

AJUSTE DE MODELOS AUTORREGRESSIVOS, NA FORMA DE MODELOS LINEARES DINÂMICOS, VIA INFERÊNCIA BAYESIANA

Autorregressive models fitting with a dynamic linear models approach via Bayesian inference

Marcelo Costa Souza¹, Thelma Sáfiadi²

RESUMO

Os modelos autorregressivos têm sido utilizados para as mais diversas aplicações, a maioria pela análise clássica, na qual os parâmetros são quantidades fixas, não podendo assumir variações ao longo do tempo. Com este trabalho objetivou-se a compreensão de modelos autorregressivos de ordem 2, AR(2), representados na forma de modelos lineares dinâmicos, utilizando como processo de estimação a inferência Bayesiana. O método de Cadeias de Markov Monte Carlo (MCMC) foi utilizado para o cálculo das estimativas a partir da implementação dos algoritmos amostrador de Gibbs e “Forward Filtering, Backward Sampling – FFBS”. Com base nos modelos AR(2), apresentaram-se o cálculo e a obtenção das distribuições condicionais completas para todos os parâmetros do modelo. Para avaliar o comportamento e a qualidade do ajuste, utilizaram-se duas cadeias de valores, cada uma com 8000 iterações, para três diferentes tamanhos de séries geradas, com 200, 500 e 800 observações. Como parte da aplicação, ajustou-se a série Canadian Lynx (NICHOLLS e QUIN, 1982) para diferentes fatores de desconto (0,90, 0,95 e 0,99), sendo o erro quadrático médio resultante utilizado para a comparação com o ajuste da mesma série, via inferência clássica. Um melhor ajuste para o modelo com fator de desconto igual a 0,99 foi observado. Considerando-se as estimativas obtidas tanto no caso simulado quanto para dados reais, obtiveram-se as previsões um passo à frente para as séries atualizada e “amostrada para trás”, e para essa última, o ajuste e o erro quadrático médio comportaram-se bem melhor. Com base nos resultados obtidos, observou-se um bom ajuste dos modelos AR(2) na forma de modelos dinâmicos, via inferência Bayesiana, além de se obter uma melhor compreensão em relação à qualidade do ajuste em diferentes situações, simuladas e reais.

Termos para indexação: FFBS, inferência bayesiana, modelos lineares dinâmicos, séries temporais.

ABSTRACT

The autoregressive models have been widely used in applications, mostly through a classical viewpoint, in which the parameters are regarded as fixed quantities, not assuming changes in time. This work aimed at fitting of autoregressive models with order 2, AR(2), specified in the form of dynamic linear models using Bayesian inference. Monte Carlo Markov Chain (MCMC) was used to obtain the estimates, via Gibbs Sampler and Forward Filtering Backward Sampling (FFBS). To evaluate the fitting, two chains with 8000 iterations each, and three different series sizes, with 200, 500 and 800 observations were sampled. The Canadian lynx series (NICHOLLS and QUIN, 1982), was fitted with different discount factors (0.90, 0.95 and 0.99), and the resulting mean square error was used to compare to the fitting using classical inference. A better fit for the model with discount equal to 0.99 was observed. One-step ahead forecasts were done to check the estimates obtained for the updated and the backward sampled series. To the latter, the fitting was better and mean square error lower. In general, it was observed a good fit of the AR(2) dynamic models via Bayesian inference, and this gives a better understanding of the fitting in different situations, both simulated and real.

Index terms: Bayesian inference, dynamic linear models, FFBS, time series.

(Recebido para publicação em 25 de agosto de 2002 e aprovado em 23 de julho de 2004)

INTRODUÇÃO

Os modelos autorregressivos têm sido utilizados em estudos envolvendo previsão nas mais diversas áreas, como, por exemplo, Economia, Geologia, Ciências Agrárias, Biologia, entre outras. A literatura apresenta diversas aplicações, assim como métodos para sua estimação e formas para se modelar os dados.

Wakefield et al. (1994) apresentam uma análise Bayesiana completa aplicada a modelos populacionais lineares e não-lineares. Segundo os autores, o amostrador parece oferecer o método mais flexível e poderoso disponível atualmente para análises de problemas envolvendo modelos populacionais. West (1995) apresenta uma revisão de variações de Modelos Lineares Dinâmicos, mencionando aspectos práticos para tratar de dados omissos, observações adicionais, erros amostrais,

1. Mestre em Estatística e Experimentação Agropecuária – Universidade Federal de Lavras/UFLA – Caixa Postal 3037 – Lavras, MG, mcsouza@ufla.net.com.br
2. Dra. em Estatística - Profa. Adjunta do DEX/UFLA, safadi@ufla.br

“outliers” e decomposição da série para investigar possíveis componentes latentes. Há uma discussão sobre cadeias de Markov Monte Carlo (MCMC) para a simulação da posteriori em modelos de espaços de estados, utilizados na análise de registros de isótopos de oxigênio no fundo de oceanos. Como aplicação, Glickman e Stern (1998) utilizam modelos de espaço de estados para desenvolver um modelo preditivo para os scores da Liga Nacional de Futebol Americano (NFL). Com a utilização de tais modelos, os parâmetros de interesse, como a medida da força de um time, podiam variar com o tempo, sendo assumido que esses parâmetros comportavam-se como um processo autorregressivo.

Apresenta-se neste trabalho o ajuste de um conjunto de observações tomadas em tempos equidistantes, por meio de modelos autorregressivos na forma de modelos de espaço de estados, ou modelos dinâmicos. A característica principal desses modelos reside na possibilidade de os parâmetros estarem sujeitos a variações ao longo do tempo. Tem-se ainda que os parâmetros, como quantidades desconhecidas, possuem distribuição, e que essas podem ser obtidas a partir da inferência Bayesiana, com base no conhecimento da distribuição a posteriori. A distribuição marginal para os parâmetros do modelo é obtida com base em resultados de simulação, utilizando o algoritmo amostrador de Gibbs e o “Forward Filtering Backward Sample – FFBS”.

Este artigo está organizado da seguinte maneira: Material e Métodos, onde são apresentados os conceitos de Modelos Lineares Dinâmicos (MLD). A definição das equações de observação e evolução, juntamente com os procedimentos para a atualização do modelo no tempo, mostram uma relação direta com os conceitos da estatística Bayesiana. Em seguida, são apresentados os modelos autorregressivos na forma de MLD. Também é apresentada uma descrição detalhada do algoritmo “FFBS” para o caso geral. Nos Resultados e Discussão, são apresentados os resultados obtidos. Primeiramente, são apresentados os cálculos necessários para a obtenção das distribuições condicionais completas para todos os parâmetros. Como parte das aplicações, na Aplicação com base em dados Simulados, há um exemplo realizado, considerando-se dados simulados, para, em seguida, na Aplicação com Dados Reais, apresentar uma aplicação com base na série apresentada por Nicholls e Quinn (1982).

MATERIAL E MÉTODOS

Para a análise, foi considerado o modelo autorregressivo de ordem 2, AR(2), na forma de MLD, e as

aplicações foram feitas utilizando-se dados simulados e a série Canadian Lynx.

Modelos Lineares Dinâmicos

O Modelo Linear Dinâmico,MLD, descrito por West e Harrison (1997), é caracterizado pela quádrupla $\{\mathbf{F}, \mathbf{G}, \mathbf{V} \text{ e } \mathbf{W}\}_t$, e pelas equações das observações e de evolução:

Equação de observação:

$$\mathbf{Y}_t = \mathbf{F}_t \boldsymbol{\theta}_t + v_t \quad v_t \sim N[0, \mathbf{V}_t] \quad (1)$$

Equação de sistema:

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad \boldsymbol{\omega}_t \sim N[\mathbf{0}, \mathbf{W}_t]$$

Informação inicial:

$$(\boldsymbol{\theta}_0 | D_0) \sim N[\mathbf{m}_0, \mathbf{C}_0]$$

Com base nas equações apresentadas em (1), têm-se:

\mathbf{F}_t é uma matriz conhecida ((p+1) x 1);

\mathbf{G}_t é uma matriz conhecida ((p+1) x (p+1));

\mathbf{V}_t é uma variância conhecida;

\mathbf{W}_t é uma matriz de variâncias conhecidas ((p+1) x (p+1));

em que p é o número de parâmetros.

Para compreender o intercâmbio entre as duas equações apresentadas em (1), utilizou-se a notação D_t para representar toda a informação disponível no tempo t. Em sistemas fechados $D_t = \{D_{t-1}, Y_t\}$, ou seja, a cada tempo t, a única informação nova que “entra” no sistema é a observação Y_t . Observa-se também que há uma evolução da informação, uma vez que conhecendo D_t , D_{t-1} é desnecessário, haja vista que essa quantidade já está presente em D_t .

A equação das observações define a distribuição amostral para Y_t condicional ao conhecimento da quantidade $\boldsymbol{\theta}_t$; assim, para o conhecimento de Y_t , basta o conhecimento de $\boldsymbol{\theta}_t$. A equação do sistema representa a forma com que o vetor de estado ($\boldsymbol{\theta}_t$) evolui com o tempo. A propriedade de independência condicional define uma evolução de Markov em 1 passo, de modo que, dados $\boldsymbol{\theta}_{t-1}$ e os valores de \mathbf{G}_t e \mathbf{W}_t , a distribuição de $\boldsymbol{\theta}_t$ é independente de D_{t-1} .

O erro v_t pode ser visto como uma simples perturbação aleatória que afeta a observação Y_t , mas não tem influência na série. Em contrapartida, $\boldsymbol{\omega}_t$ influencia o desenvolvimento do sistema no futuro, e a sua magnitude em relação à v_t indica o tamanho da variação do nível médio da série. Os modelos de Séries Temporais são um caso particular dos MLD quando as matrizes \mathbf{F} e \mathbf{G} (definidas pelo pesquisador) são constantes.

Equações de atualização

As equações de atualização representam toda a dinamicidade dessa classe de modelos. A partir delas, é possível apresentar a evolução da informação $(\theta_t|D_t)$ para $(\theta_{t+1}|D_{t+1})$ de maneira simples e direta, ou seja, é possível atualizar toda a informação do sistema à medida que novas observações vão sendo obtidas.

TEOREMA 1: No MLD univariado apresentado em (1), as previsões 1 passo à frente e as distribuições a posteriori são dadas, para cada t , como se segue:

- a) Posteriori em $t-1$: $(\theta_{t-1}|D_{t-1}) \sim N[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}]$
 b) Priori em t : $(\theta_t|D_{t-1}) \sim N[\mathbf{a}_t, \mathbf{R}_t]$
 $\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}$ $\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t$
 c) Previsão 1 passo à frente: $(Y_t|D_{t-1}) \sim N[f_t, Q_t]$
 $f_t = \mathbf{F}'_t \mathbf{a}_t$ $Q_t = \mathbf{F}'_t \mathbf{R}_t \mathbf{F}_t + V_t$
 d) Posteriori em t : $(\theta_t|D_t) \sim N[\mathbf{m}_t, \mathbf{C}_t]$
 $\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t / Q_t$ $e_t = Y_t - f_t$
 $\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t e_t$ $\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t Q_t \mathbf{A}_t'$

Nota-se, assim, que a partir do conhecimento da distribuição a priori $\theta_0|D_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$, é possível obter as distribuições $(\theta_1|D_1)$, ..., $(\theta_n|D_n)$.

Variância de Evolução \mathbf{W}_t

A especificação da seqüência da matriz de evolução \mathbf{W}_t é geralmente realizada com o auxílio de fatores de desconto. A especificação da estrutura dessa matriz é crucial para o sucesso da modelagem e previsão, uma vez que seus valores controlam a magnitude da variação estocástica na evolução do modelo, determinando, assim, a estabilidade no tempo.

Pelas equações de atualização do Teorema 1, têm-se:

$V[\theta_{t-1}|D_{t-1}] = \mathbf{C}_{t-1}$, que é a variância a posteriori;

$V[\theta_t|D_{t-1}] = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t$, que é a variância a priori \mathbf{R}_t ;

$V[Y_t|D_{t-1}] = \mathbf{F}'_t \mathbf{R}_t \mathbf{F}_t + V_t$, que é a variância Q_t .

Assim, quanto maior o valor de \mathbf{W}_t , menor será a precisão (aumento de Q_t) para prever futuras observações e, conseqüentemente, maior será a perda de informações. Na prática, a evolução do sistema pode ser vista a partir do conceito de fatores de desconto. Por definição, um fator de desconto, denotado por δ , assume valores no intervalo dado por $0 < \delta \leq 1$, e, geralmente, assume valores próximos de 1.

Para compreender o funcionamento do desconto, em um dado tempo t , define-se $\mathbf{P}_t = \mathbf{G}_t \mathbf{C}_t$

\mathbf{G}_t' , sendo $(\theta_{t-1}|D_{t-1}) \sim N[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}]$. A precisão associada a θ_{t-1} é \mathbf{C}_{t-1}^{-1} e, para $\mathbf{G}_t \theta_{t-1}$, é \mathbf{P}_t^{-1} . Assim, \mathbf{P}_t^{-1} representa a precisão associada a θ_t se não houver mudanças estocásticas no tempo t , o que caracteriza um modelo "globalmente" durável. Como o modelo só é localmente apropriado, a precisão atual \mathbf{R}_t^{-1} é reduzida em relação a \mathbf{P}_t^{-1} .

O conceito de desconto define essa decaída diretamente, via $\delta \mathbf{P}_t^{-1}$, ou simplesmente uma proporção δ da precisão. Assim,

$$V[\theta_t|D_{t-1}] = \mathbf{R}_t = (1/\delta) \mathbf{P}_t,$$

e, como $\mathbf{R}_t = \mathbf{P}_t + \mathbf{W}_t$, tem-se que:

$$(1/\delta) \mathbf{P}_t = \mathbf{P}_t + \mathbf{W}_t,$$

ou, então, $\mathbf{W}_t = [(1-\delta)/\delta] \mathbf{P}_t$.

Dessa forma, dados δ e \mathbf{C}_0 , a série $\{\mathbf{W}_t\}$ fica totalmente identificada e quanto mais próximo de 1 for o fator de desconto, mais durável é o modelo (menor será a perda de informação).

Modelos Autorregressivos na forma de MLD

Considere que a série Y_t possa ser dada como a soma de um nível (μ_t), um componente autorregressivo e um ruído, ou seja, $Y_t = \mu_t + X_t + v_t$, em que

$$a) \mu_t = \mu_{t-1} + \omega_t; \quad \omega_t \sim N[0, W^*],$$

$$b) X_t = \sum_{j=1}^p \phi_j X_{t-j} + \varepsilon_t; \quad \varepsilon_t \sim N[0, U],$$

$$c) v_t \sim N[0, V].$$

O nível médio da série, μ_t está na forma de um passeio aleatório, ou seja, é igual ao nível médio no instante anterior (μ_{t-1}), acrescido de um ruído aleatório ω_t , e X_t representa a parte autorregressiva de ordem p .

Expressando as equações de (a) a (c) na forma de MLD, com base nas equações de observação e evolução, têm-se:

$$Y_t = \mathbf{F}' \theta_t + v_t \tag{2}$$

$$\theta_t = \mathbf{G} \theta_{t-1} + \omega_t$$

em que:

$$\theta_t = (\mu_t, X_t, X_{t-1}, \dots, X_{t-p+1})'$$
 é vetor de estados,

$$\mathbf{F} = (1, 1, 0, \dots, 0)'$$

$$\omega_t = (\omega_t, \varepsilon_t, 0, \dots, 0)'$$

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \phi_1 & \phi_2 & \phi_3 & \dots & \phi_p \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix},$$

$$\mathbf{W}^* = \begin{bmatrix} \mathbf{W} & 0 & 0 & \dots & 0 \\ 0 & \mathbf{U} & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \text{ e } \omega^* \sim N(0, \mathbf{W}^*)$$

Com base no modelo (2), os parâmetros a serem estimados são ϕ , \mathbf{V} , \mathbf{W} , \mathbf{U} , além dos vetores de estados. Define-se $\Theta_n = \{\theta_0, \theta_1, \dots, \theta_n\}$ e $\alpha_n = \{\phi, \mathbf{V}, \mathbf{W}, \mathbf{U}\}$. Pelo amostrador de Gibbs, infere-se que a distribuição a posteriori conjunta

$$p(\Theta_n, \alpha_n | D_n)$$

pode ser simulada iteragindo entre duas posteriores condicionais completas

$$p(\Theta_n | \alpha_n, D_n) \leftrightarrow p(\alpha_n | \Theta_n, D_n). \quad (3)$$

Esse esquema envolve a amostragem de Θ_n (1º bloco) e, separadamente, de α_n , a partir das posteriores condicionais multivariadas completas (2º bloco).

Como a análise é padrão condicional a α_n , a simulação do 1º componente de (3) será acessível como uma distribuição normal multivariada a partir da implementação do algoritmo "Forward Filtering, Backward Sampling", sendo o segundo componente muito dependente da forma do modelo.

RESULTADOS E DISCUSSÃO

Com base na teoria apresentada em Modelos Lineares Dinâmicos e Modelos Autorregressivos na Forma de MLD, o comportamento do modelo autorregressivo de 2ª ordem, AR(2), foi avaliado com relação às estimativas dos parâmetros, erros de previsão, convergência e previsões um passo à frente. Considerando o modelo

$$Y_t = \mu_t + X_t + v_t,$$

o procedimento para a simulação segue aquele descrito em Modelos Autorregressivos na Forma de MLD, sendo os parâmetros do modelo que devem ser estimados: $\alpha = \{\phi_1, \phi_2, \mathbf{V}, \mathbf{W}, \mathbf{U}\}$ e Θ_n , em que $\Theta_n = (\theta_0, \theta_1, \dots, \theta_n)$.

O método utilizado para o procedimento de inferência e estimação dos parâmetros foi o amostrador de Gibbs, o qual, a partir de resultados de simulação, forneceu como resultados: (1) as estimativas dos parâmetros, (2) verificação de convergência e (3) os gráficos dos valores gerados.

Outro resultado explorado foi a utilização das estimativas para se obter previsões um passo à frente. Essas previsões foram obtidas de duas maneiras: (1) a partir das equações de atualização e (2) a partir dos vetores de estado provenientes da "amostragem para trás", ao aplicar o algoritmo FFBS.

Distribuições condicionais completas

A amostragem de $\Theta_n | \alpha_n, D_n$, apresentada em Modelos Autorregressivos na forma de MLD, foi realizada mais facilmente considerando o algoritmo FFBS descrito a seguir:

O objetivo foi amostrar Θ_n pela simulação sequencial dos vetores de estado individuais $\theta_n, \theta_{n-1}, \dots, \theta_0$ da seguinte maneira:

1 – Amostra-se θ_n de $\theta_n | D_n \sim N[\mathbf{m}_t, \mathbf{C}_t]$, e então,

2 – para cada $t = n-1, n-2, \dots, 1, 0$ amostra-se θ_t de $p(\theta_t | \theta_{t+1}, D_t)$, em que o valor condicional de θ_{t+1} é o vetor que acabou de ser amostrado.

No que segue, será considerado o caso particular AR(2), e a extensão para o caso de um AR(p) é imediata. Com base no vetor θ_n amostrado, o interesse concentra-se na obtenção do vetor $\theta_t | \theta_{t+1}, D_t$, para $t = n-1, \dots, 1, 0$.

A partir do fato de que

$$\theta_{t+1} = (\mu_{t+1}, X_{t+1}, X_t), \text{ e de que}$$

$$\theta_t = (\mu_t, X_t, X_{t-1}),$$

o segundo elemento de θ_t está definido, uma vez que o vetor anterior, θ_{t+1} , é conhecido. Assim, é necessário amostrar o primeiro e último elementos de θ_t , para depois completar com o valor de X_t , já conhecido a partir do vetor amostrado anteriormente.

Assim, o primeiro passo é computar os momentos desses dois valores a partir de $p(\mu_t, X_{t-1} | X_t, D_t)$, o que é exatamente a distribuição condicional bivariada para os dois elementos de θ_t a partir da distribuição conjunta completa $\theta_t | D_t \sim N[\mathbf{m}_t, \mathbf{C}_t]$.

Para amostrar esses elementos de θ_t , definem-se as seguintes matrizes:

$$\Sigma_{11} = \begin{bmatrix} \text{var}(\mu_t) & \text{cov}(\mu_t, X_{t-1}) \\ \text{cov}(X_{t-1}, \mu_t) & \text{var}(X_{t-1}) \end{bmatrix}$$

$$\Sigma_{12} = \begin{bmatrix} \text{cov}(\mu_t, X_t) \\ \text{cov}(X_t, X_{t-1}) \end{bmatrix} = \Sigma'_{21},$$

$$= [\text{cov}(\mu_t, X_t) \quad \text{cov}(X_t, X_{t-1})]$$

e $\Sigma_{22} = \text{var}(X_t)$, obtidos de C_t .

Definindo θ_t^- como o vetor $(\mu_t, X_{t-1})'$, o qual deve ser amostrado, têm-se:

$$\theta_t^- | X_t, D_t, \sim N_2(\mu_2, \Sigma_2), \text{ onde,}$$

$$\mu_2 = \begin{bmatrix} \bar{\mu}_t \\ \bar{X}_{t-1} \end{bmatrix} + \Sigma_{12} \Sigma_{22}^{-1} (X_t - \bar{X}_t) \quad \text{e} \quad (4)$$

$$\Sigma_2 = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

Dessa forma, tem-se que

$$p(\theta_t^- | D_t, X_t) \propto |\Sigma_2|^{-1/2} \exp\left\{-\frac{1}{2}[(\theta_t^- - \mu_2)' \Sigma_2^{-1} (\theta_t^- - \mu_2)]\right\}. \quad (5)$$

Fazendo $e_{t+1} = X_{t+1} - \phi_1 X_t$, têm-se $\mu_{t+1} \sim N(\mu_t, W)$ e $e_{t+1} \sim N(\phi_2 X_{t-1}, U)$. (6)

A primeira parte de (6) vem do fato de que $\mu_{t+1} = \mu_t + \omega_t$, com $\omega_t \sim N(0, W)$. Para o segundo termo, $X_{t+1} = \phi_1 X_t + \phi_2 X_{t-1} + \varepsilon_{t+1}$, com $\varepsilon_{t+1} \sim N(0, U)$. Como $e_{t+1} = X_{t+1} - \phi_1 X_t$, tem-se que $e_{t+1} = \phi_2 X_{t-1} + \varepsilon_{t+1}$, o que fornece $e_{t+1} \sim N(\phi_2 X_{t-1}, U)$. Assim, de (5) e (6), obtém-se:

$$p(\theta_t^- | \theta_{t+1}, D_n) \propto \exp\left\{-\frac{1}{2}[(\theta_t^- - \mu_2)' \Sigma_2^{-1} (\theta_t^- - \mu_2)]\right\} x$$

$$x \exp\left\{-\frac{1}{2} \left[\frac{(\mu_{t+1} - \mu_t)^2}{W} + \frac{(X_{t+1} - \phi_1 X_t - \phi_2 X_{t-1})^2}{U} \right]\right\}. \quad (7)$$

O segundo termo da expressão (7) pode ser reescrito como:

$$(\mu_t - \mu_{t+1}) W^{-1} (\mu_t - \mu_{t+1}) +$$

$$+ (X_{t-1} - (X_{t+1} - \phi_1 X_t) \phi_2^{-1}) \phi_2^2 U^{-1} (X_{t-1} - (X_{t+1} - \phi_1 X_t) \phi_2^{-1}).$$

Escrevendo na forma matricial, tem-se:

$$\begin{bmatrix} \mu_t \\ X_{t-1} \end{bmatrix} - \begin{bmatrix} \mu_{t+1} \\ (X_{t+1} - \phi_1 X_t) \phi_2^{-1} \end{bmatrix} \begin{bmatrix} W & 0 \\ 0 & U \phi_2^{-2} \end{bmatrix}^{-1} \begin{bmatrix} \mu_t \\ X_{t-1} \end{bmatrix} - \begin{bmatrix} \mu_{t+1} \\ (X_{t+1} - \phi_1 X_t) \phi_2^{-1} \end{bmatrix}$$

Definindo

$$\mu_{\theta_t^-} = \begin{bmatrix} \mu_{t+1} \\ (X_{t+1} - \phi_1 X_t) \phi_2^{-1} \end{bmatrix} \quad \text{e} \quad \Sigma_{\theta_t^-} = \begin{bmatrix} W & 0 \\ 0 & U \phi_2^{-2} \end{bmatrix},$$

tem-se:

$$p(\theta_t^- | \theta_{t+1}, D_n)$$

$$\propto \exp\left\{-\frac{1}{2}[(\theta_t^- - \mu_2)' \Sigma_2^{-1} (\theta_t^- - \mu_2) + (\theta_t^- - \mu_{\theta_t^-})' \Sigma_{\theta_t^-}^{-1} (\theta_t^- - \mu_{\theta_t^-})]\right\}. \quad (8)$$

Considerando a seguinte componente de (8):

$$(\theta_t^- - \mu_2)' \Sigma_2^{-1} (\theta_t^- - \mu_2) + (\theta_t^- - \mu_{\theta_t^-})' \Sigma_{\theta_t^-}^{-1} (\theta_t^- - \mu_{\theta_t^-}),$$

e efetuando as multiplicações dos termos, ela pode ser reescrita como:

$$p(\theta_t^- | \theta_{t+1}, D_n)$$

$$\propto \exp\left\{-\frac{1}{2} \left[\theta_t^- - (\Sigma_2^{-1} + \Sigma_{\theta_t^-}^{-1})^{-1} (\Sigma_2^{-1} \mu_2 + \Sigma_{\theta_t^-}^{-1} \mu_{\theta_t^-}) \right]' (\Sigma_2^{-1} + \Sigma_{\theta_t^-}^{-1}) x \right\}$$

$$x \left\{ \theta_t^- - (\Sigma_2^{-1} + \Sigma_{\theta_t^-}^{-1})^{-1} (\Sigma_2^{-1} \mu_2 + \Sigma_{\theta_t^-}^{-1} \mu_{\theta_t^-}) \right\},$$

a qual tem a forma de uma distribuição normal para θ_t^- . Portanto,

$$\theta_t^- | \theta_{t+1}, D_n \sim N(\mu_f, \Sigma_f), \quad (9)$$

$$\text{onde } \mu_f = (\Sigma_2^{-1} + \Sigma_{\theta_t^-}^{-1})^{-1} (\Sigma_2^{-1} \mu_2 + \Sigma_{\theta_t^-}^{-1} \mu_{\theta_t^-}),$$

$$\text{e } \Sigma_f = (\Sigma_2^{-1} + \Sigma_{\theta_t^-}^{-1})^{-1}.$$

Dessa forma, amostra-se um valor para θ_t^- a partir de (9), e então, completa-se com o valor de X_t , já amostrado em θ_{t+1} . Esse procedimento é repetido para $t = n-1, \dots, 0$, até que o vetor $\Theta_n = \{\theta_n, \theta_{n-1}, \dots, \theta_1, \theta_0\}$ seja amostrado.

Com isso, o primeiro bloco está amostrado e a partir desse, pode-se utilizar as informações obtidas (conjunto de todos os vetores θ) para proceder à amostragem do segundo bloco, o qual é composto pelas seguintes distribuições condicionais completas:

a) $\phi | V, W, U, \Theta_n, D_n,$

b) $V | \phi, W, U, \Theta_n, D_n,$

c) $W | \phi, U, V, \Theta_n, D_n,$

d) $U | \phi, V, W, \Theta_n, D_n.$

a) Amostragem de $\phi | V, U, W, \Theta_n, D_n:$

$$P(\phi | V, U, W, \Theta_n, D_n) \propto p(\phi |$$

$$D_0) (p(Z_0 | \alpha, D_0) \prod_{t=1}^n \exp(-(X_t - Z'_{t-1} \phi)^2 / 2U)$$

$$\text{Definido } \mathbf{A} = \begin{bmatrix} \Sigma \mathbf{X}_{t-1}^2 & \Sigma \mathbf{X}_{t-1} \mathbf{X}_{t-2} \\ \Sigma \mathbf{X}_{t-2} \mathbf{X}_{t-1} & \Sigma \mathbf{X}_{t-2}^2 \end{bmatrix}, \text{ e}$$

$$\mathbf{B} = \begin{bmatrix} \Sigma \mathbf{X}_t \mathbf{X}_{t-1} \\ \Sigma \mathbf{X}_t \mathbf{X}_{t-2} \end{bmatrix} \text{ tem-se:}$$

$$(\phi|V, U, W, \Theta_n, D_n) \sim N(\mathbf{A}^{-1} \mathbf{B}, U \mathbf{A}^{-1}).$$

b) Amostragem de $V|\phi, W, U, \Theta_n, D_n$:

$$V \mid \phi, W, U, \Theta_n, D_n \sim \text{GI} \left(\frac{n+2\alpha}{2}, \frac{1}{2} \left[\sum_{t=1}^n (Y_t - \mathbf{F}' \theta_t)^2 + 2\beta \right] \right).$$

O procedimento para obter as outras duas distribuições condicionais completas $W|\phi, U, V, \Theta_n, D_n$ e $U|\phi, V, W, \Theta_n, D_n$ é similar ao desenvolvido com relação à distribuição para $V|\phi, W, U, \Theta_n, D_n$, conduzindo até as seguintes distribuições:

$$W \mid \phi, V, U, \Theta_n, D_n \sim \text{GI} \left(\frac{n+2\alpha}{2}, \frac{1}{2} \left[\sum_{t=1}^n (\mu_t - \mu_{t-1})^2 + 2\beta \right] \right).$$

$$U \mid \phi, V, W, \Theta_n, D_n \sim \text{GI} \left(\frac{n+2\alpha}{2}, \frac{1}{2} \left[\sum_{t=3}^n (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2})^2 + 2\beta \right] \right).$$

Nas duas aplicações a serem apresentadas, trabalhou-se com 8000 iterações para cada uma das duas cadeias. Para reduzir o efeito desses valores iniciais, foram eliminadas as primeiras 20% observações, o que reduziu para 6400 o número de observações para cada uma.

Como o amostrador é um processo Markoviano, as amostras não são tomadas independentemente, e

sim, com uma correlação de 1ª ordem. Para suavizar esse problema, retirou-se 1 observação a cada 25 de um total de 6400 observações de cada cadeia. Assim, essa nova série possui 256 observações, totalizando 512, ao analisar as duas cadeias.

As estimativas foram obtidas com base nessas 512 observações e utilizadas para compor as previsões um passo à frente das duas maneiras descritas anteriormente.

Os parâmetros utilizados na distribuição a priori para as variâncias foram $\alpha = \beta = 2$. A distribuição atribuída para $\theta_0|D_0$ possui média e variância dadas por:

$$\mathbf{m}_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \text{ e } \mathbf{C}_0 = \begin{bmatrix} 0,3 & 0 & 0 \\ 0 & 0,8 & 0 \\ 0 & 0 & 0,7 \end{bmatrix}.$$

O software utilizado para esses procedimentos de simulação e ajuste da série real foi o MATLAB 6.0. Para a análise da série com base na inferência clássica, utilizou-se o software MINITAB v.12, o qual estima os parâmetros a partir do método da máxima verossimilhança.

Aplicação com base em dados simulados

Na aplicação com dados simulados, foram utilizadas duas cadeias em paralelo para avaliar a convergência e os diferentes tamanhos de série, com 200, 500 e 800 observações. As séries foram geradas segundo os valores $\phi_1 = 0,6$, $\phi_2 = -0,4$, $V = 0,5$, $W = 0,9$, $U = 0,9$. Na Tabela 1 apresenta-se a comparação das estimativas obtidas para cada uma das três situações avaliadas.

TABELA 1 – Comparação das estimativas dos parâmetros e do erro quadrático médio obtido para cada tamanho de série.

		Tamanho da Série					
Valor Real		200 Observações		500 Observações		800 Observações	
		Média	DP	Média	DP	Média	DP
ϕ_1	0,6	0,6028	0,2825	0,5295	0,2060	0,5011	0,1325
ϕ_2	-0,4	-0,0871	0,2656	-0,2636	0,2415	-0,4995	0,1670
V	0,5	0,4111	0,1463	0,3109	0,0920	0,2756	0,0680
W	0,9	6,5660	0,8267	4,3794	0,3583	3,5731	0,2549
U	0,9	0,3547	0,1294	0,2767	0,0844	0,2529	0,0707
EQM – atualizada		6,8434		4,8541		4,2399	
EQM – amostrada		0,3721		0,3090		0,2764	

Na Figura 1 apresentam-se os gráficos para a série simulada e para a série prevista a partir das equações de atualização e da amostragem para trás, considerando a situação com 800 observações.

Aplicação com dados reais

Na aplicação com dados reais, utilizou-se a série Canadian lynx (NICHOLLS e QUIN, 1982). A

série possui 100 observações anuais no período de 1821 a 1920; o seu gráfico pode ser visto na Figura 2.

Os parâmetros a serem estimados no caso da aplicação com dados reais foram ϕ_1 , ϕ_2 , V e U . A matriz W foi trabalhada a partir da utilização de fatores de descontos. Os valores adotados para δ foram 0,90, 0,95 e 0,99. A comparação dos resultados pode ser vista na Tabela 2.

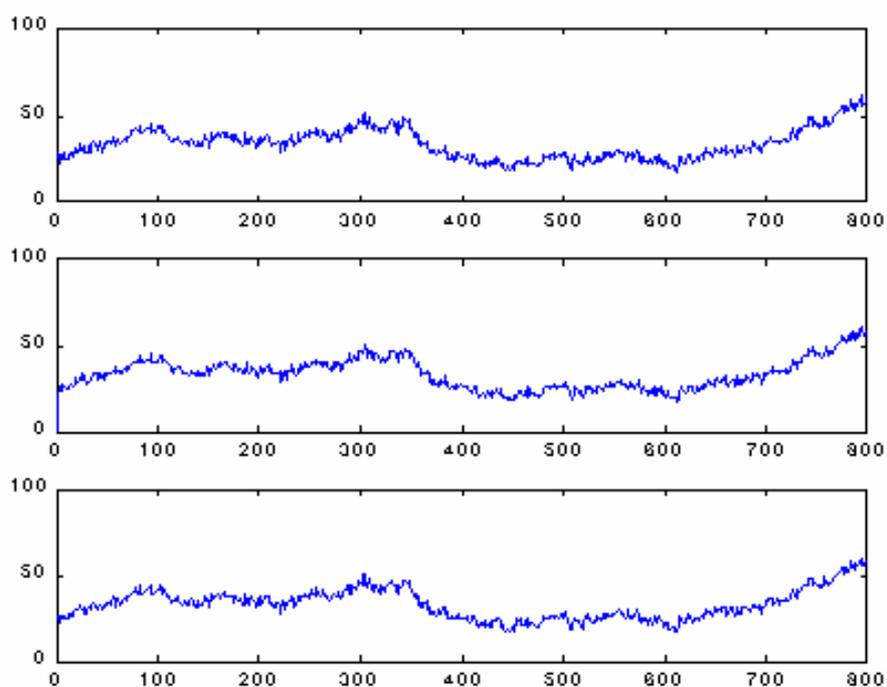


FIGURA 1 – Gráficos da série original, atualizada e “amostrada para trás”, respectivamente, a partir das estimativas obtidas.

TABELA 2 – Estimativas obtidas para ϕ_1 e ϕ_2 e erro quadrático médio, segundo a estimação utilizada.

	$\delta = 0,90$	$\delta = 0,95$	$\delta = 0,99$	Inf. Clássica
ϕ_1	1,1718	1,2453	1,2592	1,3841
ϕ_2	-0,3671	-0,5070	-0,5893	-0,7546
EQM	0,0749	0,0640	0,0400	0,0570

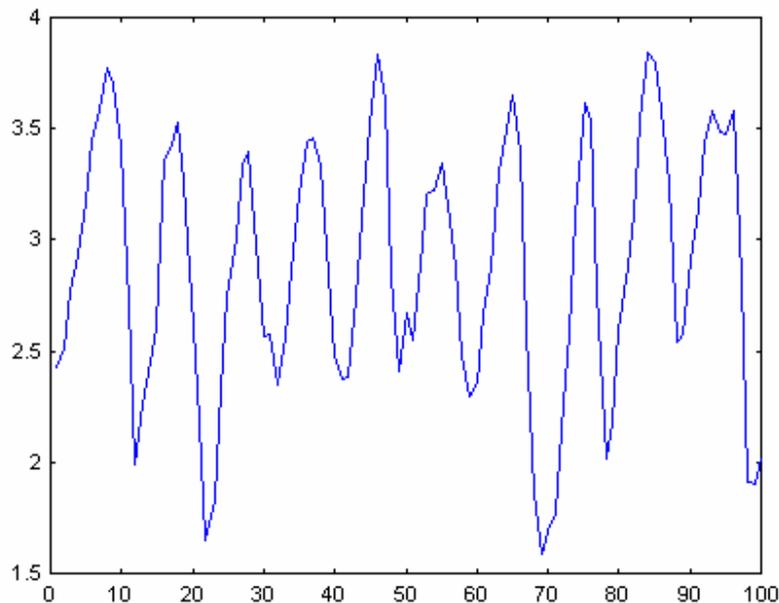


FIGURA 2 – Gráfico da série Canadian lynx.

Considerações Finais

A convergência, medida pelo teste de Gelman e Rubin (1992), não apresentou problemas para nenhuma situação avaliada. Nos resultados de simulação e na aplicação com dados reais, utilizaram-se 8000 iterações em cada cadeia, e os resultados obtidos a partir de 5000 iterações apresentaram estimativas de R um pouco mais distantes do valor de referência igual a 1, quando comparadas com as obtidas para 8000.

Pelos altos valores apresentados pela variância W, infere-se uma dificuldade de utilização na prática devido à dificuldade de especificação a priori a respeito de seu comportamento.

O procedimento para simulação na aplicação com dados reais só se comportou de maneira satisfatória com a utilização de fatores de desconto para se trabalhar com a matriz W. Quando se propôs não utilizá-los, o procedimento de simulação não evoluiu.

O ajuste da série real comportou-se muito bem com relação às previsões um passo à frente; entretanto, previsões para “lags” maiores não foram avaliadas.

CONCLUSÕES

Com base na teoria apresentada e nos resultados obtidos a partir de procedimentos de simulação e aplicação em dados reais, pode-se concluir que:

a) É possível obter distribuições condicionais completas para os parâmetros avaliados e aplicar o algoritmo FFBS para o caso de um modelo AR(2).

b) As séries que possuem um maior número de observações foram mais bem ajustadas. As estimativas dos parâmetros aproximaram-se dos valores reais, com exceção do valor de W, que assumiu valores bem maiores do que o real.

c) Os dados utilizados na aplicação ajustaram-se muito bem com os Modelos Lineares Dinâmicos. O resultado da comparação entre a estimação a partir de fatores de desconto iguais a 0,90, 0,95 e 0,99 com a estimação a partir da inferência clássica foi favorável à estimação utilizando $\delta = 0,99$, a qual apresentou o menor erro quadrático médio.

d) As séries obtidas com base na “amostragem para trás” apresentaram um melhor ajuste da série original, quando comparadas com o ajuste das séries obtidas pelas equações de atualização. Esse fato ocorreu tanto com a série gerada quanto com a série real.

e) A modelagem de Séries Temporais via Modelos Lineares Dinâmicos por meio de Inferência Bayesiana mostrou-se aplicável, de implementação direta e ajustou-se bem tanto à série simulada quanto à série real.

REFERÊNCIAS BIBLIOGRÁFICAS

- GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequence. **Statistical Science**, Hayward, v. 7, n. 4, p. 457-511, 1992.
- GLICKMAN, M. E.; STERN, H. S. A state-space model for national football league scores. **Journal of the American Statistical Association**, Alexandria, v. 93, n. 441, p. 25-35, Mar. 1998.
- NICHOLLS, D. F.; QUINN, B. G. **Random coefficients autorregressive models**: an introduction: lecture notes in statistics. New York: Springer, 1982. v. 11.
- WAKEFIELD, J. C.; SMITH, A. F. M.; RACINE-POON, A. E.; GELFAND, A. E. Bayesian analysis of linear and non-linear population models by using the Gibbs Sampler. **Journal of the Royal Statistical Society serie C**, London, v. 43, n. 1, p. 201-221, 1994.
- WEST, M. Bayesian time series. In: WORKSHOP ON MAXIMUM ENTROPY AND BAYESIAN METHODS, 15., 1995, Santa Fé. **Proceedings...** Novo México: ISDS, 1995. p. 1-10.
- WEST, M.; HARRISON, J. **Bayesian forecasting and dynamic models**. New York: Springer, 1997. 680 p.