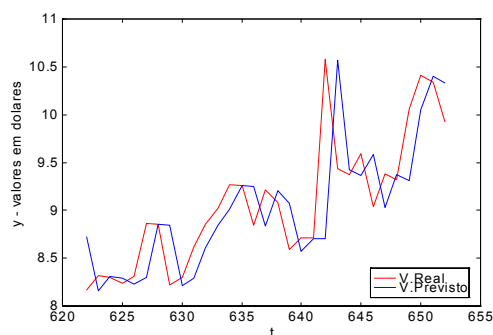


Figura 2. Comparação diária de preços reais e previstos de ações da Vale do Rio Doce

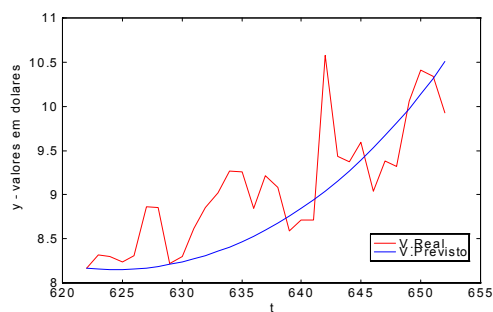


Figura 3. Tendência de preços de ações da Vale do Rio Doce em 30 dias consecutivos

Considerando o modelo hierárquico com a variância desconhecida, visualiza-se, na Figura 4, a evolução de preços reais e de preços previstos um passo (dia) à frente ($y_{t+k} | y^{(t+k-1)}$; $k=1, \dots, 30$), enquanto que na Figura 5 visualiza-se a tendência dos preços das ações da Siderúrgica Vale do Rio Doce, negociadas nos últimos 30 dias do período analisado, sendo que os valores de y_{t+k} são calculados considerando todos os dados anteriores, ou seja, $y_{t+k} | y^{(t)}$, observando que nas duas figuras, os últimos 30 dias foram deixados fora dos dados usados para o ajuste do modelo.

Na Figura 5, é possível observar que, pelo fato de se acrescentar uma incerteza na variância, ocorre uma discrepância nos valores previstos a partir da amostra gerada em relação aos valores reais dos dados. Fato este que pode ser confirmado pelo erro de previsão pois, nesse caso, tem-se um erro médio de $\bar{e}_t = 0,375\%$.

Na Figura 6, são comparados os preços reais e previstos, considerando um passo (dia) à frente,

($y_{t+k} | y^{(t+k-1)}$; $k=1, 2, \dots, 10$), onde os últimos 10 dias foram deixados fora do conjunto de observações usadas no ajuste do modelo dinâmico.

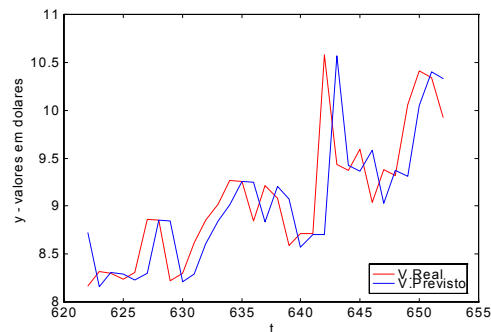


Figura 4. Comparação de preços reais e previstos nos últimos 30 dias do período

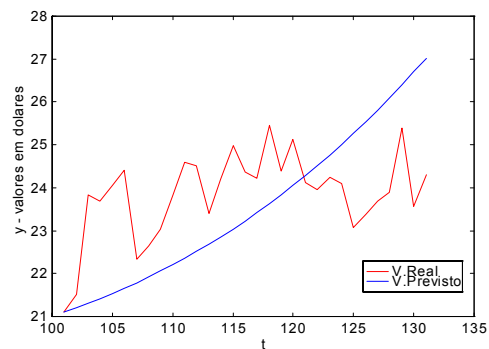


Figura 5. Tendência de preço real e preços estimados

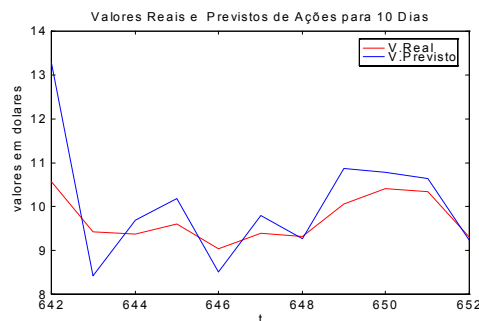


Figura 6. Comparação de preços real e preço estimados de ações da Vale do Rio Doce

Na Figura 6, é possível observar que os valores previstos estão bem próximos dos valores reais com pequeno erro de previsão, cujo valor médio percentual é: $\bar{e}_t = 5,23\%$.

Mediante os resultados obtidos nesse exemplo, vale ressaltarmos que a dependência das variáveis (coeficiente) do modelo auto-regressivo não apresenta mudanças muito acentuadas na previsão das observações futuras, mesmo quando acrescentamos uma incerteza na variância do ruído.

O uso de modelos hierárquico e dinâmico para modelagem de processos auto-regressivos com parâmetros aleatórios são alternativas eficientes para analisarmos e fazermos previsões futuras da evolução de valores econômicos.

Nesse caso, a análise bayesiana, usando método de simulação de Monte Carlo via Cadeia de Markov, tem implementação razoavelmente fácil, não exigindo desenvolvimento de aproximações numéricas, e as inferências obtidas são bastante precisas. Entretanto, conforme resultados gráficos obtidos, pode-se observar que o modelo hierárquico, no qual é considerado a variância do ruído (τ^2) conhecida, apresenta os valores previstos mais próximos dos valores reais no período considerado. A abordagem bayesiana é estendida para os modelos mais realistas, nos quais têm-se incertezas sobre as variâncias dos ruídos.

Nesses modelos mais gerais, os erros de previsão crescem como era de se esperar; mesmo assim, os modelos mostram-se mais adequados para a previsão de séries não estacionárias.

Referências

- BOX, G.E.P. *et al.* Time series analysis: forecasting and control. 3. ed. New Jersey: Holden Day, 1994.
- BROEMILING, D.L.; COOK, P. Bayesian estimation of the mean of an autoregressive process. *Journal of Applied Statistics*, Oklahoma, v.20, n1, p.25-38, 1993.
- CASELLA, G.; GEORGE, E. J. Explaining the Gibbs-Sampler, Belmont. *Amer. Statis. Soc.*, Alexandria, v. 46, n.3, p.167-174, 1992.
- CHIB, S.; GRENBERG, E.J. Understanding the Metropolis-Hasting Algorithm. *Amer. Statis. Soc.*, Alexandria, v. 49, n.4, p.35-51, 1995.
- DIAZ, J. Bayesian Forecasting for AR(1) Models With Normal Coefficients. *Commu. Statist.-Theory Meth.*, Voorburg, v.196, p.2229-2246, 1990.
- GELMAN, A.; RUBIN, B.D. Inference from Iterative Simulation Using Multiple Sequences. *Stat. Sci.*, Hayward, v.4, p.457-511, 1992.
- KALMAN, R.E. A New Approach to Linear Filtering and Prediction Problem. *Journal of Basic Eng, Trans. ASME, Series. D*, New York, v. 82, p 35-45, 1960.
- LIU, S.I. Comparison of Forecasts for Arma Models Between a Random Coeficent Approach and a Bayesian Approach. *Commu. Statist Theory Meth.*, Voorburg, v. 24, n.2, p.319-333, 1995.
- MEINHOLD, R.J.; SINGPURVALLA, N.D. Understanding the Kalman Filter, *Amer. Statis. Soc.*, Alexandria, v. 37, n, p.229-246, 1983.
- SOYER, R. *Random coefficient autoregressive process and their ramifications: applications to reliability growth. assessment.* 1985. (Dissertation) - School Eng. Appl. Sci., George Washington Univ, Washington D.C., 1985.
- WEST, M.; HARRISON, J. Bayesian Forecasting and Dynamic Models, 2. ed. New York: Springer. 1997.

Received on January 26, 2002.

Accepted on June 18, 2002.