



UNIVERSIDADE FEDERAL DE LAVRAS

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

2002

53204

37685MFN

MARCELO COSTA SOUZA

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

Dissertação apresentada à
Universidade Federal de Lavras
como parte das exigências do
Curso de Mestrado em
Estatística e Experimentação
Agropecuária, área de
concentração em Teoria e
Métodos Estatísticos, para
obtenção do título de “Mestre”

Orientadora

Prof^ª Thelma Sáfadi

LAVRAS
MINAS GERAIS – BRASIL
2002

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Souza, Marcelo Costa

Ajuste de modelos autorregressivos na forma de modelos lineares dinâmicos
via inferência Bayesiana / Marcelo Costa Souza. -- Lavras : UFLA, 2002.

112 p. : il.

Orientadora: Thelma Sáfyadi.

Dissertação (Mestrado) – UFLA.

Bibliografia.

1. Modelo linear dinâmico. 2. Série temporal. 3. Inferência Bayesiana. 4.
Amostrador de Gibbs. I. Universidade Federal de Lavras. II. Título.

CDD-519.542


MARCELO COSTA SOUZA

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

Dissertação apresentada à
Universidade Federal de Lavras,
como parte das exigências do
Curso de Mestrado em
Estatística e Experimentação
Agropecuária, área de
concentração em Teoria e
Métodos Estatísticos, para
obtenção do título de “Mestre”

APROVADA em 21 de fevereiro de 2002

Prof. Heyder Diniz Silva	UFU
Prof. Daniel Furtado Ferreira	UFLA
Prof. Júlio Silvio de Sousa Bueno Filho	UFLA


Profª Thelma Sáfadi
UFLA
(Orientadora)

LAVRAS
MINAS GERAIS – BRASIL
2002

**Aos meus pais
Gilnei e Solange**

DEDICO

AGRADECIMENTO

A Deus pela minha existência.

Aos meus pais Gilnei e Solange, pela força, incentivo e apoio nas minhas escolhas.

À professora e amiga Thelma pela orientação, paciência e exemplo de profissionalismo.

Ao Professor Joel pela amizade e coordenação do curso de Mestrado em Estatística.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudo concedida.

Aos professores Augusto, Daniel, Delly, Eduardo, Júlio, Lucas, Luiz Henrique, Mário pela orientação nas disciplinas do curso.

Aos meus colegas de turma, pela convivência e experiências vividas, em especial a minha prima Livia.

Aos funcionários do DEX/UFLA, pela convivência e disponibilidade.

Aos meus irmãos Júnior, Ana e Flávia e à minha namorada Fernanda, pelo apoio no desenvolvimento desta dissertação.

ÍNDICE

RESUMO	i
ABSTRACT	ii
1 – INTRODUÇÃO	1
2 – REFERENCIAL TEÓRICO	4
2.1 – Estatística Bayesiana.....	4
2.2 – Simulação Estocástica via Cadeias de Markov.....	7
2.3 – Exemplos de processos de estimação Bayesiana.....	12
2.4 – Modelos Lineares Dinâmicos.....	16
2.5 – Modelos Lineares Dinâmicos em Séries Temporais.....	27
2.6 – Simulação em Modelos Lineares Dinâmicos.....	33
2.7 – Modelos Autorregressivos na Forma de Modelos Lineares Dinâmicos.....	38
3 – METODOLOGIA	43
4 – RESULTADOS E DISCUSSÃO	47
4.1 – Distribuições Condicionais Completas.....	47
4.2 – Aplicação com dados simulados.....	59
4.3 – Aplicação com dados reais.....	78
5 – CONCLUSÕES	93
6 – REFERÊNCIAS BIBLIOGRÁFICAS	94
ANEXO A – DISTRIBUIÇÕES DE PROBABILIDADE	98
ANEXO B – PROGRAMA PARA SIMULAÇÃO	100

RESUMO

SOUZA, Marcelo Costa. **Ajuste de modelos autorregressivos na forma de modelos lineares dinâmicos via inferência Bayesiana.** 2000. 112p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária)* - Universidade Federal de Lavras, Lavras.

Os modelos autorregressivos têm sido utilizados para as mais diversas aplicações, a maioria através da análise clássica, na qual os parâmetros são vistos como quantidades fixas, não podendo assumir variações ao longo do tempo. A modelagem dinâmica possibilita uma evolução desses parâmetros à medida que os dados vão sendo observados. Este trabalho tem como objetivo 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 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), apresentou-se o cálculo e a obtenção das distribuições condicionais completas para todos os parâmetros do modelo. Para avaliar o comportamento e 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 and Quin 1982) para diferentes fatores de desconto (0,90, 0,95 e 0,99), sendo que o erro quadrático médio resultante foi utilizado para a comparação com o ajuste da mesma série via inferência clássica. Um melhor ajuste para o modelo com desconto igual a 0,99 foi observado. A partir das 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”, sendo que, para esta última, o ajuste e o erro quadrático médio se comportaram 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.

* Orientadora: Thelma Sáfydi - UFLA

ABSTRACT

SOUZA, Marcelo Costa. **Autorregressive models fitting with a dynamic linear models approach via Bayesian inference**. 2000. 112p. Dissertation (Master in Statistics and Agricultural Experimentation)* - Universidade Federal de Lavras, Lavras

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. The dynamic modeling allows the parameters to change as data are observed. 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, by means of the algorithms Gibbs sampler and Forward Filtering Backward Sampling (FFBS). For AR(2) models, the calculation of the complete conditionals distributions for all the parameters was developed. To evaluate the fitting, two chains with 8000 iterations each, and three different series sizes, with 200, 500 and 800 observations were used. 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. With the estimates obtained in simulation and application, the one-step ahead forecasts were done for the updated and the backward sampled series. In the latter, the fitting was better and mean square error lower. In general, was observed a good fit of the AR(2) models in dynamic way via Bayesian inference, and a better understanding of the fitting in different situations, both simulated and real.

* Adviser: Thelma Sáfaci - UFLA

1 – INTRODUÇÃO

Os modelos autorregressivos têm sido utilizados em estudos envolvendo previsão nas mais diversas áreas, como, por exemplo, Econometria, Geologia, Ciências Agrárias e Bioestatística. A literatura apresenta diversas aplicações, assim como métodos para sua estimação e formas para modelar os dados. Buscasse, 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 destes modelos reside na possibilidade dos parâmetros do modelo estarem sujeitos a variações ao longo do tempo. Tem-se, ainda, que os parâmetros, como quantidades desconhecidas, possuem uma distribuição, e que estas podem ser obtidas a partir da inferência Bayesiana por meio do 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 trabalho está organizado da seguinte maneira:

Na seção 2.1 são apresentados alguns aspectos da estatística Bayesiana, como a sua crescente utilização no tratamento de problemas das mais diversas áreas, assim como a definição dos conceitos de distribuições a priori e a posteriori. A inferência Bayesiana é vista como uma poderosa ferramenta na estimação de parâmetros desconhecidos.

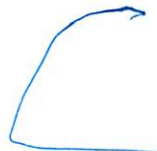
O uso da inferência Bayesiana frequentemente é acompanhado de métodos de simulação, os quais são descritos na seção 2.2. É dada uma atenção especial ao amostrador de Gibbs como caminho para se obterem as distribuições marginais a posteriori. Alguns métodos para verificação da convergência das distribuições dos parâmetros também são apresentadas.

Na seção 2.3 é apresentado um histórico da utilização de Modelos Lineares Dinâmicos (MLD) em várias áreas, algumas referências relativas ao uso do amostrador de Gibbs, além de uma evolução dos conceitos e aplicações ao longo do tempo.

Os conceitos de Modelos Lineares Dinâmicos são apresentados na seção 2.4. A definição das equações de observação e evolução, juntamente com os procedimentos para a atualização do modelo no tempo, mostra uma relação direta com os conceitos da estatística Bayesiana, em que os parâmetros do modelo são estimados a partir do intercâmbio de informações entre as distribuições a priori e a posteriori. Algumas características também são tratadas, como métodos de filtragem, fatores de desconto, e de como a estrutura das componentes utilizadas pelas equações pode auxiliar no ajuste do modelo a ser estudado.

Com o conhecimento das características dos Modelos Lineares Dinâmicos e de como elas podem ser relacionadas com a inferência Bayesiana, na seção 2.5 faz-se uma ligação com os conceitos de Séries Temporais, de forma a possibilitar o ajuste de Séries por meio das equações que caracterizam tais modelos. São apresentadas as características de modelos de Séries Temporais e de como descrevê-los na forma de Modelos Lineares Dinâmicos, sendo utilizados, mais especificamente, os modelos autorregressivos (AR).

Na seção 2.6 é apresentado um algoritmo, conhecido como “Forward Filtering, Backward Sampling – FFBS”, para auxílio na estimação dos parâmetros do MLD. Na seqüência, define-se um modelo AR na forma de Modelos Lineares Dinâmicos para o caso geral com “p” parâmetros, ou seja, um AR(P). Podem-se ver, também, as distribuições condicionais completas para os parâmetros do modelo e um esquema iterativo para a implementação do amostrador de Gibbs.



Na seção 3, acerca da metodologia utilizada, é apresentada a formulação do caso de um AR(2) com base no modelo AR(p) descrito na seção anterior. Também pode ser vista a maneira pela qual os resultados serão obtidos pelos procedimentos de simulação e aplicação em dados reais.

Na seção 4.1, pode ser visto o desenvolvimento das distribuições condicionais completas a posteriori, em que é definido o caminho para proceder com a estimação dos parâmetros utilizando dados simulados e dados reais com base em um modelo autorregressivo de ordem 2. A seção 4.2 apresenta os resultados obtidos a partir de dados simulados para diferentes tamanhos de séries. Como parte das aplicações, é apresentado, na seção 4.3, o ajuste da série anual Canadian lynx, apresentada em Nicholls & Quinn (1982), para diferentes fatores de desconto. O ajuste é comparado com o ajuste realizado utilizando inferência clássica. A seção 5 apresenta algumas conclusões com base nos resultados obtidos

Os anexos apresentam as características de algumas distribuições utilizadas, como, por exemplo, a Normal, Gama e Gama Inversa, assim como a descrição do programa utilizado para a implementação do amostrador de Gibbs e estimação dos parâmetros.

Os objetivos deste trabalho foram: (1) Apresentar as distribuições condicionais completas para o modelo AR(2), (2) Apresentar e avaliar o desempenho da modelagem de Séries Temporais via Modelos Lineares Dinâmicos e por meio da Inferência Bayesiana; (3) Apresentar o ajuste do modelo AR(2) por meio de um algoritmo implementando o amostrador de Gibbs; e (4) Comparar os resultados obtidos a partir da análise Bayesiana com aqueles da inferência clássica.

2 – REFERENCIAL TEÓRICO

2.1 – Estatística Bayesiana

A crescente versatilidade da estatística Bayesiana, vista a partir dos anos 80, promoveu uma maior utilização e compreensão das características de sua teoria. Como conceito inicial, e principal característica, a estatística Bayesiana atribui uma distribuição de probabilidade para qualquer quantidade desconhecida, ou seja, se não se conhece o valor de um parâmetro, deve-se atribuir uma distribuição de probabilidade para ele. Desta forma, esta distribuição pode ser definida antes que a amostra seja observada, ou seja, a priori.

Segundo West & Harrison (1997), a previsão sob ponto de vista Bayesiano envolve a determinação da informação predita em termos das distribuições de probabilidade que representam e resumem o conhecimento incerto atual.

Dois distribuições são características da estatística Bayesiana: (1) distribuição a priori e (2) distribuição a posteriori, sendo que é com base nesta última que se realizam todas as inferências necessárias.

A análise Bayesiana é representada pela combinação da informação a priori do parâmetro θ , $p(\theta)$ e da informação contida na amostra, Y , que resulta na chamada distribuição a posteriori de θ dado Y , na qual todas as decisões e inferências são realizadas.

A distribuição a posteriori, denotada por $p(\theta|Y)$, é definida como sendo a distribuição condicional de θ dada a informação amostral Y , ou seja, é obtida uma distribuição contendo todo o conhecimento em relação ao parâmetro θ , pela combinação da informação a priori com a observação da amostra.

A informação contida na amostra é representada por meio da função de verossimilhança, a qual representa a densidade conjunta de um grupo de observações $Y = (y_1, y_2, \dots, y_n)$ em função de um parâmetro θ e é representada por $L(Y|\theta)$.

A distribuição a posteriori é obtida com base no Teorema de Bayes:

$$p(\theta | Y) = \frac{L(\theta | Y) \cdot p(\theta)}{\int L(\theta | Y) p(\theta) d\theta},$$

e, como o denominador não depende de θ temos:

$$p(\theta | Y) \propto L(\theta | Y) p(\theta),$$

em que \propto representa proporcionalidade.

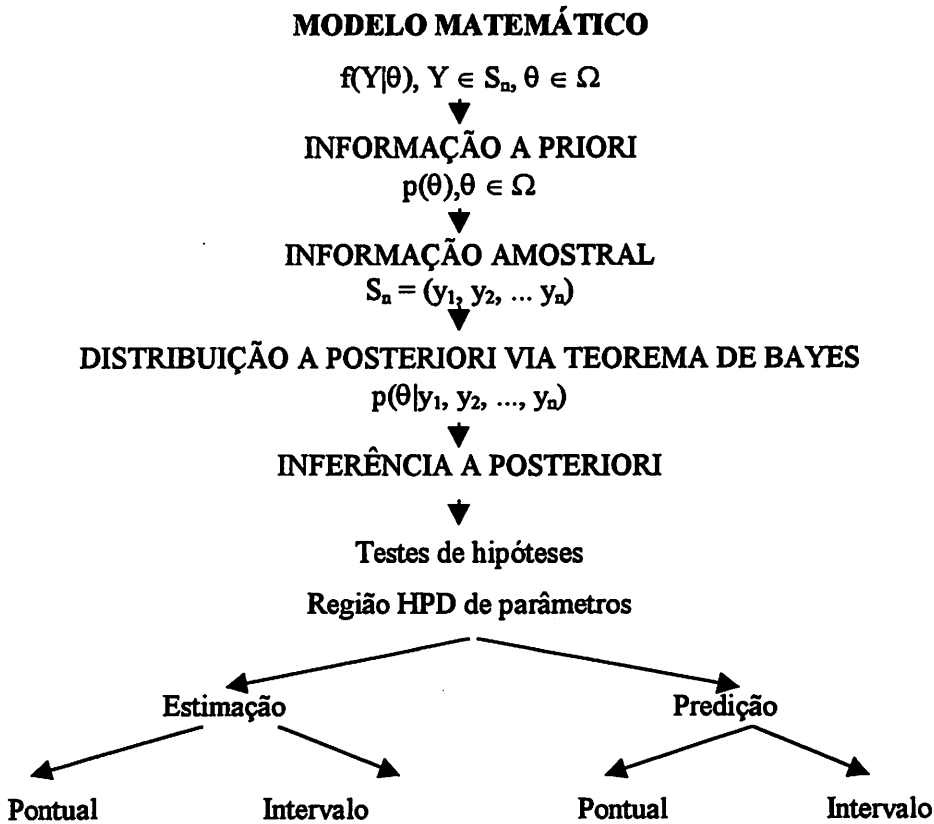
Em geral, a distribuição a posteriori $p(\theta|Y)$ não é fácil de ser calculada. Uma grande parte da literatura Bayesiana é dedicada a descobrir distribuições a priori nas quais $p(\theta|Y)$ pode ser facilmente trabalhada. Estas prioris são chamadas de prioris conjugadas, as quais têm como principal propriedade o fato de que as distribuições a priori e a posteriori pertencem à mesma classe de distribuições. Isto traz inúmeras vantagens para todo o processo de inferência decorrente essencialmente da simplificação devida à restrição da análise a um subconjunto de distribuições. Desta forma, a passagem da priori para a posteriori envolve uma simples mudança nos hiperparâmetros sem a necessidade de cálculo adicional, facilitando, assim, o processo de aprendizado à medida que novas observações são incorporadas.

Segundo West & Harrison (1997), a utilidade das prioris conjugadas, juntamente com suas propriedades atrativas, deve ser vista como de importância secundária quando comparada com a questão básica da possibilidade de uma priori conjugada ser escolhida, fornecendo uma aproximação razoável da “verdadeira” priori.

Outro ponto a ser considerado é a possibilidade da distribuição marginal a posteriori não possuir a forma de uma distribuição conhecida. Esta situação pode ser evitada utilizando prioris conjugadas, embora existam recursos baseados em cadeias de Markov para tratar destas possíveis posterioris. A idéia é que, desde que a distribuição a posteriori supostamente contém toda informação disponível sobre θ , qualquer inferência com relação a θ deve consistir somente das características desta distribuição. Jeffreys (1961), Zellner (1971) e Box & Tiao (1973) apresentam uma discussão mais aprofundada sobre os aspectos da inferência Bayesiana.

A representação probabilística de todo o conhecimento incerto é a essência da inferência Bayesiana, seja tal conhecimento relacionado ao futuro, quantidades observáveis ou parâmetros desconhecidos. Os conceitos de priori e posteriori são sempre relativos à observação considerada no momento.

Broemeling (1985) apresenta o seguinte diagrama para ilustrar a técnica utilizada na inferência Bayesiana:



2.2 – Simulação estocástica via Cadeias de Markov

O termo simulação refere-se ao tratamento de problemas reais a partir de reproduções em ambientes controlados pelo pesquisador, sendo que, muitas vezes, esses ambientes são os equipamentos computacionais. Alguns problemas apresentam componentes aleatórios, os quais não podem ser descritos de forma exata e sim baseados em informações probabilísticas. Para estes problemas, o

processo de simulação é estocástico, ou seja, baseado em distribuições de probabilidades.

Os métodos de simulação apresentam mecanismos determinísticos e iterativos. Existem vários algoritmos que utilizam a simulação com base em cadeias de Markov, dentre os quais podem ser citados os algoritmos EM, amostrador de Gibbs e Metropolis-Hasting.

Segundo Gamerman (1997), a teoria dos processos estocásticos, na qual a cadeia de Markov está inserida, é geralmente definida como a parte dinâmica da teoria das probabilidades, através da qual se estuda uma coleção de variáveis aleatórias, sob o ponto de vista de sua interdependência e de seu comportamento limite.

Amostrador de Gibbs

Geralmente, é de interesse do pesquisador obter as distribuições marginais a posteriori para os parâmetros em estudo. Suponha uma densidade conjunta $g(x, y_1, \dots, y_n)$ e que se está interessado nas características da sua densidade marginal

$$g(x) = \int \dots \int g(x, y_1, \dots, y_n) dy_1 \dots dy_n \quad (2.1)$$

tais como a média e a variância. A aproximação mais natural e direta seria calcular $g(x)$ e usá-la para obter as características desejadas. Entretanto, há muitos casos em que a integração de funções como em (2.1) é extremamente difícil de se obter, tanto analiticamente quanto numericamente. Em tais casos, o amostrador de Gibbs fornece um método alternativo para a obtenção de $g(x)$.

Em vez de computar ou aproximar $g(x)$ diretamente, o amostrador permite de maneira eficiente gerar uma amostra $x_1, \dots, x_m \sim g(x)$ sem a

necessidade de conhecer $g(x)$. Atuando como alternativa, o amostrador de Gibbs gera uma amostra grande o suficiente, de forma que a média, a variância ou qualquer outra característica de $g(x)$ possa ser calculada com o grau de confiança desejado.

Para a utilização do amostrador de Gibbs, é necessário o conhecimento das distribuições condicionais completas para cada parâmetro, as quais representam a distribuição do parâmetro condicional ao conhecimento de todos os outros. Seja:

$$g_i(\theta_i|\theta_{-i}) = g_i(\theta_i|\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$$

a distribuição condicional completa para θ_i dado o conhecimento dos outros $k-1$ parâmetros de θ . Para obtê-las, basta considerar os $k-1$ termos da posteriori conjunta como constantes conhecidas para, assim, determinar a distribuição condicional completa do vetor θ_i . A utilidade do amostrador aumenta à medida que a dimensão do problema aumenta, uma vez que ele evita o cálculo de integrais da forma (2.1)

Para descrever o algoritmo, suponha que a distribuição de interesse seja $g(\theta)$, em que $\theta = (\theta_1, \theta_2, \dots, \theta_k)$. Considere ainda que as densidades condicionais completas a posteriori $g_i(\theta_i|\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k) = g_i(\theta_i|\theta_{-i})$, $i = 1, \dots, k$ estão disponíveis. O problema que se coloca é o de geração de uma amostra de $g(\theta)$.

Segundo Gamerman (1997), o algoritmo pode ser descrito da seguinte forma:

- i) Inicialize o contador de iterações da cadeia $j = 1$ e arbitre valores iniciais $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_k^{(0)})$.
- ii) Obtenha um novo valor $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_k^{(j)})$ a partir de $\theta^{(j-1)}$ com base em sucessivas gerações de valores.

$$\theta_1^{(j)} \sim g_1(\theta_1|\theta_2^{(j-1)}, \dots, \theta_k^{(j-1)})$$

$$\theta_2^{(j)} \sim g_2(\theta_2|\theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_k^{(j-1)})$$

$$\theta_k^{(j)} \sim g_k(\theta_k|\theta_1^{(j)}, \dots, \theta_{k-1}^{(j)})$$

iii) Mude o contador j para $j + 1$ e retorne a (ii) até a convergência.

À medida que o número de iterações aumenta, a cadeia se aproxima de sua condição de equilíbrio, assim, assume-se que a convergência é atingida em uma iteração cuja distribuição esteja arbitrariamente próxima da distribuição de equilíbrio $g_i(\theta_i)$.

O esquema anterior define uma cadeia de Markov, pois os acontecimentos na iteração j dependem da história do processo apenas com base nos valores da iteração $j - 1$. Além disso, a cadeia é homogênea, pois o núcleo de transição não varia com j .

Convergência

Um aspecto importante na utilização do amostrador de Gibbs é a definição do número de iterações necessárias para que a cadeia atinja o equilíbrio.

A monitoração informal da convergência, apresentada por Gelfand e Smith (1990), sugere técnicas gráficas para verificação da convergência. Após um número suficientemente grande, N , de iterações, forma-se uma amostra de θ . O mesmo procedimento pode ser obtido após $N + k$ iterações. Se não houver diferença visual entre os gráficos obtidos após N e $N + k$ iterações, conclui-se pela convergência das cadeias. O valor de k não pode ser muito pequeno, pois a

correlação inerente à cadeia de Markov estará exercendo sua influência, não permitindo que se afirme se a similaridade é devida à convergência ou à correlação do processo. Por outro lado, valores de k muito grandes são desnecessários, pois se há suspeita de convergência após N iterações, não há necessidade de ir muito além na cadeia apenas para uma verificação. Tipicamente, valores entre 10 e 50 iterações são apropriados para o valor de k .

Vale ressaltar que estas técnicas de monitoração devem ser utilizadas com muita cautela e sempre acompanhadas de alguma fundamentação teórica.

A monitoração formal da convergência pode ser vista com base no método proposto por Gelman e Rubin (1992). Este método consiste na verificação a partir de várias cadeias em paralelo começando de diferentes pontos. Após a convergência, todas as cadeias devem ter o mesmo comportamento qualitativo e quantitativo. As cadeias são inicializadas em pontos que devem ser sobredispersos em relação à distribuição a posteriori, sendo que o número de cadeias deve ser maior do que 1.

Considerando m cadeias paralelas e uma função real $t(\theta)$, têm-se m trajetórias para t . Desta forma, podem ser obtidas as variâncias entre as cadeias (E) e a variância dentro das cadeias (D), dadas por:

$$E = \frac{n}{m-1} \sum_{i=1}^m (\bar{t}_i - \bar{t})^2 \quad \text{e} \quad D = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (t_i^{(j)} - \bar{t}_i)^2,$$

em que \bar{t}_i é a média das observações da cadeia i , \bar{t} é a média das observações de todas as cadeias e $t_i^{(j)}$ é a j -ésima observação da i -ésima cadeia. Sob convergência, todos os valores são gerados da posteriori e a variância de t pode ser estimada de forma não viesada por:

$$\hat{V}[t(\theta)] = \left(1 - \frac{1}{n}\right)D + \left(\frac{1}{n}\right)E$$

Se as cadeias ainda não tiverem atingido a convergência, então esta estimativa é maior que $V[t(\theta)]$, pois os valores iniciais ainda estarão influenciando. Por outro lado, D fornece estimativa menor que $V[t(\theta)]$, pois uma cadeia só não terá coberto toda a variabilidade de $t(\theta)$. Um indicador de convergência, dado pela redução potencial de escala $R = \sqrt{\hat{V}[t(\theta)]/D}$, deve ser sempre próximo do valor um. À medida que n cresce, ambos os estimadores acabarão convergindo para $V[t(\theta)]$ e R convergirá para 1; assim, R pode ser utilizado como indicador de convergência pela avaliação de sua proximidade a 1.

2.3 – Exemplos de processos de estimação Bayesiana

Esta seção apresenta considerações sobre diversos artigos e trabalhos da literatura relacionados aos processos de estimação via inferência Bayesiana, utilização e aplicação do amostrador de Gibbs, modelos de espaço de estados, entre outros.

Gelfand & Smith (1990) apresentam três alternativas para aproximações baseadas em Monte Carlo para o cálculo de estimativas numéricas de distribuições de probabilidade marginais. Os três métodos – substituição estocástica, amostrador de Gibbs e amostragem-reamostragem por importância – são comparados em relação a diferentes estruturas de probabilidades conjuntas freqüentemente encontradas em aplicações.

Gelfand et al (1990) ilustram a utilização de inferência Bayesiana em modelos com distribuições normais com base no amostrador de Gibbs. É

apresentada uma revisão do amostrador como método para se obter a distribuição marginal a posteriori, assim como a densidade preditiva. Como ilustração, é utilizado um conjunto de modelos normais, tais como os utilizados em componentes de variância, médias ordenadas e não ordenadas e curvas de crescimento hierárquicas, de forma a demonstrar a versatilidade e facilidade de implementação do amostrador de Gibbs na prática. Segundo os autores, a aproximação é direta na especificação da distribuição, trivial de ser implementada computacionalmente, fornecendo resultados através dos quais podem ser calculadas as quantidades necessárias. Ainda segundo os autores, o potencial da metodologia é enorme, representando fielmente as análises de um número de problemas vistos, até então como intratáveis sob o ponto de vista Bayesiano.

Casella & George (1992) apresentam simples explicações de como e por que o amostrador de Gibbs funciona. O amostrador é apresentado como uma técnica para a geração de variáveis aleatórias indiretamente a partir de uma distribuição marginal, sem ter calculado a sua densidade. Segundo os autores, embora seja direto descrevê-lo, o mecanismo que comanda o esquema iterativo parece misterioso. Ainda segundo os autores, o amostrador pode ser visto como uma implementação prática do fato que o conhecimento da distribuição condicional é suficiente para determinar a distribuição conjunta (se ela existir). Embora a teoria por trás do amostrador de Gibbs venha da teoria de cadeias de Markov, há também uma conexão com a teoria de “dados incompletos”, tal como as que formam as bases para o algoritmo EM.

Smith & Gelfand (1992) apresentam uma discussão com respeito à iniciação aos conceitos Bayesianos, os quais podem causar impacto negativo devido às integrações numéricas necessárias em todas aplicações. Eles apresentam uma perspectiva na inferência Bayesiana com base na amostragem-reamostragem, a qual pode ser facilmente implementada. Segundo os autores,

embora o mais usual seja gerar uma amostra a partir de uma densidade, também é possível gerar uma densidade a partir de uma amostra. Pelo Teorema de Bayes, obtém-se uma amostra a posteriori a partir de uma amostra a priori, com a verossimilhança agindo como uma probabilidade de reamostragem.

Carter & Kohn (1994) mostram como utilizar o Amostrador de Gibbs em inferência Bayesiana em modelos de espaço de estados, apresentando uma aproximação que gera simultaneamente todos os vetores de estado usando o filtro de Kalman. São apresentados os conceitos de Modelos Lineares Dinâmicos e a definição de suas componentes e quantidades desconhecidas, as quais são estimadas com base no uso do amostrador de Gibbs. Neste artigo, é utilizada a aproximação do amostrador de modo que todos os vetores de estado são gerados de uma só vez, aproveitando a ordenação do tempo de modelos de espaço de estados. São utilizados exemplos para comparar convergência da distribuição a posteriori quando os vetores de estado são gerados um a um, ou simultaneamente, sendo que, para este último, a convergência é mais rápida e apresenta menores variâncias.

Fruhwirth-Schnatter (1994) ilustra a forma com que métodos de “Data Augmentation” são ferramentas úteis para analisar MLD com hiperparâmetros desconhecidos. Para implementação prática, uma aproximação Monte Carlo é sugerida e baseada no algoritmo conhecido como “Forward Filtering, Backward Sampling – FFBS”, juntamente com uma discussão sobre como este método está relacionado com o Amostrador de Gibbs

Wakefield et al (1994) apresentam uma análise Bayesiana completa aplicada em modelos populacionais lineares e não lineares. É ressaltada a importância da utilização do amostrador de Gibbs como alternativa para obter as distribuições marginais a posteriori, uma vez que soluções analíticas para estes tipos de estudos, com estrutura complexa e com muitos parâmetros, se tornam quase que impraticáveis. Segundo os autores, o amostrador parece oferecer o

método mais flexível e poderoso disponível atualmente para análises de problemas envolvendo modelos populacionais. Entretanto, a aproximação pode apresentar uma performance menos eficiente em certas circunstâncias, tais como parametrização desnecessária (alta correlação a posteriori ou convergência lenta) ou posterioris com multimodalidade (regiões desconectas do espaço paramétrico).

West (1995) apresenta uma revisão de variações de Modelos Lineares Dinâmicos, mencionando aspectos práticos para tratar de dados omissos, observações adicionais, erros amostrais, “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 estado, utilizados na análise de registros de isótopos de oxigênio no fundo dos oceanos. É apresentado o Modelo Linear Dinâmico com suas equações de observação e evolução, utilizando como exemplo aplicações envolvendo componentes autorregressivos. A obtenção das posterioris para a análise da série temporal é feita por meio de métodos MCMC, mais especificamente variações do amostrador de Gibbs e do algoritmo Metropolis-Hasting, em que são apresentados métodos para amostragem das quantidades desconhecidas.

Fruhwirth-Schnatter (1995) apresenta como discriminar entre diferentes modelos de espaço de estados para uma dada série temporal a partir de uma aproximação Bayesiana, a qual escolhe o modelo que minimiza a perda esperada. A implementação deste procedimento necessita de uma análise Bayesiana completa tanto para os vetores de estado quanto para os hiperparâmetros desconhecidos, a qual é realizada por métodos MCMC.

Como aplicação, Glickman & 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 estes parâmetros se comportavam como um processo autorregressivo. Os autores utilizam MCMC, mais especificamente o amostrador de Gibbs, como uma ferramenta computacional para o cálculo das distribuições a posteriori para os parâmetros do modelo. As previsões com base no modelo proposto se comportaram de maneira satisfatória em relação aos resultados obtidos no decorrer da liga.

2.4 – Modelos Lineares Dinâmicos

Segundo West & Harrison (1997), os princípios empregados na previsão Bayesiana e modelagem dinâmica envolvem:

- a) Modelos paramétricos com parâmetros de significado dinâmico;
- b) Representação probabilística da informação sobre os parâmetros;
- c) Definição de um modelo seqüencial utilizando independência condicional;
- d) Componentes dos modelos robustas e condicionalmente independentes;
- e) Previsões derivadas como distribuições de probabilidades;
- f) Facilidade de incorporação de informações importantes;
- g) Controle da qualidade do modelo.

Para fins de notação, considere que a quantidade D_t representa todo o conhecimento adquirido até o tempo t . Assim, em um sistema fechado, tem-se que $D_t = \{Y_t, D_{t-1}\}$, ou seja, a cada tempo t , a única informação nova é a observação Y_t .

No tempo t , a informação referente ao tempo anterior D_{t-1} funciona como uma priori ao observar Y_t . Entretanto, ela também funciona como uma posteriori em relação ao tempo $t-1$. Desta forma, considerando θ_t como a quantidade desconhecida, vê-se que a priori $p(\theta_t|D_{t-1})$ e a posteriori $p(\theta_t|D_t)$ representam uma transferência concisa da informação entre dois tempos consecutivos. Além disso, a distribuição $(\theta_0|D_0)$ funciona como uma priori, uma vez que se atribui uma distribuição para o vetor θ_0 antes que a amostra seja observada.

Na utilização da aproximação Bayesiana, é interessante selecionar uma parametrização, θ_{t-1} , tal que toda informação histórica relevante em prever observações futuras esteja contida na informação sobre θ_{t-1} . A representação desta informação em termos de distribuição de probabilidades é dada por $p(\theta_{t-1}|D_{t-1})$; ou seja, dado que se conhece D_{t-1} , a posteriori $p(\theta_{t-1}|D_{t-1})$ é suficiente em prever o futuro. Nota-se que todo o contexto no tempo $t-1$ está representado pelos parâmetros e pela distribuição de probabilidade.

Desta forma, é necessário relacionar a informação disponível com o futuro, de modo que distribuições com $(Y_{t+k}|D_{t-1})$ possam ser derivadas. A estrutura crucial que possibilita a efetiva modelagem dinâmica é a independência condicional, em que a característica principal é que dado o presente, o futuro é independente do passado. Assim, em um dado tempo t , dado o conhecimento de θ_t , o passado, presente e o futuro são mutuamente independentes. Da mesma forma, dado D_t , toda informação com relação ao futuro está contida na distribuição a posteriori $(\theta_t|D_t)$. As previsões podem ser derivadas via:

$$p(Y_t, \theta_t|D_{t-1}) = p(Y_t|\theta_t, D_{t-1})p(\theta_t|D_{t-1}).$$

A previsão 1 passo à frente é simplesmente a distribuição marginal $p(Y_t|D_{t-1})$.

O MLD geral é caracterizado pelo conjunto de quádrupla $\{F, G, V, W\}_t$, para cada tempo t , onde

- A) F_t é uma matriz conhecida ($np \times r$);
- B) G_t é uma matriz conhecida ($np \times np$);
- C) V_t é uma matriz de variâncias conhecida ($r \times r$);
- D) W_t é uma matriz de variâncias conhecida ($np \times np$);

em que r = número de observações do vetor Y_t ($r \times 1$) e np = número de parâmetros.

A relação entre Y_t e os parâmetros do vetor θ_t ($np \times 1$) é feita seqüencialmente a partir do par de equações mostradas abaixo:

$$\text{Equação de observação: } Y_t = F_t' \theta_t + v_t \quad v_t \sim N[0, V_t],$$

$$\text{Equação de sistema: } \theta_t = G_t \theta_{t-1} + \omega_t \quad \omega_t \sim N[0, W_t],$$

e completado pela informação inicial $(\theta_0 | D_0) \sim N[m_0, C_0]$.

A equação das observações define a distribuição amostral para Y_t condicional ao conhecimento da quantidade θ_t ; assim, dado θ_t , Y_t é independente de todas as outras observações e valores dos parâmetros. No tempo t :

- a) F_t é a matriz de planejamento;
- b) θ_t é o vetor de estado;
- c) $\mu_t = F_t' \theta_t$ é a resposta média, ou nível;
- d) v_t é o erro observacional.

A equação do sistema representa a forma com que o vetor de estado evolui com o tempo. A propriedade de independência condicional define uma

evolução de Markov em 1 passo, de modo que, dado θ_{t-1} e os valores de G_t e W_t , a distribuição de θ_t é independente de D_{t-1} . No tempo t :

- e) G_t é a matriz de evolução;
- f) ω_t é o erro de evolução.

DEFINIÇÃO 2.1 – Dois subconjuntos da classe de MLD geral são de interesse especial:

- i) Se $\{F, G\}_t$ são constantes para todo t , então o modelo é definido como MLD de séries temporais, ou MLDST;
- ii) Um MLD cujas variâncias de observação e de evolução são constantes para todo t é conhecido como MLD constante.

Esta importante subclasse de MLD inclui essencialmente todos os modelos clássicos de séries temporais lineares. Considera-se, agora, o caso de um MLD univariado, onde Y_t representa um único valor.

DEFINIÇÃO 2.2: Para cada tempo t , o MLD geral univariado é definido por:

$$\text{Equação de observação: } Y_t = F_t \theta_t + v_t \quad v_t \sim N[0, V_t]$$

$$\text{Equação de sistema: } \theta_t = G_t \theta_{t-1} + \omega_t \quad \omega_t \sim N[0, W_t]$$

$$\text{Informação inicial: } (\theta_0 | D_0) \sim N[m_0, C_0]$$

No MLD univariado, modela-se cada observação Y_t pela equação de observação condicional ao conhecimento do vetor de estados θ_t . Os erros v_t e ω_t são independentes e mutuamente independentes. Modelos mais gerais podem ser

obtidos permitindo-se que as seqüências de erros ω_t e v_t possam ser correlacionadas e, neste caso, é sempre possível reformular este tipo de modelo de modo que a suposição de independência seja satisfeita, o que faz com que nada seja perdido pela imposição desta restrição. 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, ω_t influencia o desenvolvimento do sistema no futuro e a sua magnitude em relação à v_t nos diz o tamanho da variação do nível médio da série.

Equações de atualização

As equações de atualização representam toda a dinamicidade desta 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. A vantagem da utilização de modelos com distribuição normal é que em qualquer momento do tempo, toda a informação existente relativa ao sistema é representada e suficientemente sumarizada pela distribuição a posteriori para o vetor de estado atual. O Teorema 2.1 apresenta estas equações, além da previsão 1 passo à frente para $(Y_t|D_{t-1})$.

TEOREMA 2.1: No MLD univariado apresentado na definição 2.2, 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 a frente: $(Y_t|D_{t-1}) \sim N[f_t, Q_t]$

$$f_t = F_t' a_t \qquad Q_t = F_t' R_t F_t + V_t$$

d) Posteriori em t: $(\theta_t|D_t) \sim N[m_t, C_t]$

$$A_t = R_t F_t / Q_t \qquad e_t = Y_t - f_t$$

$$m_t = a_t + A_t e_t \qquad C_t = R_t - A_t Q_t A_t'$$

Nota-se, assim, que a partir do conhecimento da distribuição a priori $\theta_0|D_0 \sim N(m_0, C_0)$, é possível obter as distribuições $(\theta_1|D_1)$, ..., $(\theta_n|D_n)$.

Algumas identidades são facilmente derivadas com base nas quantidades descritas nas equações de atualização, as quais são mostradas a seguir:

- a) $A_t = R_t F_t / Q_t = C_t F_t / V_t$
- b) $C_t = R_t - A_t Q_t A_t' = R_t (I - F_t A_t')$
- c) $C_{t-1} = R_{t-1} + F_{t-1} V_{t-1}^{-1} F_{t-1}'$
- d) $Q_t = (1 - F_t' A_t)^{-1} V_t$
- e) $F_t' A_t = 1 - V_t Q_{t-1}$

onde A_t é o vetor adaptativo no tempo t.

As demonstrações para as equações apresentadas no Teorema 2.1 podem ser encontradas em West & Harrison (1997, pág 104).

Variância observacional V_t

Sabe-se que um MLD fica completamente definido a partir do conhecimento da quádrupla $\{F, G, V, W\}_t$. Dessas quatro componentes, F_t e G_t são definidas pelo pesquisador dependendo do tipo de situação em que o estudo está inserido. A variância de evolução também é, geralmente, definida pelo pesquisador com base no uso de fatores de desconto, como será visto mais adiante. O fator remanescente, a variância observacional V_t , é freqüentemente desconhecida e grande em relação à variância do sistema W_t .

Assumindo variância observacional constante e trabalhando em termos da precisão $\phi = V^{-1}$, tem-se que a análise Bayesiana é desenvolvida com a utilização de uma priori específica para $\phi = V^{-1}$, uma vez que V é desconhecida.

DEFINIÇÃO 2.3: Para cada tempo t , escrevendo $\phi = V^{-1}$, o MLD geral é definido por:

$$\text{Equação de observação: } Y_t = F'_t \theta_t + v_t \quad v_t \sim N[0, V]$$

$$\text{Equação do sistema: } \theta_t = G_t \theta_{t-1} + \omega_t \quad \omega_t \sim N[0, W_t]$$

$$\text{Informação inicial: } (\theta_0 | D_0, \phi) \sim N[m_0, C_0]$$

$$(\phi | D_0) \sim G[n_0/2, n_0 S_0/2]$$

Vê-se que agora é acrescentada uma distribuição para a quantidade $\phi = V^{-1}$. Desta forma, ela também será atualizada, fornecendo estimativas para o parâmetro V . Os valores de m_0 , C_0 , n_0 , S_0 e as matrizes F_t , G_t , e W_t são especificadas pelo pesquisador. Tem-se ainda que $E[\phi | D_0] = 1/S_0$, onde S_0 é a estimativa pontual a priori para a variância observacional V .

TEOREMA 2.2: Para o MLD apresentado na definição 2.3, os resultados das distribuições obtidas a cada tempo t , $t \geq 1$ são dados por:

a) Condicional ao conhecimento de V

$$(\theta_{t-1}|D_{t-1}, V) \sim N[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}]$$

$$(\theta_t|D_{t-1}, V) \sim N[\mathbf{a}_t, \mathbf{R}_t]$$

$$(Y_t|D_{t-1}, V) \sim N[\mathbf{f}_t, Q_t]$$

$$(\theta_t|D_t, V) \sim N[\mathbf{m}_t, \mathbf{C}_t]$$

onde, $\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t$; $\mathbf{f}_t = \mathbf{F}_t' \mathbf{a}_t$; $Q_t = 1 + \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t$; $\mathbf{e}_t = Y_t - \mathbf{f}_t$;

$$\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t' / Q_t; \quad \mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1},$$

com $\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t \mathbf{e}_t$ e $\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t' / Q_t$.

b) Atualização da distribuição para ϕ

$$(\phi|D_{t-1}) \sim G[n_{t-1}/2, n_{t-1} S_{t-1}/2],$$

$$(\phi|D_t) \sim G[n_t/2, n_t S_t/2],$$

onde $n_t = n_{t-1} + 1$ e $n_t S_t = n_{t-1} S_{t-1} + \mathbf{e}_t^2 / Q_t$.

c) Incondicional a V

$$(\theta_{t-1}|D_{t-1}) \sim t_{n_{t-1}}[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}],$$

$$(\theta_t|D_{t-1}) \sim t_{n_{t-1}}[\mathbf{a}_t, \mathbf{R}_t],$$

$$(Y_t|D_{t-1}) \sim t_{n_{t-1}}[\mathbf{f}_t, Q_t],$$

$$(\theta_t|D_t) \sim t_{n_t}[\mathbf{m}_t, \mathbf{C}_t],$$

onde: $\mathbf{C}_t = S_t \mathbf{C}_t$, $\mathbf{R}_t = S_{t-1} \mathbf{R}_t$, $Q_t = S_{t-1} Q_t$.

d) Definição operacional das equações de atualização

Sejam $Q_t = F_t' R_t F_t + S_{t-1}$ e $A_t = R_t F_t / Q_t$, assim:

$$n_t = n_{t-1} + 1 \quad \text{e} \quad S_t = S_{t-1} + \frac{S_{t-1}}{n_t} \left(\frac{e_t^2}{Q_t} - 1 \right)$$

$$m_t = a_t + A_t e_t \quad \text{e} \quad C_t = \frac{S_t}{S_{t-1}} (R_t - A_t A_t' Q_t)$$

No tempo t , a média da priori de ϕ é $E[\phi | D_{t-1}] = n_{t-1} / d_{t-1} = 1 / S_{t-1}$, onde $S_{t-1} = d_{t-1} / n_{t-1}$ é uma estimativa pontual a priori para $V = \phi^{-1}$. Da mesma forma, a estimativa a posteriori é $S_t = d_t / n_t$.

As demonstrações para as equações apresentadas no Teorema 2.2 podem ser encontradas em West & Harrison (1997, pág 110).

Variância de Evolução W_t

A especificação da seqüência da matriz de evolução W_t é geralmente realizada com o auxílio de fatores de desconto. A especificação da estrutura desta 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.

A matriz W_t leva a um aumento na incerteza, ou perda de informação, entre um vetor de estado no tempo t e $t + 1$.

Pelas equações de atualização do Teorema 2.1, tem-se:

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

$V[\theta_t|D_{t-1}] = G_t C_{t-1} G_t' + W_t$, que é a variância a priori R_t .

$V[Y_t|D_{t-1}] = Q_t = F_t' R_t F_t + V_t$

A matriz W_t indica quão durável é o sistema, uma vez que quanto maior for a sua magnitude, menor será a precisão em relação às previsões um passo à frente. Se $W_t = 0$, o modelo é confiável. Caso contrário, para $W_t \rightarrow \infty$, o modelo se torna sem controle e, por consequência, sem confiança. W_t mede a rapidez com que a informação atual D_t decai com k passos. Segundo West & Harrison (1997), alguns problemas surgem na utilização da matriz W_t :

- a) Não é invariante com relação à escala da medida como especificado em F_t ,
- b) É ambígua,
- c) Efeitos dos componentes devem satisfazer restrições, e isto demanda que W_t também satisfaça as restrições,
- d) A durabilidade local do MLD varia com o tempo, assim, não há um valor “ótimo” para W_t para todo t ,
- e) A maioria dos pesquisadores tem dificuldade em quantificar os elementos da variância e covariância.

Para contornar estes problemas 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 $0 < \delta \leq 1$, sendo que, geralmente, assume valores próximos de 1.

Para compreender o funcionamento do desconto em um dado tempo t , defina $P_t = G_t C_{t-1} G_t'$. Pelas equações de atualização, tem-se que $(\theta_{t-1}|D_{t-1}) \sim N[m_{t-1}, C_{t-1}]$. A precisão associada a θ_{t-1} é C_{t-1}^{-1} , e para $G\theta_{t-1}$, é P_t^{-1} . Assim, P_t^{-1} representa a precisão associada com 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 R^1_t , é reduzida em relação a P^1_t .

O conceito de desconto define esta decaída diretamente, via δP^1_t , ou simplesmente uma proporção δ da precisão. Assim,

$$V[\theta_t|D_{t-1}] = R_t = (1/\delta)P_t,$$

e como $R_t = P_t + W_t$ tem-se que:

$$(1/\delta)P_t = P_t + W_t,$$

$$\text{ou então, } W_t = [(1-\delta)/\delta]P_t.$$

Desta forma, dado δ e C_0 , a série $\{W_t\}$ fica totalmente identificada e quanto mais próximo de 1 for o fator de desconto, mais durável é o modelo.

Filtragem

Geralmente é interessante olhar para trás no tempo e ver o que aconteceu, ou seja, o interesse agora está nos vetores de estado $\theta_t, \theta_{t-1}, \dots$. A análise retrospectiva contribui para o aumento do conhecimento, desenvolvimento e performance do modelo. Este processo é chamado de filtragem.

A distribuição de $(\theta_{t-k}|D_t)$ para $k \geq 1$ é chamada de filtragem k – passos. Um outro conceito relacionado é o da suavização, em que a resposta média μ_t pode ser filtrada utilizando a distribuição $(\mu_{t-k}|D_t)$ para $k \geq 1$.

TEOREMA 2.3: - No MLD univariado $\{F_t, G_t, V_t, W_t\}$, defina para todo t

$$B_t = C_t G'_{t+1} R^{-1}_{t+1}.$$

Então, para todo k ($1 \leq k < t$), as distribuições marginais filtradas são definidas por:

$$(\theta_{t-k}|D_t), \sim N[a_t(-k), R_t(-k)]$$

$$\text{onde } a_t(-k) = m_{t-k} + B_{t-k}[a_t(-k+1) - a_{t-k+1}]$$

$$\text{e } R_t(-k) = C_{t-k} + B_{t-k}[R_t(-k+1) - R_{t-k+1}]B'_{t-k}$$

$$\text{com valores iniciais } a_t(0) = m_t, R_t(0) = C_t,$$

$$\text{e } a_{t-k}(1) = a_{t-k+1} \text{ e } R_{t-k}(1) = R_{t-k+1}$$

As demonstrações para as equações apresentadas no Teorema 2.3 podem ser encontradas em West & Harrison (1997, pág 114).

2.5 – Modelos Lineares Dinâmicos em Séries Temporais

Para apresentar o ajuste de Séries Temporais com base nos Modelos Lineares Dinâmicos, começa-se com a definição para modelos de regressão, alguns conceitos de Séries Temporais, como estacionaridade e invertibilidade, para então definir os modelos autorregressivos na forma de MLD.

O MLD de regressão múltipla consiste em modelar Y_t por regressão, com relação a n variáveis independentes X_1, X_2, \dots, X_n , sendo que os valores da

i -ésima variável X_i em cada tempo t são assumidos conhecidos e denotados por X_{it} . O vetor F_t é dado por $F_t' = (X_{t1}, \dots, X_{tm})$.

Assim, para todo tempo t , o MLD é definido por:

$$\text{Eq. observações: } Y_t = F_t' \theta_t + v_t \quad ; v_t \sim N[0, V_t];$$

$$\text{Eq. evolução: } \theta_t = \theta_{t-1} + \omega_t \quad ; \omega_t \sim N[0, W_t];$$

A observação Y_t pode ser vista ainda como a combinação de um efeito médio além de um ruído aleatório $\mu_t + v_t$, onde μ_t é dado pela quantidade $F_t' \theta_t$. Quando a variância de evolução $W_t = 0$, para todo t , tem-se que o vetor de estados θ_t é constante para todo o tempo t , ou seja, tem-se apenas a equação de observação, como visto a seguir:

$$\text{Eq. observações: } Y_t = F_t' \theta + v_t \quad ; v_t \sim N[0, V_t];$$

e θ funciona como um vetor de regressão fixo.

Segundo West & Harrison (1997), a estrutura de MLD permite várias formas de relacionamento com base na escolha apropriada das variáveis regressoras e suas combinações. Alguns exemplos são:

- a) Regressão linear em variáveis quantitativas,
- b) Regressão múltipla em variáveis quantitativas,
- c) Variáveis “lagged”,
- d) Superfícies polinomiais,
- e) Variáveis classificatórias,
- f) Vários fatores,
- g) Interações fator por fator,
- h) Outras formas de interação.

A performance na previsão é alcançada a partir da identificação de estabilidade nas relações de regressão. Assim, modelos com pequeno ruído de evolução são desejados. Segundo West & Harrison (1997), embora previsões a “curto prazo” possam ser acuradas e precisas, previsões a “médio” e “longo prazo” podem ser ruins localmente e muito difusas. Como consequência, as matrizes W_t , determinadas pelo pesquisador serão relativamente pequenas.

Estruturar W_t por fatores de desconto pode ser feito seguindo as estruturas apresentadas anteriormente na seção 2.4, na parte referente à estrutura desta matriz.

A regressão estática é obtida quando $W_t = 0$ para todo t , fato este implicado por um $\delta = 1$. Neste caso, $R_t = C_{t-1}$ para qualquer t e não há decaimento de informação no vetor de estado entre as observações.

Estacionaridade e invertibilidade

Considere a equação de observação dada por $Y_t = \mu_t + X_t + v_t$, onde $\mu_t = F_t' \theta_t$ é a função de resposta média e v_t é a variância observacional; X_t é um novo termo introduzido para descrever qualquer variação adicional não explicada por μ_t e pelo modelo de evolução para o vetor de estado θ_t .

Esta variação adicional pode ser representada por um modelo para X_t como um processo estocástico no tempo, cujos valores apresentam algum tipo de dependência. Uma característica bastante importante na modelagem do componente X_t é a presença de estacionaridade, que significa assumir simetria em relação ao tempo. Para um dado modelo M , as condições para que ele seja estacionário são:

$$1 - p(X_{t_1}, \dots, X_{t_k}|M) = p(X_{s+t_1}, \dots, X_{s+t_k}|M),$$

ou seja, a distribuição conjunta, dado M , de qualquer conjunto de k valores é invariante com respeito à arbitrariedade da escolha do eixo do tempo;

$$2 - E[X_t|M] = \mu, \text{ constante no tempo};$$

$$3 - V[X_t|M] = W, \text{ constante no tempo};$$

$$4 - C[X_t, X_{t-s}|M] = \gamma_s = \rho_s W, \text{ que é independente do tempo.}$$

Como $\rho_0 = 1$, então $\gamma_0 = W$, onde ρ é a função de autocorrelação e γ a função de autocovariância.

Uma vantagem importante de se ter uma série estacionária é que X_t pode ser decomposta na soma de uma média e uma combinação linear de valores em uma seqüência, ε_t , com média zero e não correlacionada, dada por

$$X_t = \sum_{r=0}^{\infty} \psi_r \varepsilon_{t-r}, \quad (2.2)$$

$$\text{onde, } E[\varepsilon_t|M] = 0;$$

$$V[\varepsilon_t|M] = U;$$

$$C[\varepsilon_t, \varepsilon_s|M] = 0, \text{ qualquer } t \neq s \text{ e } U$$

e para alguma seqüência de coeficientes ψ_0, ψ_1, \dots , com $\psi_0 = 1$. Como

$$V[X_t|M] = W, \text{ } W \text{ pode ser escrito como } U \sum_{r=0}^{\infty} \psi_r^2, \text{ de modo que a soma do}$$

quadrado dos coeficientes deve convergir e a autocovariância γ_s deve ser escrita em termos da soma de produtos de ψ_r , que também deve convergir.

O operador retardo (B) utilizado para simplificação de notação é definido pela relação $B^r X_t = X_{t-r}$; assim, (2.2) fica definida por:

$$X_t = \mu + \sum_{r=0}^{\infty} \psi_r B^r \varepsilon_t = \mu + \psi(B) \varepsilon_t$$

onde $\psi(\cdot)$ é uma função polinomial dada por $\psi(x) = 1 + \psi_1x + \psi_2x^2 + \dots$, a qual também pode ser escrita por $\phi(B)(X_t - \mu) = \varepsilon_t$, onde $\phi(\cdot)$ é outra função polinomial em que $\phi(x) = 1 - \phi_1x - \phi_2x^2 - \dots$, satisfazendo $\psi(x)\phi(x) = 1$.

Se esta representação é possível, ou seja, $\psi(x)\phi(x) = 1$, então $\psi(\cdot)$ é dita como sendo invertível.

Modelos autorregressivos

Na equação a seguir:

$$X_t = \mu + \sum_{r=1}^p \phi_r (X_{t-r} - \mu) + \varepsilon_t .$$

suponha que $\phi_r = 0$ para $r > p$; para $p \geq 1$ e inteiro. Assim, X_t é conhecido como um processo autorregressivo de ordem p e denotado por $X_t \sim \text{AR}(p)$.

O modelo autorregressivo de ordem p é um processo Markoviano com $p(X_t | X_{t-1}, \dots, X_{t-p}, X_{t-p-1}, \dots, M) = p(X_t | X_{t-1}, \dots, X_{t-p}, M)$, onde os valores de X com mais de p “passos” atrás não exercem influência no seu valor.

Considere um AR (1); assim, $p = 1$ e $X_t = \mu + \phi_1(X_{t-1} - \mu) + \varepsilon_t$. Além disso, $W = \phi_1^2 W + U$, então, se $|\phi_1| < 1$, $W = U/(1 - \phi_1^2)$. Desta forma, para que um processo AR(1) seja estacionário, deve-se ter $|\phi_1| < 1$, o que faz com que $\rho_s = \phi^s$.

Aqui, $\phi(x) = 1 - \phi_1x$, com $\psi(x) = 1 + \phi_1x + \phi_1^2x^2 + \dots = 1/\phi(x)$, fazendo com que seja invertível.

O modelo autorregressivo-médias móveis de ordem p e q denotado por ARMA(p,q) para X_t é definido por:

$$X_t = \mu + \sum_{r=1}^p \phi_r (X_{t-r} - \mu) + \sum_{r=1}^q \psi_r \varepsilon_{t-r} + \varepsilon_t , \quad \varepsilon_t \sim N[0, U].$$

Todos os modelos ARMA podem ser escritos na forma de MLD de várias maneiras, dependendo da forma das componentes F , G e θ .

Na prática, ao utilizar modelos ARMA, pode ser necessário transformar a série Y_t com o objetivo de se chegar a uma série X_t que seja, aproximadamente, estacionária e com média zero. A diferença de 1ª ordem em Y_t pode ser definida por $(1 - B)Y_t = Y_t - Y_{t-1}$ e as diferenças de ordem superior, $d > 1$, são derivadas de forma similar. Se o componente sazonal de período “s” está presente, então Y_t deve ser diferenciada primeiramente em relação ao componente sazonal da forma $(1 - B^s)Y_t$. Assim, a série X_t pode ser vista como $X_t = (1 - B)^d(1 - B^s)Y_t$ para algum “d” e “s”.

Uma vantagem ao se utilizarem os conceitos de MLD é que se pode modelar a série original Y_t diretamente, representando a tendência, regressão, sazonalidade e componentes residuais sem a necessidade da transformação nos dados. Para exemplificar, considere a série Y_t modelada em função de uma tendência polinomial μ_t com um processo correlacionado X_t e erro observacional aditivo. Assim, tem-se:

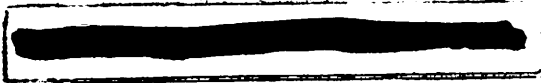
$$Y_t = \mu_t + X_t + v_t$$

com μ_t e X_t podendo ser representados em termos de componentes de MLD. Suponha μ_t como tendo uma tendência polinomial de 1ª ordem e ω_t o erro de evolução desta tendência. Assuma ainda que X_t pode ser representado por um processo AR(p). Tem-se, então:

$$Y_t = F'\theta_t + v_t, \quad v_t \sim N[0, V]$$

$$\theta_t = G\theta_{t-1} + \omega_t, \quad \omega_t \sim N[0, W]$$

Com $\theta_t = (\mu_t, X_t, X_{t-1}, \dots, X_{t-p+1})'$ sendo o vetor de estado $(p+1) \times 1$, $F = (1, 1, 0, \dots, 0)'$,



$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \phi_1 & \phi_2 & \phi_3 & \dots & \phi_p \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} \text{ e } \mathbf{W}^* = \begin{bmatrix} W & 0 & 0 & \dots & 0 \\ 0 & U & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

Note que o processo autorregressivo é não observável ou latente. Desta forma, a análise com base em um MLD leva a um aprendizado seqüencial em X_t , em adição ao componente de tendência. Assim, a teoria normal padrão se aplica condicionalmente ao conhecimento dos valores específicos dos parâmetros ϕ 's. Se estes parâmetros são desconhecidos, o problema de estimação e inferência se torna essencialmente não linear, de modo que aproximações analíticas ou métodos numéricos, como o amostrador de Gibbs, se tornam necessários.

2.6 – Simulação em Modelos Lineares Dinâmicos

Suponha que θ representa um conjunto de quantidades de interesse e que, com base na informação inicial a priori e nos dados observados, tem-se a distribuição a posteriori $p(\theta|Y)$.

Suponha que, por simulação, são realizadas amostras Monte Carlo de tamanho m : $\theta_1, \theta_2, \dots, \theta_m$. Inferências do tipo médias, histogramas, etc... são calculadas com base simplesmente na lei dos grandes números, sob condições geralmente aplicáveis, de modo que as amostras convergem para a esperança “exata” a posteriori, ou seja:

$$m^{-1} \sum_{i=1}^m g(\theta_i) = \int g(\theta) p(\theta) d\theta = E[g(\theta)]$$

quando $m \rightarrow \infty$ e para qualquer função $g(\cdot)$ especificada.

Na prática, a simulação direta pode se tornar impossível para as distribuições a posteriori cujas formas podem ser matematicamente complicadas. Na década de 90, a estatística Bayesiana desenvolveu diversos esquemas baseados em métodos MCMC para tratar com simulação a posteriori quando os métodos diretos são impraticáveis.

O método MCMC consiste em simular sucessivamente valores do vetor θ . Assim, uma seqüência de valores simulados $\theta_1, \theta_2, \dots$ é gerada por:

- 1 – Especificando um valor inicial θ_1 e,
- 2 – Amostra de valores sucessivos de uma distribuição de transição com densidade $f(\theta_i|\theta_{i-1})$, para $i = 2, 3, \dots$. θ_i é gerado condicionalmente independente de $\theta_{i-2}, \theta_{i-3}, \dots$

Os diferentes métodos MCMC diferem na escolha da distribuição de transição $f(\cdot|\cdot)$. É importante ressaltar que a seqüência MCMC dos valores de θ são dependentes de 1ª ordem, e não amostras aleatórias.

A cadeia de Markov converge no sentido de que a seqüência de valores parece estacionária e os θ_i 's são distribuídos marginalmente de acordo com uma distribuição única e estacionária. O esquema MCMC escolhe densidades de transição $f(\cdot|\cdot)$, de modo a garantir que a distribuição estacionária é o verdadeiro "alvo" a posteriori; assim, a simulação é feita até que a escolha do valor inicial seja irrelevante e o processo tenha se tornado estacionário.

Cadeias de Markov Monte Carlo aplicadas em modelos dinâmicos

Sabe-se que um MLD é definido pela seqüência de distribuições de observação e evolução, $p(Y_t|\theta_t, D_{t-1})$ e $p(\theta_t|\theta_{t-1}, D_{t-1})$, sendo que a especificação do sistema é completada pela priori $p(\theta_0|D_0)$.

Com base no conjunto de observações Y_1, Y_2, \dots, Y_n , o interesse se concentra na posteriori conjunta $p(\theta_1, \dots, \theta_n|D_n)$. Uma versão do amostrador de Gibbs propõe que se reamostrem iterativamente posterioris condicionais completas:

$$p(\theta_t|\theta_{(-t)}, D_n), \text{ para } t = 1, \dots, n.$$

Pelo Teorema de Bayes, vê-se que:

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

como função de θ_t .

Um problema técnico que pode surgir com a aplicação do amostrador de Gibbs com muitos parâmetros e variáveis desconhecidas é que a convergência pode ser lenta. Nos MLD estudados anteriormente, o amostrador de Gibbs se comporta possibilitando uma maneira natural e simples para amostrar a distribuição a posteriori dos parâmetros do modelo e dos vetores de estado em um intervalo de tempo fixo.

A importância e utilidade desta estrutura são evidentes considerando MCMC baseado no amostrador de Gibbs. Defina $\Theta_n = \{\theta_0, \theta_1, \dots, \theta_n\}$ para os n vetores de estado do MLD para as observações Y_1, \dots, Y_n junto com o vetor inicial θ_0 e $\alpha_n = \{\phi, V, W \text{ e } U\}$.

O amostrador de Gibbs sugere que a distribuição a posteriori completa

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

pode ser simulada interagindo entre duas posteriores condicionais

$$p(\Theta_n | \alpha_n, D_n) \leftrightarrow p(\alpha_n | \Theta_n, D_n). \quad (2.3)$$

Este esquema envolve a amostragem de Θ_n e, separadamente α_n , a partir das posteriores condicionais multivariadas completas, em vez de fazer seqüencialmente a partir de elementos escalares individuais como na definição original do amostrador.

Como a análise é padrão condicional a α_n , a simulação do 1º componente de (2.3) será acessível como uma distribuição normal multivariada, sendo que o segundo componente depende muito da forma do modelo.

Em alguns casos, a amostragem direta de $p(\alpha_n | \Theta_n, D_n)$ pode ser complicada, de forma que a divisão de α_n em k subconjuntos pode ser feita sem maiores problemas. Assim, tem-se:

$$\begin{aligned} \Theta_n | \alpha_n, D_n & \\ & \leftrightarrow \alpha_{n,1} | \alpha_{n,(-1)}, \Theta_n, D_n \\ & \dots \\ & \leftrightarrow \alpha_{n,k} | \alpha_{n,(-k)}, \Theta_n, D_n \end{aligned} \quad (2.4)$$

“Forward Filtering Backward Sampling – FFBS”

A amostragem de $\Theta_n | \alpha_n, D_n$ em (2.3) é realizada mais facilmente considerando o algoritmo denominado de “Forward Filtering, Backward Sampling – FFBS”, o qual é descrito a seguir.

Algoritmo FFBS

Este algoritmo é utilizado para amostrar o conjunto de vetores de estado $\Theta_n = \{\theta_0, \theta_1, \dots, \theta_n\}$ a partir da posteriori normal multivariada completa $p(\Theta_n | D_n)$. Pode-se escrever

$$p(\Theta_n | D_n) = p(\theta_n | D_n) p(\theta_{n-1} | \theta_n, D_{n-1}) \dots p(\theta_1 | \theta_2, D_1) p(\theta_0 | \theta_1, D_0).$$

Assim, amostra-se Θ_n pela simulação seqüencial dos vetores de estado individuais $\theta_n, \theta_{n-1}, \dots, \theta_0$ da seguinte maneira.

- 1 – Amostre θ_n de $\theta_n | D_n \sim N[\mathbf{m}_t, \mathbf{C}_t]$ e, então,
- 2 – para cada $t = n-1, n-2, \dots, 1, 0$, amostre θ_t de $p(\theta_t | \theta_{t+1}, D_t)$, onde o valor condicional de θ_{t+1} é o vetor que acabou de ser amostrado.

As distribuições condicionais do item 2 são obtidas segundo:

$$\theta_t | \theta_{t+1}, D_t \sim N[\mathbf{h}_t, \mathbf{H}_t]$$

$$\text{onde } \mathbf{h}_t = \mathbf{m}_t + \mathbf{B}_t(\theta_{t+1} - \mathbf{a}_{t+1}),$$

$$\mathbf{H}_t = \mathbf{C}_t - \mathbf{B}_t \mathbf{R}_{t+1} \mathbf{B}_t',$$

$$\mathbf{B}_t = \mathbf{C}_t \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{-1}.$$

Assim, o processo de amostrar Θ_n começa realizando a análise padrão de $t = 0$ até $t = n$, computando e salvando as quantidades m_t , C_t , a_t , R_t e B_t em cada estágio. No tempo $t = n$, um vetor θ_n é amostrado e volta-se seqüencialmente no tempo, computando os elementos h_t e H_t em cada passo e gerando um novo valor para θ_t .

2.7 – Modelos Autorregressivos na forma de MLD

Definem-se agora modelos autorregressivos na forma de modelos lineares dinâmicos. Considere a série Y_t descrita a seguir:

$$Y_t = \mu_t + X_t + v_t \quad (2.5)$$

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

Tem-se que o nível médio da série, μ_t , está na forma de um passeio aleatório e X_t representa a parte autorregressiva de ordem p .

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

$$Y_t = F' \theta_t + v_t$$

$$\theta_t = G \theta_{t-1} + \omega^* t$$

onde:

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

$F = (1, 1, 0, \dots, 0)'$,

$\omega^*_t = (\omega_t, \varepsilon_t, 0, \dots, 0)'$,

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \phi_1 & \phi_2 & \phi_3 & \dots & \phi_p \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}, W^* = \begin{bmatrix} W & 0 & 0 & \dots & 0 \\ 0 & U & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \text{ e } \omega^* \sim N(0, W^*).$$

Nota-se, aqui, que os parâmetros da componente autorregressiva ajudam a compor a matriz G . Os parâmetros do modelo (2.5) são definidos por:

$$\alpha_1 = \phi = \{\phi_1, \phi_2, \dots, \phi_p\}$$

$$\alpha_2 = \{V, W, U\},$$

$$\text{além de } \Theta_n = \{\theta_0, \theta_1, \dots, \theta_n\}$$

e para cada t , $Z_t = (X_t, X_{t+1}, \dots, X_{t+p+1})$.

O procedimento para a simulação a ser apresentado agora difere um pouco em relação ao algoritmo FFBS apresentado anteriormente, o qual se referia à situação de um modelo autorregressivo de 1ª ordem. Para os modelos de ordem igual ou superior a 2, são necessárias algumas modificações, como será descrito a seguir.

Como definido, junto com os parâmetros $\alpha = \{\phi, V, W, U\}$, deve-se simular os vetores de estado $\Theta_n | \alpha, D_n$.

Para amostrar $\Theta_n | \alpha, D_n$, utiliza-se o algoritmo FFBS descrito na seção 2.6. Quando se deseja calcular θ_t dado o conhecimento de θ_{t+1} , apenas o primeiro e último elementos de θ_t são desconhecidos, pois

$$\theta_{t+1} = (\mu_{t+1}, X_{t+1}, X_t, X_{t-1}, \dots, X_{t-p+2})$$

$$\theta_t = (\mu_t, X_t, X_{t-1}, \dots, X_{t-p+2}, X_{t-p+1})$$

o que faz com que seja necessário amostrar os elementos μ_t e X_{t-p+1} , uma vez que os outros elementos de θ_t já estão obtidos em θ_{t+1} . Para proceder esta amostragem, adota-se a seguinte estratégia:

1 – Computam-se os momentos $p(\mu_t, X_{t-p+1} | X_t, X_{t-1}, \dots, X_{t-p+2}, D_t)$, que nada mais são do que a distribuição bivariada condicional para o 1º e último elementos de θ_t a partir da conjunta completa $\theta_t | D_t \sim N[m, C]$.

2 – Computa-se $e_{t+1} = X_{t+1} - \sum_{j=1}^{p-1} \phi_j X_{t+1-j}$ com base nos elementos de θ_{t+1} .

Desta forma, tem-se agora um par de “observações” independentes

$$\mu_{t+1} \sim N[\mu_t, W] \quad \text{e} \quad e_{t+1} \sim N[\phi_p X_{t-p+1}, U]$$

nos dois parâmetros (μ_t, X_{t-p+1}) .

3 – Utiliza-se a verossimilhança correspondente para atualizar a priori já computada em (1), ou seja:

$$p(\mu_t, X_{t-p+1} | X_t, \dots, X_{t-p+2}, D_t) * \exp\left(-\frac{(\mu_{t+1} - \mu_t)^2}{2W}\right) * \exp\left(-\frac{(e_{t+1} - \phi_p X_{t-p+1})^2}{2U}\right)$$

que é a marginal bivariada apropriada para o 1º e último elementos de θ_t .

Repete-se este processo para $t = n-1, \dots, 0$, resultando em um conjunto completo de vetores de estado amostrados Θ_n que representa uma amostra de $p(\Theta_n | \alpha, D_n)$.

A amostragem da componente $p(\alpha_n | \Theta_n D_n)$ é separada em $\alpha_1 = \{\phi\}$ e $\alpha_2 = \{V, U, W\}$. Assume-se que $p(V, U, W | D_0)$ é independente, ou seja, igual a $p(V | D_0)p(U | D_0)p(W | D_0)$, assim, tem-se que os procedimentos para obtenção das distribuições condicionais completas para estes parâmetros são apresentados a seguir.

1 – Amostragem de $\phi | V, U, W, \Theta_n, D_n$. Condicional ao conhecimento de Θ_n , o qual fornece todos os valores do processo autorregressivo para $t = 0, \dots, n$, tem-se que a distribuição condicional completa para o vetor de parâmetros ϕ é proporcional a:

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

onde $Z_t = (X_t, X_{t-1}, \dots, X_{t-p+1})$, que nada mais é do que o vetor de estados sem a componente referente ao nível da série μ_t . A verossimilhança condicional possui a forma da normal em ϕ , e tem-se ainda que $p(Z_0 | \alpha, D_0)$ não depende de α , de forma que uma priori normal para $p(\phi | D_0)$ leva a uma posteriori condicional normal.

2 – Amostragem de $V | \phi, \Theta_n, D_n$. Com base nos valores de θ_t , computam-se os resíduos $v_t = Y_t - F'\theta$ para cada $t = 1, \dots, n$, e então amostra-se a posteriori proporcional a

$$p(V | D_0)V^{-n/2} \exp(-\sum_{t=1}^n v_t^2 / 2V),$$

de forma que uma priori Gama Inversa para V é conjugada, o que faz com que a posteriori também possua a forma de uma Gama Inversa.

3 – Amostragem de $W|\phi, \Theta_n, D_n$. Similar ao processo 2, calcula-se $\omega_t = \mu_t - \mu_{t-1}$ para cada t , e então, amostra-se a posteriori proporcional a

$$p(W | D_0)W^{-n/2} \exp\left(-\sum_{t=1}^n \omega_t^2/2W\right),$$

onde uma priori Gama inversa também leva a uma posteriori Gama Inversa.

4 – Amostragem de $U|\phi, \Theta_n, D_n$. Da mesma forma como apresentado em 2 e 3,

obtem-se $u_t = X_t - \sum_{j=1}^p \phi_j X_{t-j}$, e então amostra-se a posteriori condicional

proporcional a

$$p(U | D_0)U^{-n/2} \exp\left(-\sum_{t=1}^n u_t^2/2U\right),$$

onde uma priori Gama Inversa também leva a uma posteriori Gama Inversa.

3 – METODOLOGIA

Com base na teoria apresentada na seção 2, a performance de um modelo autorregressivo de 2ª ordem, AR(2), foi avaliada com relação às estimativas dos parâmetros, erros de previsão, convergência e previsões um passo à frente. Seja a observação Y_t representada como (2.5) e descrita para o caso de um AR(2), ou seja:

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

onde μ_t representa o nível médio da série, X_t o componente autorregressivo de ordem 2, e v_t uma perturbação aleatória.

O nível médio da série pode ser visto como $F'\theta_t = \mu_t = \mu_{t-1} + \omega_t$; onde $\omega_t \sim N[0, W]$, e o componente autorregressivo, como $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$, onde $\varepsilon_t \sim N[0, U]$ e $v_t \sim N[0, V]$.

Expressando Y_t na forma de MLD, tem-se:

$$Y_t = F'\theta_t + v_t$$

$$\theta_t = G\theta_{t-1} + \omega^*_t$$

onde:

$$\theta_t = (\mu_t, X_t, X_{t-1})'$$

$$F = (1, 1, 0)'$$

$$\omega^*_t = (\omega_t, \varepsilon_t, 0)'$$

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \phi_1 & \phi_2 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{W}^* = \begin{bmatrix} W & 0 & 0 \\ 0 & U & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ e } \omega^*_t \sim N(0, \mathbf{W}^*).$$

Este modelo serviu de base para a obtenção dos resultados. O procedimento para a simulação segue aquele descrito na seção 2.7, sendo que os parâmetros do modelo que devem ser estimados são $\alpha = \{\phi_1, \phi_2, V, W, U\}$ e Θ_n , onde $\Theta_n = (\theta_0, \theta_1, \dots, \theta_n)$.

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

Outro resultado a ser explorado foi a utilização das estimativas para se obterem previsões um passo à frente. Estas 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.

Desta forma, pode-se comparar estas duas séries com a série original. De maneira similar, pode-se comparar os erros de previsão para os dois casos e identificar a série que melhor se ajustou com base no erro quadrático médio.

Na aplicação com dados simulados, foram utilizadas duas cadeias em paralelo para avaliar a convergência e diferentes tamanhos de série, iguais a 200, 500 e 800 observações. As séries foram geradas segundo os valores $\phi_1 = 0,6$, $\phi_2 = -0,4$, $V = 0,5$, $W = 0,9$, $U = 0,9$. A distribuição atribuída para $\theta_0|D_0$ possui média e variância dadas por:

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

Com relação à aplicação com dados reais, a série Canadian lynx, Nicholls & Quin (1982), foi utilizada, com o amostrador sendo aplicado para fornecer as estimativas de ϕ_1 , ϕ_2 , V e U. A matriz W foi trabalhada a partir de fatores de descontos. A apresentação dos resultados começa com fator de desconto igual a 0,90, depois com valor igual a 0,95 e, finalmente, com 0,99.

A série possui 100 observações anuais no período de 1821 a 1920, sendo que o seu gráfico pode ser visto na FIGURA 1.

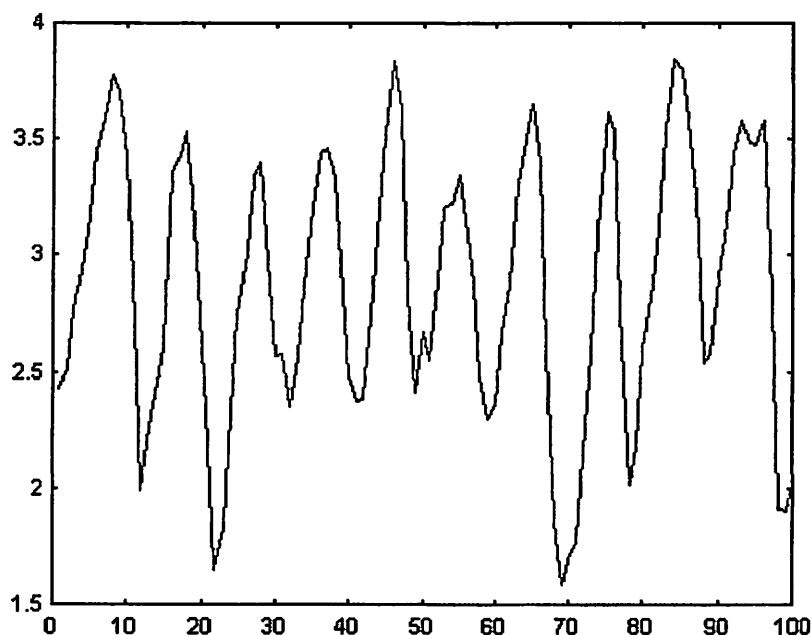


FIGURA 1 – Gráfico da série Canadian lynx.

Os resultados obtidos na aplicação com a série Canadian Lynx foram comparados com os obtidos na estimação a partir da inferência clássica.

Nas duas aplicações, os resultados foram obtidos por simulação, na qual o programa se estrutura da seguinte maneira: (1) definição dos valores dos

parâmetros e geração da série, (2) definição de duas cadeias de valores iniciais distintos para avaliar a convergência, (3) equações de atualização, (4) “amostragem para trás (FFBS)”, (5) amostragem do 2º bloco, (6) descarte das observações iniciais, (7) obtenção de valores alternados, (8) implementação do teste de Gelman e Rubin e (9) estatísticas descritivas e gráficos. Estes passos estão mais detalhados no decorrer dos resultados.

Na utilização do amostrador de Gibbs, trabalhou-se com 8000 iterações para cada uma das duas cadeias. Para reduzir o efeito destes valores iniciais, foram eliminadas as primeiras vinte por cento observações, o que reduziu para 6400 o número de observações para cada uma.

Dado que o amostrador é um processo Markoviano, tem-se que as amostras não são tomadas independentemente, e sim com uma correlação de 1ª ordem. Para suavizar este problema, retirou-se 1 observação a cada 25 de um total de 6400 observações de cada cadeia. Assim, esta nova série possui 256 observações, totalizando 512 ao analisar as duas cadeias.

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

É importante ressaltar que todas as quantidades a serem estimadas são aleatórias, sendo que três delas são variâncias. A convergência pode se tornar um pouco lenta e as estimativas perderem um pouco da precisão.

O software utilizado para estes 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

4 – RESULTADOS E DISCUSSÕES

Esta seção apresenta o desenvolvimento das distribuições necessárias para a estimação dos parâmetros utilizando o amostrador de Gibbs. Sabe-se que o amostrador é realizado considerando os blocos $\Theta_n|\alpha, D_n$ e $\alpha_n|\Theta_n, D_n$. Assim, em um primeiro momento é amostrado $\Theta_n|\alpha, D_n$, para só então amostrar $\alpha_n|\Theta_n, D_n$.

Após a definição das distribuições necessárias, apresenta-se a comparação dos resultados a partir de dados simulados (seção 4.2) e a partir de dados reais (seção 4.3).

4.1 – Distribuições condicionais completas

Começando pelo bloco $\Theta_n|\alpha, D_n$, e seguindo os passos do algoritmo FFBS descrito anteriormente, deve-se proceder com a filtragem começando no tempo 0 e caminhando até o tempo n. No tempo n é possível obter uma amostra do vetor $\theta_n|D_n \sim N(m_n, C_n)$.

Com base neste vetor amostrado, o interesse se concentra na obtenção do vetor $\theta_t|\theta_{t+1}, D_t$, para $t = n-1, \dots, 1, 0$. Para isto, necessita-se da distribuição do primeiro e último valores de θ_t . O primeiro passo é computar os momentos destes 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 $\theta_t|D_t \sim N[m_t, C_t]$, onde:

$$\mathbf{m}_t = \begin{bmatrix} \bar{\mu}_t \\ \bar{X}_t \\ \bar{X}_{t-1} \end{bmatrix}, \text{ e } \mathbf{C}_t = \begin{bmatrix} \text{var}(\mu_t) & \text{cov}(\mu_t, X_t) & \text{cov}(\mu_t, X_{t-1}) \\ \text{cov}(X_t, \mu_t) & \text{var}(X_t) & \text{cov}(X_t, X_{t-1}) \\ \text{cov}(X_{t-1}, \mu_t) & \text{cov}(X_{t-1}, X_t) & \text{var}(X_{t-1}) \end{bmatrix}.$$

A partir do de fato que $\theta_{t+1} = (\mu_{t+1}, X_{t+1}, X_t)$, e de que

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

tem-se que 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. Para amostrar estes elementos de θ_t , defina 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 \mathbf{C}_t .

Observe que Σ_{11} representa a matriz de variâncias e covariâncias para o primeiro e último valores do vetor θ_t , que é o que se deseja amostrar.

Com base nestas matrizes, pode-se reorganizar o vetor θ_t de forma que o primeiro e segundo elementos sejam, respectivamente, μ_t e X_{t-1} , ou seja:

$$\theta_{t+1} | \theta_{t+2}, \mathbf{D}_{t+1} \sim N(\mathbf{m}'_t, \mathbf{C}'_t), \text{ para}$$

$$\mathbf{m}'_t = \begin{bmatrix} \bar{\mu}_t \\ \bar{X}_{t-1} \\ \bar{X}_t \end{bmatrix}, \text{ e } \mathbf{C}'_t = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Definindo θ_t^- , lê-se “teta t menos”, como o vetor $(\mu, X_{t-1})'$, o qual deve ser amostrado, tem-se:

$$\theta_t^- | X_t, D_t \sim N_2(\mu_2, \Sigma_2),$$

onde

$$\mu_2 = \begin{bmatrix} \bar{\mu}_t \\ \bar{X}_{t-1} \end{bmatrix} + \Sigma_{12} \Sigma_{22}^{-1} (X_t - \bar{X}_t), \quad (3.1)$$

$$\Sigma_2 = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}. \quad (3.2)$$

A obtenção de μ_2 e Σ_{11} se baseia na propriedade da distribuição normal multivariada (Anexo A), a qual afirma que qualquer subconjunto de valores de um vetor com distribuição normal também possui distribuição normal com média e variância obtidas de maneira similar a (3.1) e (3.2).

Desta forma, tem-se que

$$p(\theta_t^- | D_t, X_t) \propto |\Sigma_2|^{-1/2} \exp\left\{-\frac{1}{2} \left[(\theta_t^- - \mu_2)' \Sigma_2^{-1} (\theta_t^- - \mu_2) \right]\right\}. \quad (3.3)$$

Fazendo $e_{t+1} = X_{t+1} - \phi_1 X_t$,

$$\text{tem-se } \mu_{t+1} \sim N(\mu, W) \text{ e } e_{t+1} \sim N(\phi_2 X_{t-1}, U). \quad (3.4)$$

A primeira parte de (3.4) vem do fato de que $\mu_{t+1} = \mu_t + \omega_t$, com $\omega_t \sim N(0, W)$. Para o segundo termo, tem-se que $X_{t+1} = \phi_1 X_t + \phi_2 X_{t-1} + \varepsilon_{t+1}$, 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 nos fornece $e_{t+1} \sim N(\phi_2 X_{t-1}, U)$.

Assim, unindo as expressões (3.3) e (3.4), tem-se que:

$$p(\theta_t^- | \theta_{t+1}, D_n) \propto \exp\left\{-\frac{1}{2}[(\theta_t^- - \mu_2)' \Sigma_2^{-1}(\theta_t^- - \mu_2)]\right\} x \exp\left\{-\frac{1}{2}\left[\frac{(\mu_{t+1} - \mu_t)^2}{W} + \frac{(X_{t+1} - \phi_1 X_t - \phi_2 X_{t-1})^2}{U}\right]\right\}. \quad (3.5)$$

O segundo termo da expressão (3.5) pode ser reescrito como:

$$\begin{aligned} & \left[\frac{(\mu_{t+1} - \mu_t)^2}{W} + \frac{(X_{t+1} - \phi_1 X_t - \phi_2 X_{t-1})^2}{U} \right] = \\ & = \frac{(\mu_t - \mu_{t+1})^2}{W} + \frac{(X_{t-1} - (X_{t+1} - \phi_1 X_t) \phi_2^{-1})^2 \phi_2^2}{U} = \\ & = (\mu_t - \mu_{t+1}) W^{-1} (\mu_t - \mu_{t+1}) + \\ & \quad + (X_{t-1} - (X_{t+1} - \phi_1 X_t) \phi_2^{-1}) \phi_2^2 U^{-1} (X_{t-1} - (X_{t+1} - \phi_1 X_t) \phi_2^{-1}). \end{aligned}$$

Escrevendo na forma matricial, tem-se:

$$\left[\begin{pmatrix} \mu_t \\ X_{t-1} \end{pmatrix} - \begin{pmatrix} \mu_{t+1} \\ (X_{t+1} - \phi_1 X_t) \phi_2^{-1} \end{pmatrix} \right]' \begin{bmatrix} W & 0 \\ 0 & U \phi_2^{-2} \end{bmatrix}^{-1} \left[\begin{pmatrix} \mu_t \\ X_{t-1} \end{pmatrix} - \begin{pmatrix} \mu_{t+1} \\ (X_{t+1} - \phi_1 X_t) \phi_2^{-1} \end{pmatrix} \right]$$

Definindo

$$\mu_{\theta_i^-} = \left(\begin{array}{c} \mu_{t+1} \\ (X_{t+1} - \phi_1 X_t - \phi_2 X_{t-1}) \phi_2^{-1} \end{array} \right) \quad \text{e} \quad \Sigma_{\theta_i^-} = \begin{bmatrix} W & 0 \\ 0 & U \phi_2^{-2} \end{bmatrix}, \text{ tem-se:}$$

$$\frac{(\mu_{t+1} - \mu_t)^2}{W} + \frac{(X_{t+1} - \phi_1 X_t - \phi_2 X_{t-1})^2}{U} = (\theta_t^- - \mu_{\theta_i^-})' \Sigma_{\theta_i^-}^{-1} (\theta_t^- - \mu_{\theta_i^-}).$$

Voltando em (3.5), tem-se agora:

$$\begin{aligned} p(\theta_t^- | \theta_{t+1}, D_n) &\propto \exp \left\{ -\frac{1}{2} \left[(\theta_t^- - \mu_2)' \Sigma_2^{-1} (\theta_t^- - \mu_2) \right] \right\} x \\ &x \cdot \exp \left\{ -\frac{1}{2} \left[(\theta_t^- - \mu_{\theta_i^-})' \Sigma_{\theta_i^-}^{-1} (\theta_t^- - \mu_{\theta_i^-}) \right] \right\}. \end{aligned} \quad (3.6)$$

Reescrevendo (3.6), tem-se:

$$\begin{aligned} p(\theta_t^- | \theta_{t+1}, D_n) \\ \propto \exp \left\{ -\frac{1}{2} \left[(\theta_t^- - \mu_2)' \Sigma_2^{-1} (\theta_t^- - \mu_2) + (\theta_t^- - \mu_{\theta_i^-})' \Sigma_{\theta_i^-}^{-1} (\theta_t^- - \mu_{\theta_i^-}) \right] \right\}. \end{aligned} \quad (3.7)$$

Considerando a seguinte componente de (3.7),

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

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

$$\begin{aligned} & \theta_t^- \Sigma_2^{-1} \theta_t^- - \theta_t^- \Sigma_2^{-1} \mu_2 - \mu_2 \Sigma_2^{-1} \theta_t^- + \mu_2 \Sigma_2^{-1} \mu_2 + \\ & + \theta_t^- \Sigma_{\theta_t^-}^{-1} \theta_t^- - \theta_t^- \Sigma_{\theta_t^-}^{-1} \mu_{\theta_t^-} - \mu_{\theta_t^-} \Sigma_{\theta_t^-}^{-1} \theta_t^- + \mu_{\theta_t^-} \Sigma_{\theta_t^-}^{-1} \mu_{\theta_t^-}. \end{aligned} \quad (3.8)$$

Esta ainda pode ser reescrita como:

$$\begin{aligned} & = \theta_t^- (\Sigma_2^{-1} + \Sigma_{\theta_t^-}^{-1}) \theta_t^- - \theta_t^- (\Sigma_2^{-1} \mu_2 + \Sigma_{\theta_t^-}^{-1} \mu_{\theta_t^-}) - \\ & - (\mu_2 \Sigma_2^{-1} + \mu_{\theta_t^-} \Sigma_{\theta_t^-}^{-1}) \theta_t^- + (\mu_2 \Sigma_2^{-1} \mu_2 + \mu_{\theta_t^-} \Sigma_{\theta_t^-}^{-1} \mu_{\theta_t^-}). \end{aligned} \quad (3.9)$$

A partir da equação (3.9), pode-se obter uma representação de θ_t^- na forma de uma distribuição normal bivariada. Completando quadrados, tem-se:

$$\begin{aligned} & \left[\theta_t^- - (\Sigma_2^{-1} + \Sigma_{\theta_t^-}^{-1})^{-1} (\Sigma_2^{-1} \mu_2 + \Sigma_{\theta_t^-}^{-1} \mu_{\theta_t^-}) \right] (\Sigma_2^{-1} + \Sigma_{\theta_t^-}^{-1}) \left[\theta_t^- - (\Sigma_2^{-1} + \Sigma_{\theta_t^-}^{-1})^{-1} (\Sigma_2^{-1} \mu_2 + \Sigma_{\theta_t^-}^{-1} \mu_{\theta_t^-}) \right] \\ & - (\mu_2 \Sigma_2^{-1} + \mu_{\theta_t^-} \Sigma_{\theta_t^-}^{-1}) (\Sigma_2^{-1} + \Sigma_{\theta_t^-}^{-1})^{-1} (\Sigma_2^{-1} \mu_2 + \Sigma_{\theta_t^-}^{-1} \mu_{\theta_t^-}) + \\ & + (\mu_2 \Sigma_2^{-1} \mu_2 + \mu_{\theta_t^-} \Sigma_{\theta_t^-}^{-1} \mu_{\theta_t^-}) \end{aligned} \quad (3.10)$$

Como as duas últimas componentes de (3.10) não dependem de θ_t^- , tem-se

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

$$\begin{aligned} & \propto \exp \left\{ -\frac{1}{2} \left[\theta_t^- - (\Sigma_2^{-1} + \Sigma_{\theta_t^-}^{-1})^{-1} (\Sigma_2^{-1} \mu_2 + \Sigma_{\theta_t^-}^{-1} \mu_{\theta_t^-}) \right] (\Sigma_2^{-1} + \Sigma_{\theta_t^-}^{-1}) x \right. \\ & \left. x \left[\theta_t^- - (\Sigma_2^{-1} + \Sigma_{\theta_t^-}^{-1})^{-1} (\Sigma_2^{-1} \mu_2 + \Sigma_{\theta_t^-}^{-1} \mu_{\theta_t^-}) \right] \right\}, \end{aligned}$$

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

Com base na forma obtida acima, chega-se a:

$$\theta_t^- | \theta_{t+1}, D_n \sim N(\boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f), \quad (3.11)$$

$$\text{onde } \boldsymbol{\mu}_f = \left(\boldsymbol{\Sigma}_2^{-1} + \boldsymbol{\Sigma}_{\theta_t^-}^{-1} \right)^{-1} \left(\boldsymbol{\Sigma}_2^{-1} \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{\theta_t^-}^{-1} \boldsymbol{\mu}_{\theta_t^-} \right),$$

$$\text{e } \boldsymbol{\Sigma}_f = \left(\boldsymbol{\Sigma}_2^{-1} + \boldsymbol{\Sigma}_{\theta_t^-}^{-1} \right)^{-1}.$$

Desta forma, amostra-se um valor para θ_t^- a partir de (3.11), e então completa-se com o valor de X_t já amostrado em θ_{t+1} . Este 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 pode-se agora utilizar as informações obtidas – conjunto de todos os vetores θ – para proceder a 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 $\phi|V, U, W, \Theta_n, D_n$: Com base na seção 2.7, tem-se que a distribuição condicional completa para o vetor de parâmetros ϕ é proporcional a

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

Serão consideradas, para a análise, as prioris $\phi|D_0 \sim$ uniforme e $Z_0|\alpha_0, D_0 \sim$ constante. Desta forma, como o primeiro e segundo termos são constantes, tem-se:

$$\begin{aligned} & \prod_{t=3}^n \exp(-(X_t - Z'_{t-1} \phi)^2 / 2U) = \\ & = \exp \left\{ -\frac{1}{2U} \sum_{t=3}^n \left[\begin{pmatrix} X_{t-1} & X_{t-2} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \right] \left[\begin{pmatrix} X_{t-1} & X_{t-2} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \right] - 2X_t \begin{pmatrix} X_{t-1} & X_{t-2} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} + X_t^2 \right\} \\ & = \exp \left\{ -\frac{1}{2U} \sum_{t=3}^n \left[\begin{pmatrix} \phi_1 & \phi_2 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ X_{t-2} \end{pmatrix} \begin{pmatrix} X_{t-1} & X_{t-2} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} - 2 \begin{pmatrix} X_t X_{t-1} & X_t X_{t-2} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} + X_t^2 \right] \right\} \end{aligned}$$

Aplicando o somatório, obtém-se:

$$\exp \left\{ -\frac{1}{2U} \begin{pmatrix} \phi_1 & \phi_2 \end{pmatrix} \begin{bmatrix} \sum X_{t-1}^2 & \sum X_{t-1} X_{t-2} \\ \sum X_{t-2} X_{t-1} & \sum X_{t-2}^2 \end{bmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} - 2 \begin{pmatrix} \sum X_t X_{t-1} & \sum X_t X_{t-2} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} + \sum X_t^2 \right\}$$

onde \sum significa $\sum_{t=3}^n$.

Definido $\mathbf{A} = \begin{bmatrix} \Sigma \mathbf{X}_{t-1}^2 & \Sigma \mathbf{X}_{t-1} \mathbf{X}_{t-2} \\ \Sigma \mathbf{X}_{t-2} \mathbf{X}_{t-1} & \Sigma \mathbf{X}_{t-2}^2 \end{bmatrix}$, e $\mathbf{B} = \begin{bmatrix} \Sigma \mathbf{X}_t \mathbf{X}_{t-1} \\ \Sigma \mathbf{X}_t \mathbf{X}_{t-2} \end{bmatrix}$, tem-se:

$$\exp \left\{ -\frac{1}{2U} \left[\begin{pmatrix} \phi_1 & \phi_2 \end{pmatrix} \mathbf{A} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} - 2 \mathbf{B} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} + \Sigma \mathbf{X}_t^2 \right] \right\}.$$

Portanto, completando quadrados, tem-se:

$$\begin{aligned} p(\phi | V, U, W, \Theta_n, D_n) &= \\ &= \exp \left\{ -\frac{1}{2U} (\phi - \mathbf{A}^{-1} \mathbf{B})' \mathbf{A} (\phi - \mathbf{A}^{-1} \mathbf{B}) - \mathbf{B}' \mathbf{A}^{-1} \mathbf{B} + \sum \mathbf{X}_t^2 \right\} \\ &\propto \left\{ -\frac{1}{2U} (\phi - \mathbf{A}^{-1} \mathbf{B})' \mathbf{A} (\phi - \mathbf{A}^{-1} \mathbf{B}) \right\}, \end{aligned}$$

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

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

Tem-se que a distribuição condicional completa para o parâmetro V é proporcional a:

$$p(V | D_0) V^{-n/2} \exp\left(-\sum_{t=1}^n v_t^2 / 2V\right) \quad (3.12)$$

Considera-se uma priori GI (α, β) para $V | D_0$, ou seja,

$$p(V | D_0) \propto V^{-(\alpha+1)} \exp\left(-\frac{\beta}{V}\right).$$

Desta forma,

$$\begin{aligned}
 p(V | \phi, W, U, \Theta_n, D_n) &\propto V^{-(\alpha+1)} \exp\left(-\frac{\beta}{V}\right) \cdot V^{-\frac{n}{2}} \exp\left(-\sum_{t=1}^n \frac{v_t^2}{2 \cdot V}\right) \\
 &\propto V^{-\left(\frac{n+2\alpha}{2}+1\right)} \cdot \exp\left\{-\frac{1}{V} \left(\frac{1}{2} \left(\sum_{t=1}^n (Y_t - \mathbf{F}'\theta_t)^2 + 2\beta\right)\right)\right\}, \quad (3.13)
 \end{aligned}$$

uma vez que $v_t = Y_t - \mathbf{F}'\theta_t$.

Desta forma, obtém-se a distribuição para a condicional completa $V | \phi, W, U, \Theta_n, D_n$, como uma distribuição Gama Inversa (Anexo A), ou seja,

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

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

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

Considerando $\omega_t = \mu_t - \mu_{t-1}$, tem-se que a distribuição desejada é proporcional

a:

$$p(W | D_n) W^{-n/2} \exp\left(-\sum_{t=1}^n \omega_t^2 / 2W\right) \quad (3.14)$$

Considera-se uma priori GI (α, β) para $\omega|D_0$, ou seja,

$$p(W|D_0) \propto W^{-(\alpha+1)} \exp\left(-\frac{\beta}{W}\right).$$

Desta forma,

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

$$W^{-(\alpha+1)} \exp\left(-\frac{\beta}{W}\right) \cdot W^{-\frac{n}{2}} \exp\left\{-\frac{1}{W} \left(\frac{1}{2} \sum_{t=1}^n (\mu_t - \mu_{t-1})^2\right)\right\},$$

e então:

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

$$W^{-\left(\frac{n+2\alpha}{2}+1\right)} \cdot \exp\left\{-\frac{1}{W} \left(\frac{1}{2} \left(\sum_{t=1}^n (\mu_t - \mu_{t-1})^2 + 2\beta\right)\right)\right\} \quad (3.15)$$

Dado que a forma de (3.15) se assemelha à forma da distribuição Gama Inversa, podem-se definir os parâmetros da mesma forma como realizado para a variável V.

Desta forma, obtém-se a distribuição para a condicional completa $W | \phi, V, U, \Theta_n, D_n$, como uma Gama Inversa, ou seja,

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

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

Considerando $u_t = X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2}$ tem-se que a distribuição desejada é proporcional a:

$$p(U | D_0) U^{-n/2} \exp\left(-\sum_{t=1}^n u_t^2 / 2U\right), \quad (3.16)$$

Considera-se uma priori GI (α, β) para $u|D_0$, ou seja,

$$p(U|D_0) \propto U^{-(\alpha+1)} \exp\left(-\frac{\beta}{U}\right).$$

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

$$U^{-(\alpha+1)} \exp\left(-\frac{\beta}{U}\right) U^{-\frac{n}{2}} \exp\left\{-\frac{1}{U} \left(\frac{1}{2} \sum_{t=3}^n (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2})^2\right)\right\}$$

e então:

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

$$U^{-\left(\frac{n+2\alpha}{2}+1\right)} \cdot \exp\left\{-\frac{1}{U} \left(\frac{1}{2} \left(\sum_{t=3}^n (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2})^2 + 2\beta\right)\right)\right\} \quad (3.17)$$

Desta forma, obtém-se a distribuição para a condicional completa $U | \phi, V, W, \Theta_n, D_n$ como uma Gama Inversa, como pode ser visto abaixo:

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

4.2 – Aplicação com dados simulados

Segue-se agora com a utilização das distribuições definidas na seção 4.1 para séries geradas seguindo um processo AR(2) e com tamanhos iguais a 200, 500 e 800 observações.

Análise considerando uma série com 200 observações

A FIGURA 2 apresenta o gráfico da série:

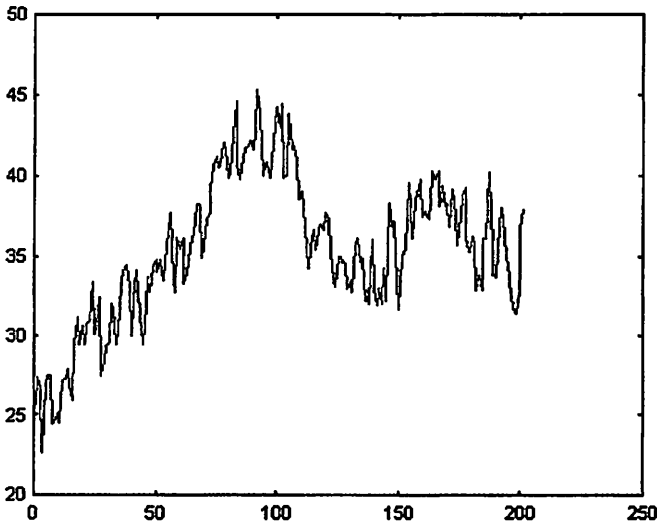


FIGURA 2: Gráfico da série simulada com 200 observações com base em um modelo AR(2).

A FIGURA 3 apresenta o gráfico e o histograma para os 12800 valores obtidos para cada um dos parâmetros (com as 20% primeiras observações já eliminadas). Os gráficos estão na seguinte ordem: Série original, ϕ_1 , ϕ_2 , V, W e U.

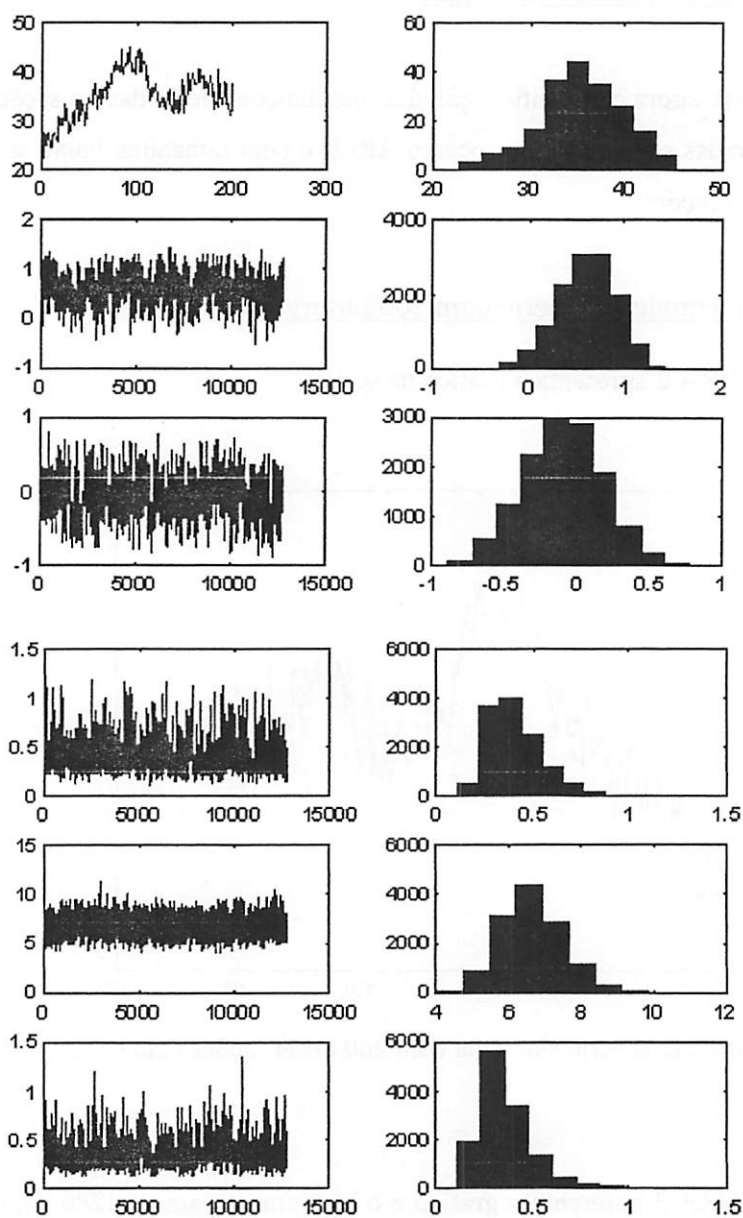


FIGURA 3: Gráfico e histograma dos valores da série gerada e dos parâmetros ϕ_1 , ϕ_2 , V , W e U , respectivamente, considerando uma série com 200 observações.

Pela FIGURA 3 nota-se que a série para todos os parâmetros parece convergir. Os histogramas para os parâmetros ϕ_1 , ϕ_2 e W sugerem a forma de uma distribuição normal, ao passo, que, para V e U, se assemelham à de uma Gama Inversa.

A análise descritiva dos parâmetros, considerando as 512 observações, pode ser observada na TABELA 1.

TABELA 1 – Resultado da inferência sobre os parâmetros, considerando uma série com 200 observações.

Parâmetros	Valor Real	Média	D.P	R	I.C (95%)
ϕ_1	0,6	0,6028	0,2825	1,0043	(0,0491 ; 1,1565)
ϕ_2	-0,4	-0,0871	0,2656	1,0234	(-0,6077 ; 0,4335)
V	0,5	0,4111	0,1463	1,0023	(0,1244 ; 0,6978)
W	0,9	6,5660	0,8267	0,9982	(4,9457 ; 8,1863)
U	0,9	0,3547	0,1294	1,0034	(0,1011 ; 0,6083)

Com base na TABELA 1, vê-se que apenas as estimativas para ϕ_1 e para V se aproximam dos valores dos parâmetros. A estimativa de ϕ_2 não se mostrou significativa, e a de W apresentou valor bem maior do que o real. O valor da estatística R atingiu valores desejados, afirmando a convergência de cada um dos cinco parâmetros.

Na FIGURA 4 são apresentados os gráficos considerando as 512 observações utilizadas para obter as estimativas.

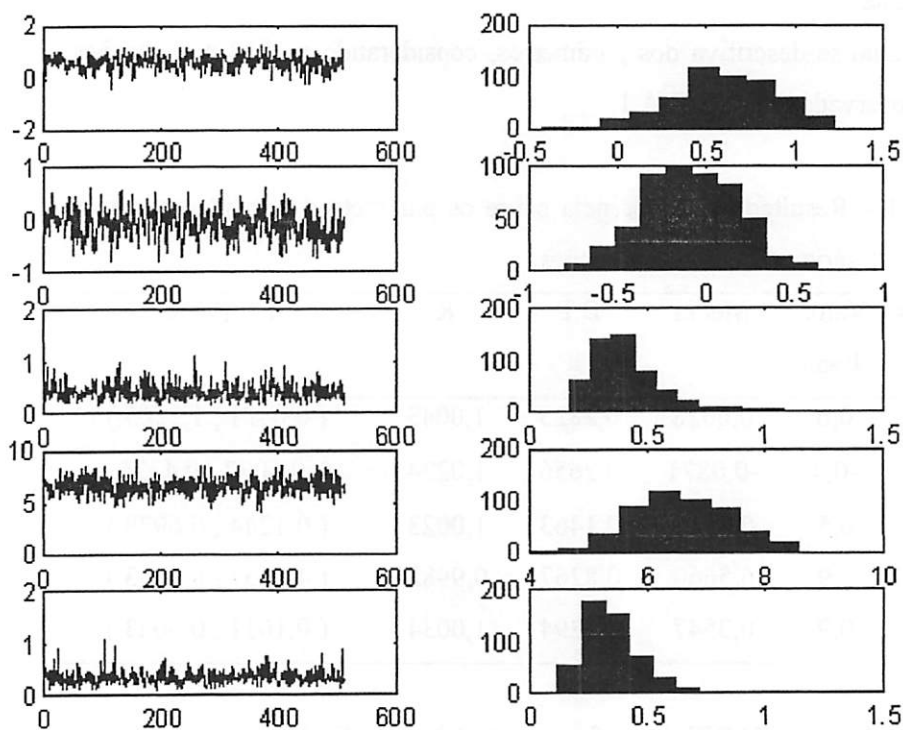


FIGURA 4: Gráfico e histograma da série utilizadas nas estimativas de ϕ_1 e ϕ_2 , V, W e U, respectivamente, considerando uma série com 200 observações.

Com base nas estimativas obtidas, a comparação das séries obtidas a partir de previsões um passo à frente pode ser vista na FIGURA 5.

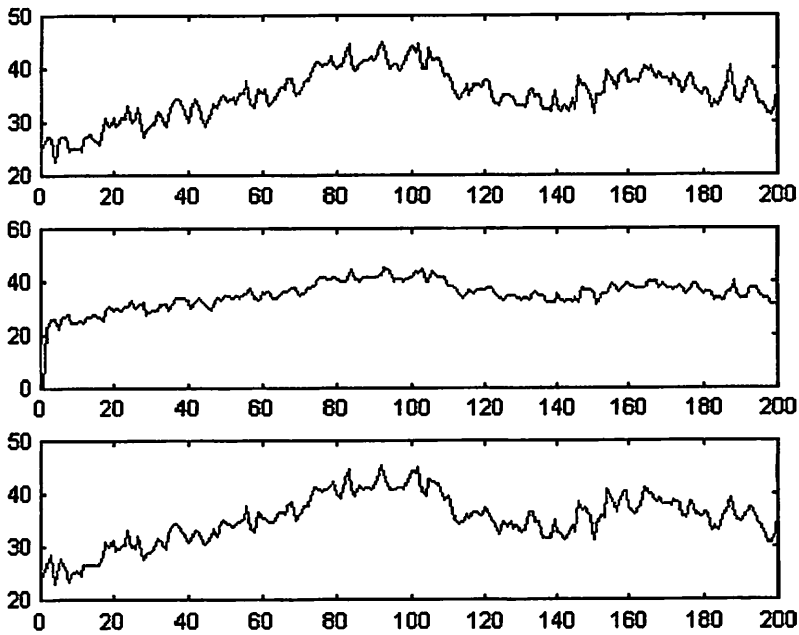


FIGURA 5: Gráfico da série original, atualizada e “amostrada para trás”, respectivamente, com base nas estimativas obtidas.

Pela FIGURA acima, nota-se um melhor ajuste em relação à série real para a série “prevista” a partir da “amostragem para trás”. As estimativas podem ser vistas na TABELA 2.

TABELA 2 – Estimativa dos erros de previsão um passo à frente para as duas séries “previstas”.

	Estimado	DP	EQM
Série atualizada	0,2119	2,6074	6,8434
Série “amostrada para trás”	-0,0280	0,6109	0,3721

Pela TABELA 2, nota-se uma grande redução no erro de previsão quando se utilizam os vetores de estados a partir da “amostragem para trás”. O gráfico dos erros de previsão para as séries pode ser visto na FIGURA 6

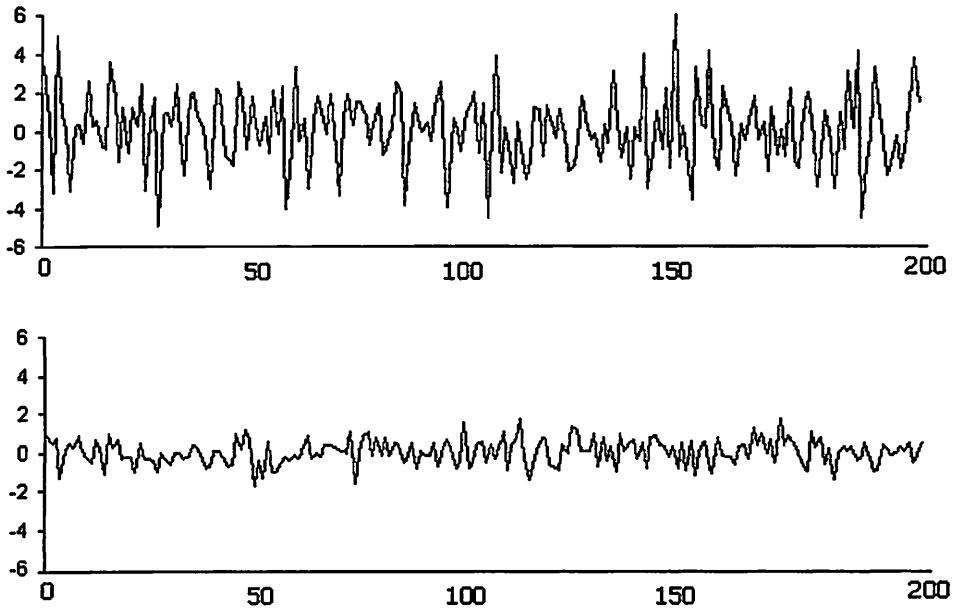


FIGURA 6 – Erros de previsão considerando a série atualizada e “amostrada para trás”, respectivamente.

Para melhor avaliar o comportamento do modelo, repetiram-se os mesmos procedimentos, para os mesmos valores iniciais, número de iterações, mas com tamanho da série alterado.

Análise considerando uma série com 500 observações

A FIGURA 7 apresenta o gráfico da série:

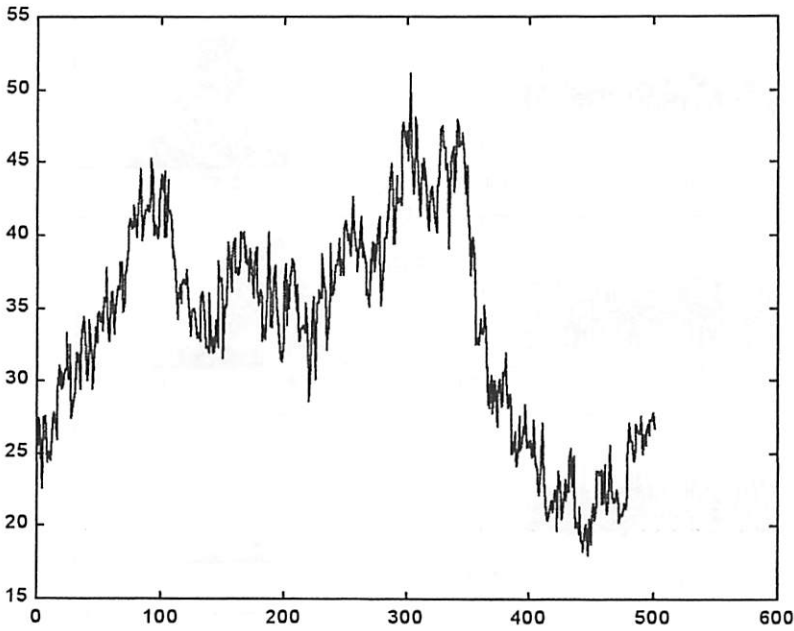


FIGURA 7: Gráfico da série simulada com 500 observações com base em um modelo AR(2).

A FIGURA 8 apresenta o gráfico e o histograma para os 12800 valores obtidos para cada um dos parâmetros (com as 20% primeiras observações já eliminadas). Os gráficos estão na seguinte ordem: Série original, ϕ_1 , ϕ_2 , V, W e U.

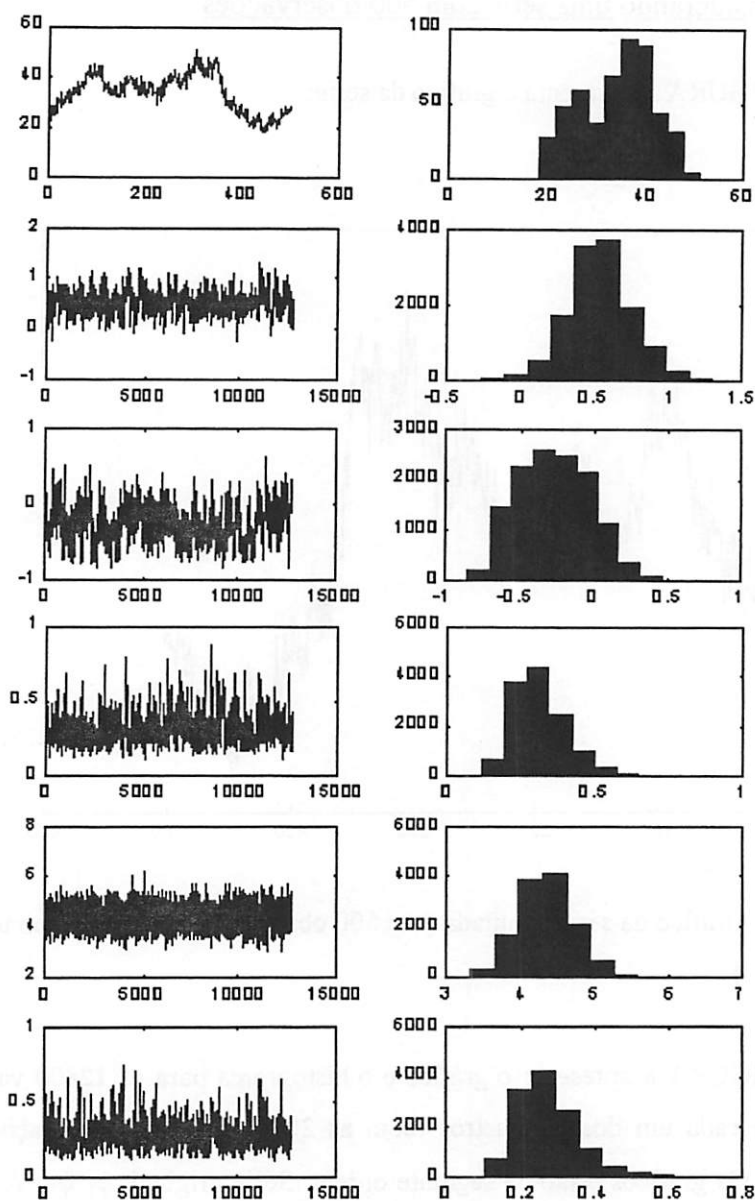


FIGURA 8: Gráfico e histograma dos valores da série gerada e dos parâmetros ϕ_1 , ϕ_2 , V , W e U , respectivamente, considerando uma série com 500 observações.

Pela FIGURA 8, nota-se que a série parece convergir para os cinco parâmetros avaliados. Os histogramas para os parâmetros ϕ_1 , ϕ_2 e W apresentam a forma de uma distribuição normal, ao passo que os obtidos para V e U se assemelham à de uma Gama Inversa.

A análise descritiva dos parâmetros, considerando as 512 observações, pode ser observada na TABELA 3

TABELA 3 – Resultado da inferência sobre os parâmetros, considerando uma série com 500 observações.

Parâmetros	Valor Real	Média	D.P	R	I.C (95%)
ϕ_1	0,6	0,5295	0,2060	0,9990	(0,1257 ; 0,9333)
ϕ_2	-0,4	-0,2636	0,2415	1,0045	(-0,7369 ; 0,2097)
V	0,5	0,3109	0,0920	1,0126	(0,1306 ; 0,4912)
W	0,9	4,3794	0,3583	1,0235	(3,6771 ; 5,0817)
U	0,9	0,2767	0,0844	0,9995	(0,1113 ; 0,4421)

Com base na TABELA 3, vê-se que, como na situação para série com 500 observações, apenas as estimativas para ϕ_1 e para V se aproximam dos valores dos parâmetros. A estimativa de ϕ_2 não se mostrou significativa, e a de W continua apresentando valor bem maior do que o real. O valor da estatística R atingiu valores desejados (um pouco alto para W), afirmando convergência de cada um dos cinco parâmetros.

Na FIGURA 9 são apresentados os gráficos considerando as 512 observações utilizadas para se obter as estimativas.

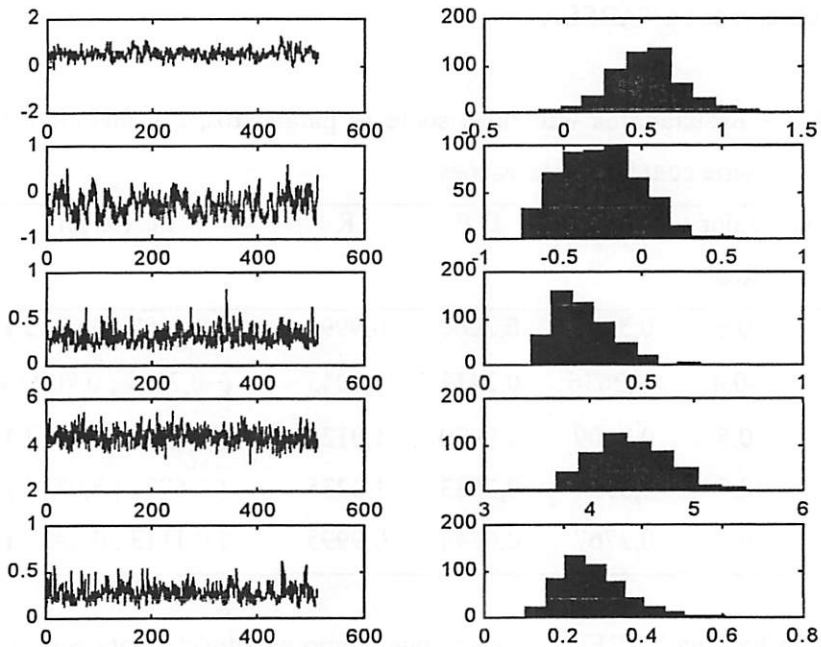


FIGURA 9: Gráfico e histograma da série utilizadas nas estimativas de ϕ_1 e ϕ_2 , V , W e U , respectivamente, considerando uma série com 500 observações.

Com base nas estimativas obtidas, a comparação das séries obtidas a partir de previsões um passo à frente pode ser vista na FIGURA 10.

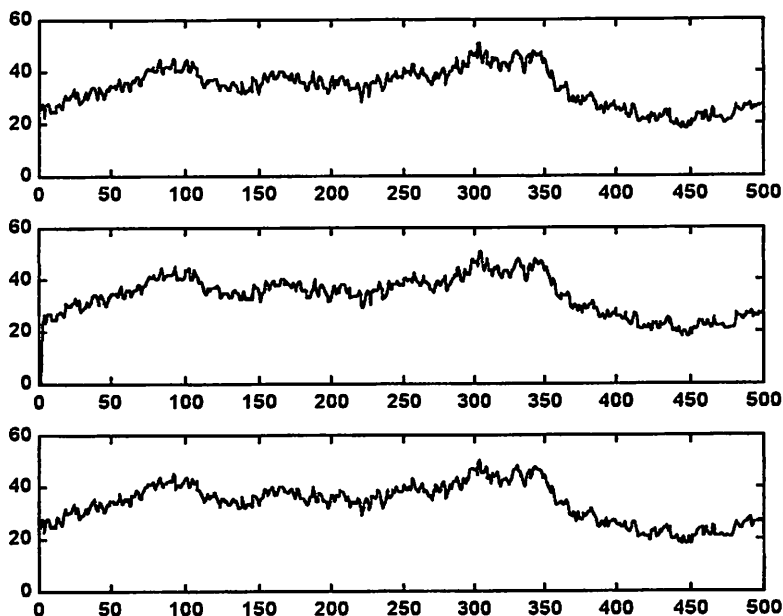


FIGURA 10: : Gráfico da série original, atualizada e “amostrada para trás”, respectivamente, com base nas estimativas obtidas.

Pela FIGURA 10, nota-se um melhor ajuste para a série “prevista” a partir dos vetores de estado obtidos na “amostragem para trás”. As estimativas podem ser vistas na TABELA 4.

TABELA 4 – Estimativa dos erros de previsão um passo à frente para as duas séries “previstas”.

	Estimativa	DP	EQM
Série atualizada	0,0625	2,2023	4,8541
Série “amostrada para trás”	-0,0093	0,5564	0,3090

Pela TABELA 4, nota-se um menor erro de previsão quando se utilizam os vetores de estados a partir da “amostragem para trás”. A diferença entre os EQM para as duas séries continua grande. O gráfico dos erros de previsão para as séries pode ser visto na FIGURA 11.

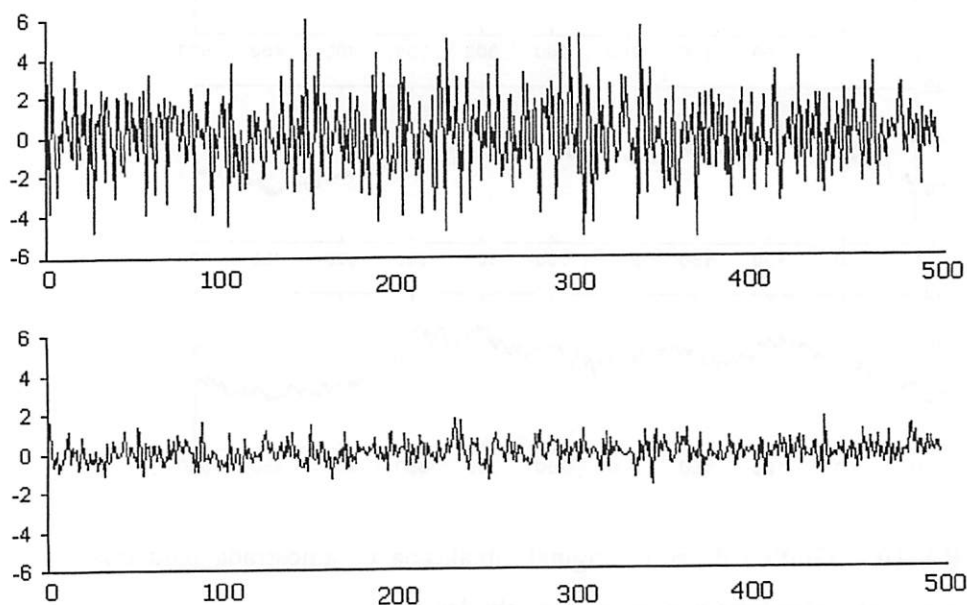


FIGURA 11 – Erros de previsão considerando a série atualizada e “amostrada para trás”, respectivamente.

Análise considerando uma série com 800 observações

A FIGURA 12 apresenta o gráfico da série:

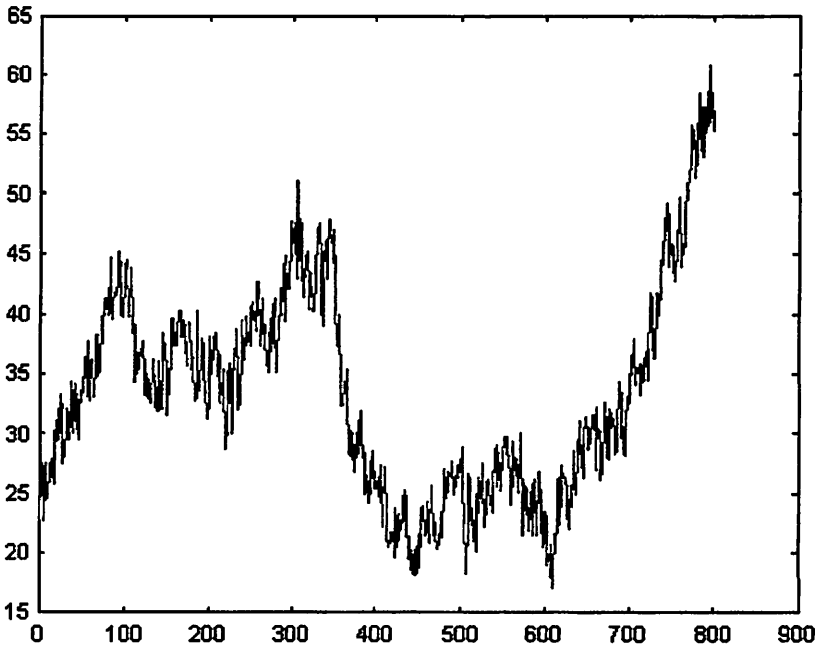


FIGURA 12: Gráfico da série simulada com 800 observações com base em um modelo AR(2).

A FIGURA 13 apresenta o gráfico e o histograma para os 12800 valores obtidos para cada um dos parâmetros (com as 20% primeiras observações já eliminadas). Os gráficos estão na seguinte ordem: Série original, ϕ_1 , ϕ_2 , V, W e U.

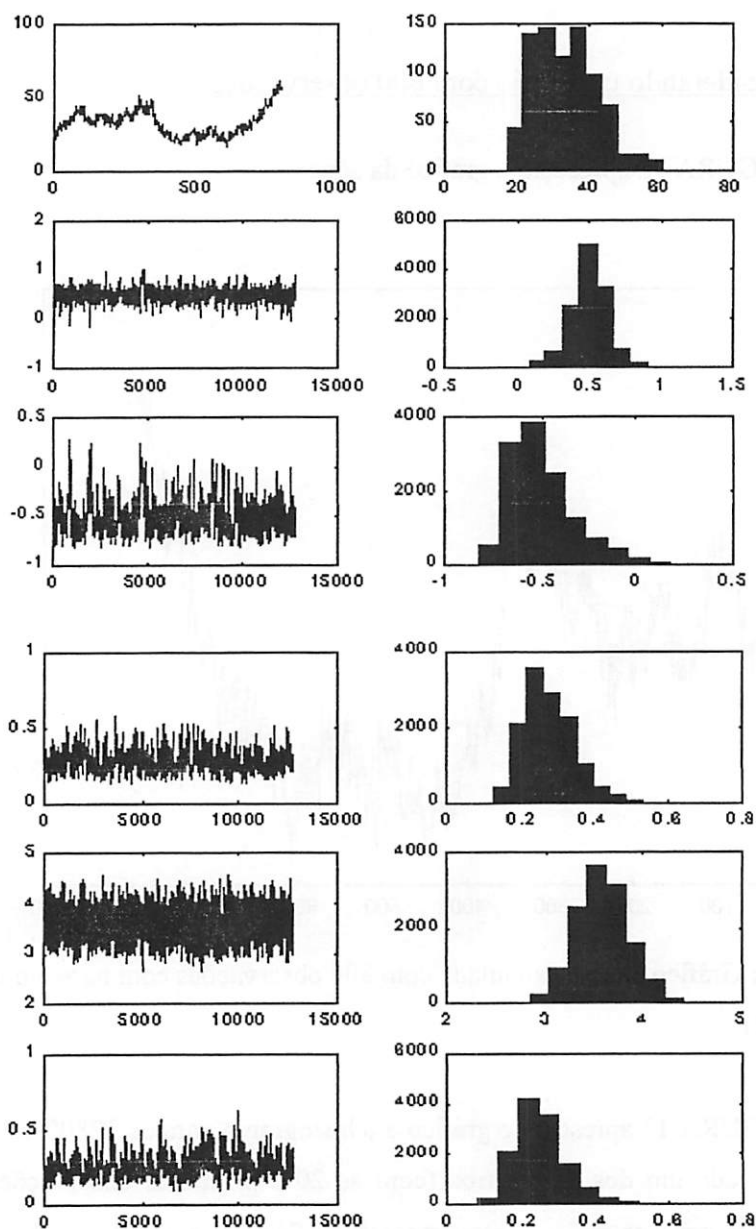


FIGURA 13: Gráfico e histograma dos valores da série gerada e dos parâmetros ϕ_1 , ϕ_2 , V , W e U , respectivamente, considerando uma série com 800 observações.

Pela FIGURA 13, nota-se que a série para todos os parâmetros parece convergir. Os histogramas para os parâmetros ϕ_1 e W sugerem uma distribuição normal; para ϕ_2 , verifica-se uma assimetria a direita. Os histogramas para V e U, assim como nas outras séries, se assemelham à de uma Gama Inversa.

A análise descritiva dos parâmetros, considerando as 512 observações, pode ser observada na TABELA 5

TABELA 5 – Resultado da inferência sobre os parâmetros, considerando uma série com 800 observações.

Parâmetros	Valor Real	Média	D.P	R	I.C (95%)
ϕ_1	0,6	0,5011	0,1325	0,9998	(0,2414 ; 0,7608)
ϕ_2	-0,4	-0,4995	0,1670	1,0148	(-0,8268 ; -0,1722)
V	0,5	0,2756	0,0680	1,0005	(0,1423 ; 0,4089)
W	0,9	3,5731	0,2549	1,0008	(3,0735 ; 4,0727)
U	0,9	0,2529	0,0707	1,0015	(0,1143 ; 0,3915)

Com base na TABELA 5, vê-se que as estimativas para ϕ_1 e ϕ_2 se aproximam dos valores dos reais. O mesmo não ocorreu para as estimativas das variâncias V e U, as quais entretanto, assumiram valores menores do que o estipulado. A variância W novamente apresentou valor bem maior do que o real. O valor da estatística R atingiu valores desejados, afirmando convergência de cada um dos cinco parâmetros.

Na FIGURA 14 são apresentados os gráficos considerando as 512 observações utilizadas para se obterem as estimativas.

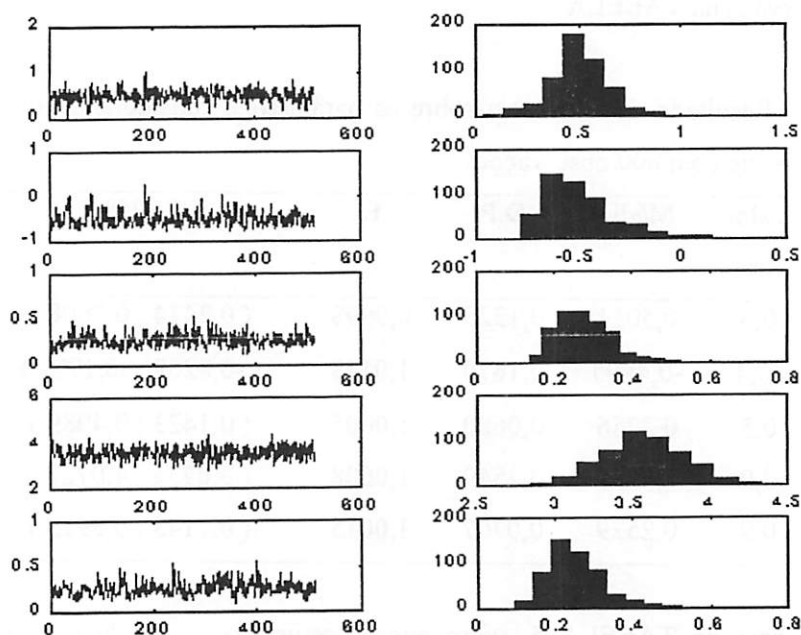


FIGURA 14: Gráfico e histograma da série utilizadas nas estimativas de ϕ_1 e ϕ_2 , V , W e U , respectivamente, considerando uma série com 800 observações.

Com base nas estimativas obtidas, a comparação das séries obtidas a partir de previsões um passo à frente pode ser vista na FIGURA 15.

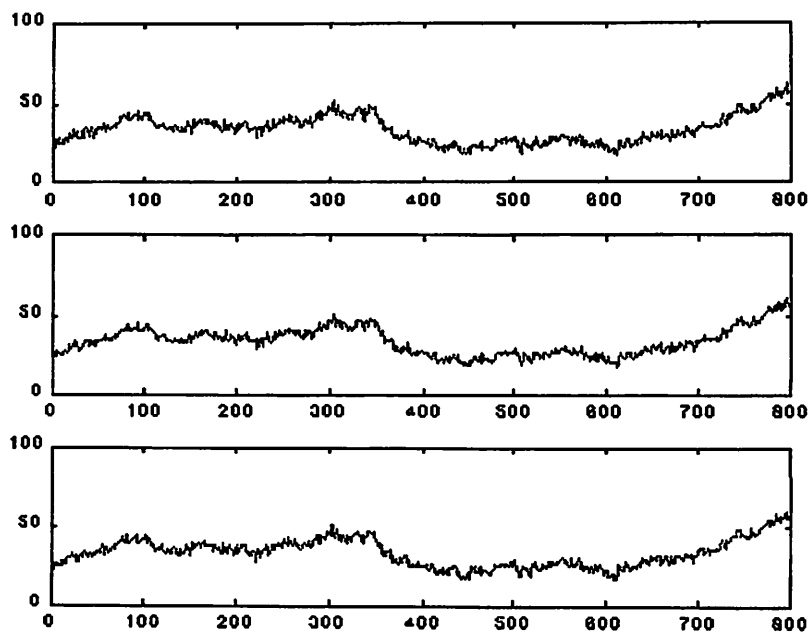


FIGURA 15: : Gráfico da série original, atualizada e “amostrada para trás”, respectivamente, com base nas estimativas obtidas.

Pela FIGURA 15, nota-se um melhor ajuste em relação à série real para a série “prevista” a partir da “amostragem para trás”. As estimativas podem ser vistas na TABELA 6.

TABELA 6 – Estimativa dos erros de previsão um passo à frente para as duas séries “previstas”.

	Estimado	DP	EQM
Série atualizada	0,0813	2,0575	4,2399
Série “amostrada para trás”	-0,0094	0,5259	0,2764

Pela TABELA 6, nota-se um menor erro de previsão quando se utilizam os vetores de estados a partir da “amostragem para trás”. O gráfico dos erros de previsão para as séries pode ser visto na FIGURA 16.

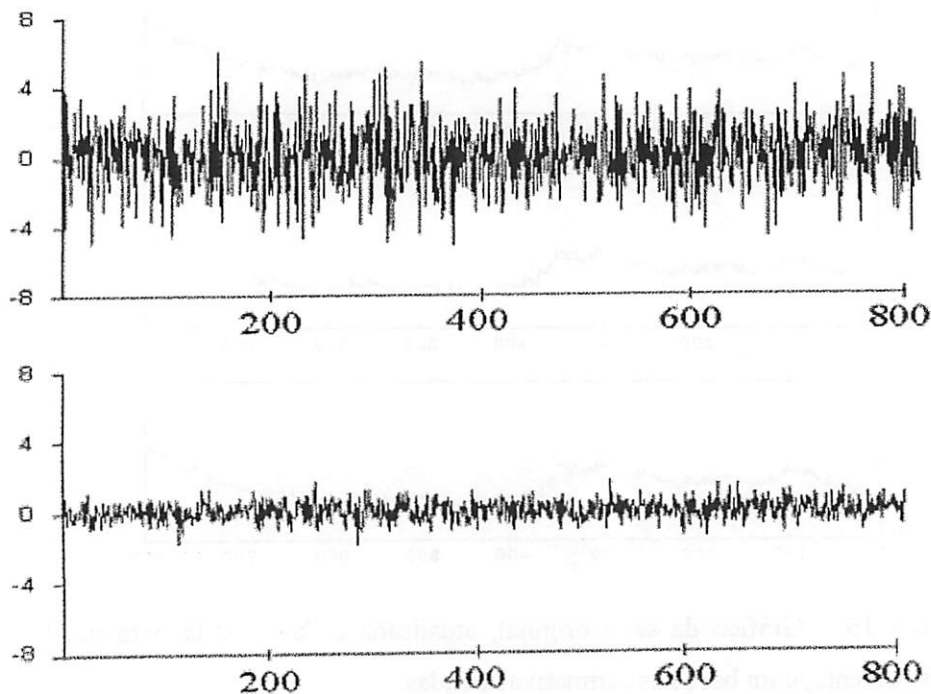


FIGURA 16 – Erros de previsão considerando a série atualizada e “amostrada para trás”, respectivamente.

A comparação dos resultados obtidos para as três séries analisadas pode ser vista na TABELA 7.

TABELA 7 – Comparação das estimativas dos parâmetros e do erro quadrático médio obtido para cada tamanho de série.

	Valor Real	Tamanho da série					
		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	

Pela TABELA 7 pode-se notar: (1) as estimativas de ϕ_1 se aproximaram do valor real para as três situações avaliadas, (2) com relação a ϕ_2 , a estimativa só foi satisfatória para o tamanho da amostra igual a 800, (3) a variância W apresentou valores bem maiores do que o real nas três situações, valor este que diminuía à medida que o tamanho da série aumentava, (4) as variâncias V e U apresentaram um comportamento similar e apresentaram estimativas inferiores aos valores reais, (5) o EQM para a série atualizada apresentou valores bem maiores do que para a “série amostrada para trás” em todas as situações e (6) os valores do EQM diminuíram à medida que o tamanho da série aumentou.

4.3 – Aplicação com dados reais

Fator de desconto igual a 0,90

A FIGURA 17 apresenta o gráfico e o histograma dos 12800 valores obtidos para cada um dos parâmetros. Os gráficos estão na seguinte ordem: Série original, ϕ_1 , ϕ_2 , V e U.

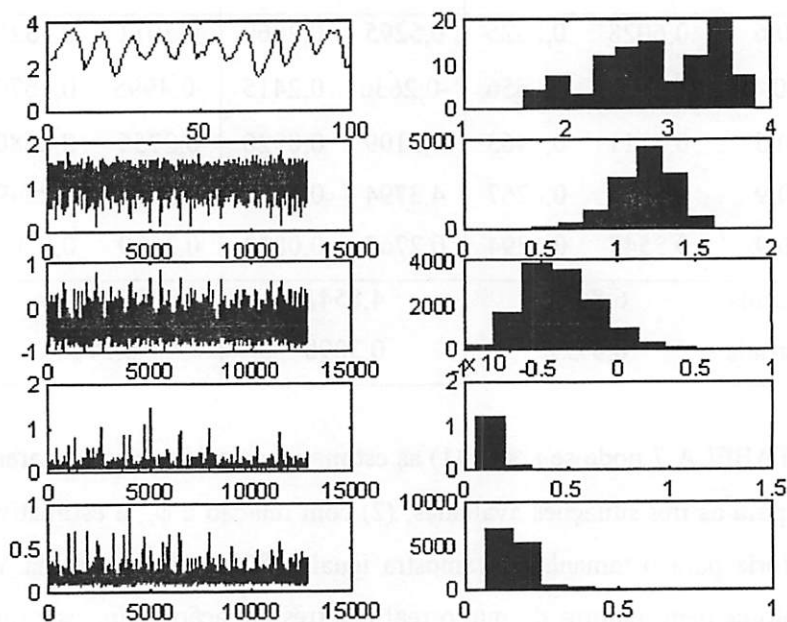


FIGURA 17: Gráfico e histograma dos valores da série $\log_{10} \text{lynx}$ e dos parâmetros ϕ_1 , ϕ_2 , V e U, respectivamente, considerando $\delta = 0,90$.

Pela FIGURA 17, nota-se que a série obtida para cada um dos parâmetros parece convergir. Os histogramas para os parâmetros ϕ_1 e ϕ_2 evidenciam a forma de uma distribuição assimétrica à esquerda e à direita, respectivamente, ao passo que os histogramas para V e U se assemelham à de uma Gama Inversa.

A análise descritiva dos parâmetros, considerando estas 512 observações, pode ser observada na TABELA 8

TABELA 8 – Resultado da inferência sobre os parâmetros, considerando 8000 observações em cada uma das duas cadeias em paralelo e $\delta = 0,90$.

Parâmetros	Estimativa	D.P	R	I.C (95%)
ϕ_1	1,1718	0,1937	1,0004	(0,7921 ; 1,5515)
ϕ_2	-0,3671	0,2319	1,0024	(-0,8216 ; 0,0874)
V	0,1205	0,0483	1,0152	(0,0258 ; 0,2152)
U	0,1623	0,0486	0,9985	(0,0670 ; 0,2576)

Com base na TABELA 8, vê-se que a estimativa de ϕ_2 não se mostrou significativa e o valor estimado de R atingiu valores desejados. Os gráficos com a série e o histograma para os parâmetros estimados com base nestas 512 observações podem ser vistos na FIGURA 18.

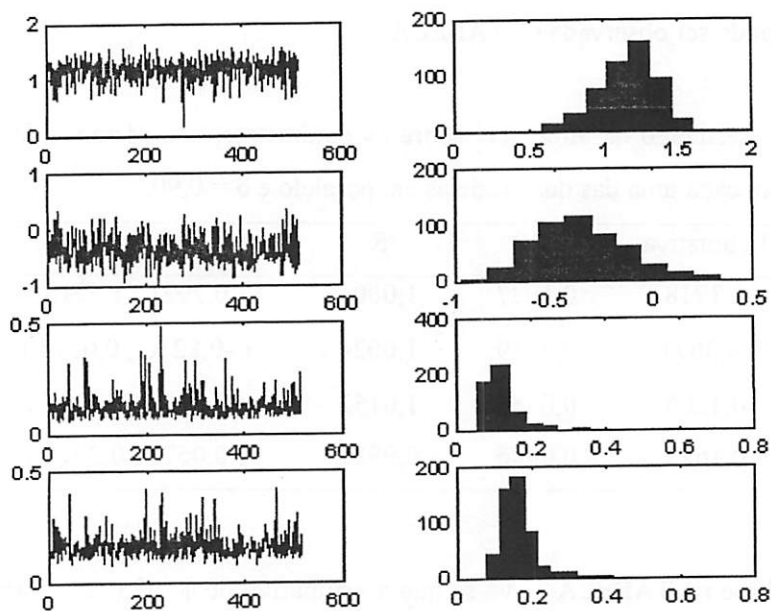


FIGURA 18: Gráfico e histograma dos valores dos parâmetros ϕ_1, ϕ_2, V e U , respectivamente, com base nas 512 observações.

A comparação das séries obtidas a partir de previsões um passo à frente pode ser vista na FIGURA 19.

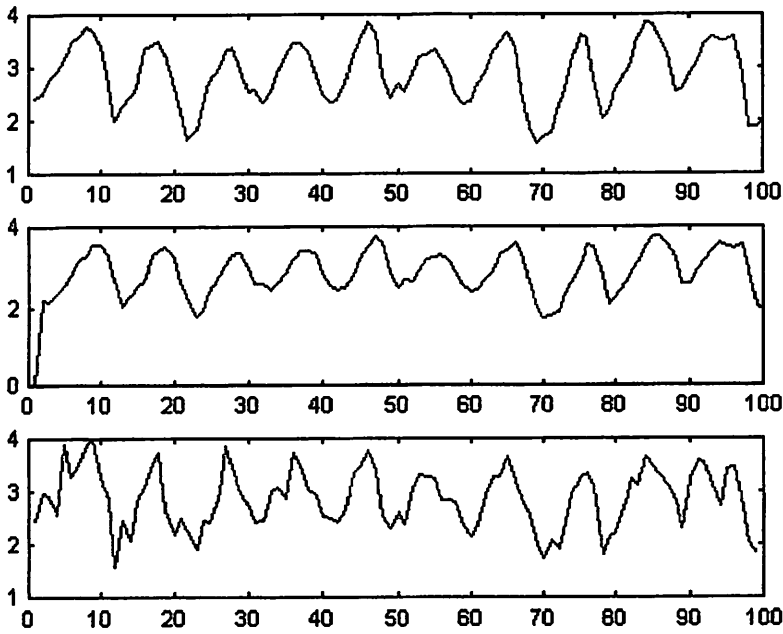


FIGURA 19: Gráfico da série original, atualizada e “amostrada para trás”, respectivamente, com base nas estimativas obtidas.

Pela FIGURA 19, nota-se um melhor ajuste em relação à série real para a série “prevista” a partir da “amostragem para trás”. As estimativas dos erros de previsão podem ser vistas na TABELA 9.

TABELA 9 – Estimativa dos erros de previsão um passo à frente para as duas séries “previstas”.

	Estimado	DP	EQM
Série atualizada	0,0532	0,4524	0,2054
Série “amostrada para trás”	0,0049	0,2764	0,0749

Pela TABELA 9, nota-se um menor erro de previsão quando se utilizam os vetores de estados a partir da “amostragem para trás”. O gráfico dos erros de previsão para as séries pode ser visto na FIGURA 20.

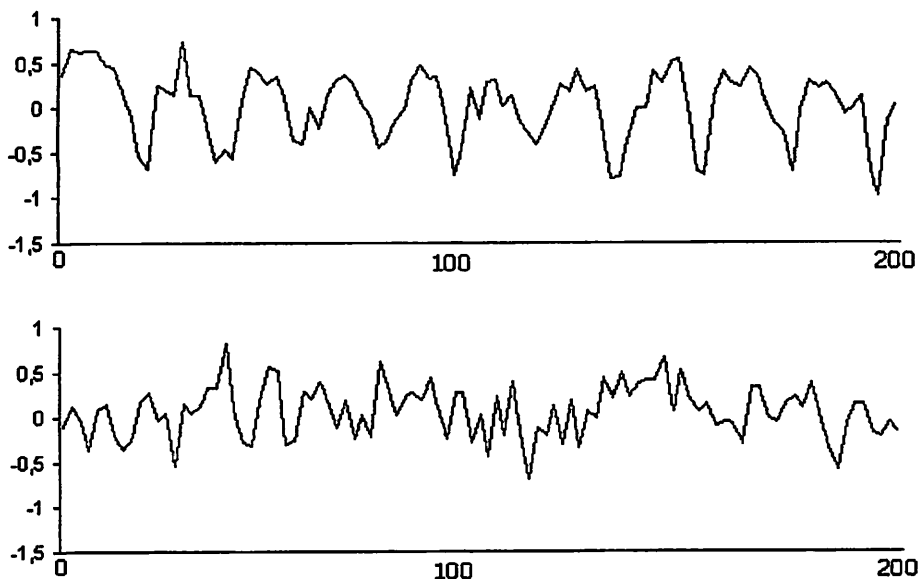


FIGURA 20 – Erros de previsão considerando a série atualizada e “amostrada para trás”, respectivamente.

Fator de desconto igual a 0,95

A FIGURA 21 apresenta o gráfico e o histograma dos 12800 valores obtidos para cada um dos parâmetros. Os gráficos estão na seguinte ordem: Série original, ϕ_1 , ϕ_2 , V e U.

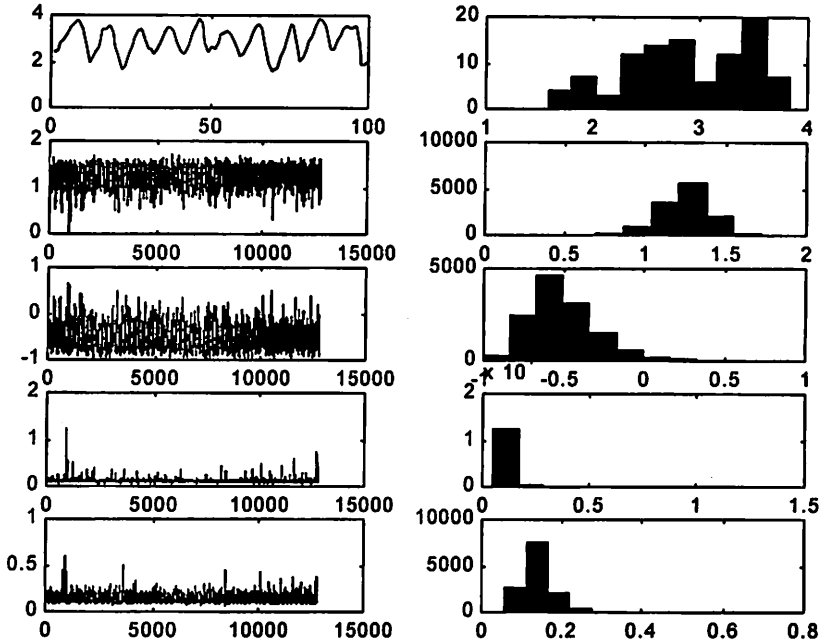


FIGURA 21: Gráfico e histograma dos valores da série \log_{10} lynx e dos parâmetros ϕ_1 , ϕ_2 , V e U, respectivamente, considerando δ igual a 0,95.

Pela FIGURA 21, nota-se que a série obtida para cada um dos parâmetros parece atingir a convergência. Os histogramas para os parâmetros ϕ_1 e ϕ_2 , evidenciam a forma de uma distribuição normal, ao passo que os histogramas para V e U se assemelham à de uma Gama Inversa.

A análise descritiva dos parâmetros, considerando a série com 512, observações pode ser observada na Tabela 10

TABELA 10 – Resultado da inferência sobre os parâmetros, considerando 8000 observações em cada uma das duas cadeias em paralelo e $\delta = 0,95$.

Parâmetros	Estimativa	D.P	R	I.C (95%)
ϕ_1	1,2453	0,1549	1,0033	(0,9417 ; 1,5489)
ϕ_2	-0,5070	0,1999	1,0074	(-0,8988 ; -0,1152)
V	0,1002	0,0280	1,0023	(0,0453 ; 0,1551)
U	0,1412	0,0366	1,0465	(0,0695 ; 0,2129)

Com base na TABELA 10, vê-se que todas as estimativas se mostraram significativas. As estimativas de R atingiram valores desejados, indicando a convergência do amostrador de Gibbs. Os gráficos com a série e o histograma podem ser vistos na FIGURA 22.

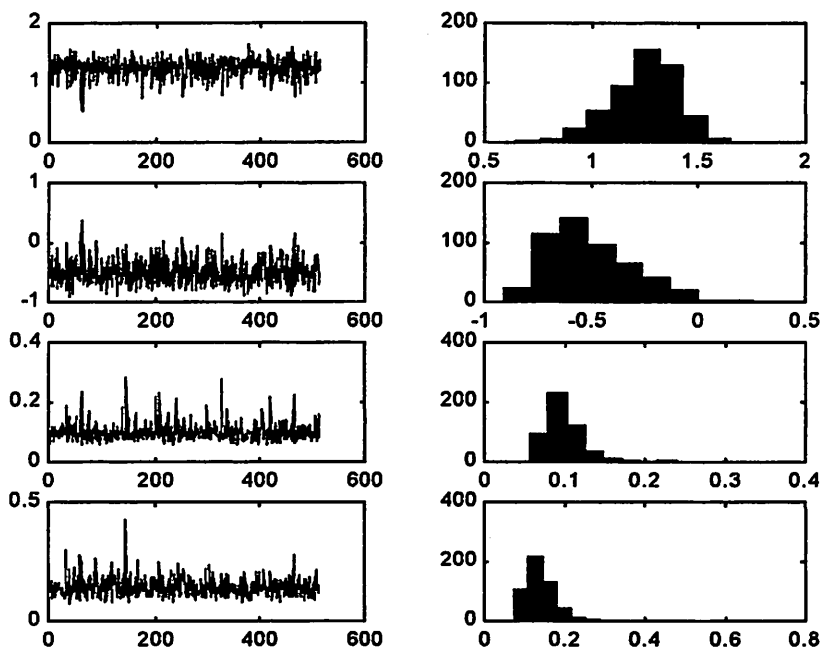


FIGURA 22: Gráfico e histograma dos valores dos parâmetros ϕ_1 , ϕ_2 , V e U, respectivamente, com base nas 512 observações.

A comparação entre as séries previstas pode ser vista na FIGURA 23.

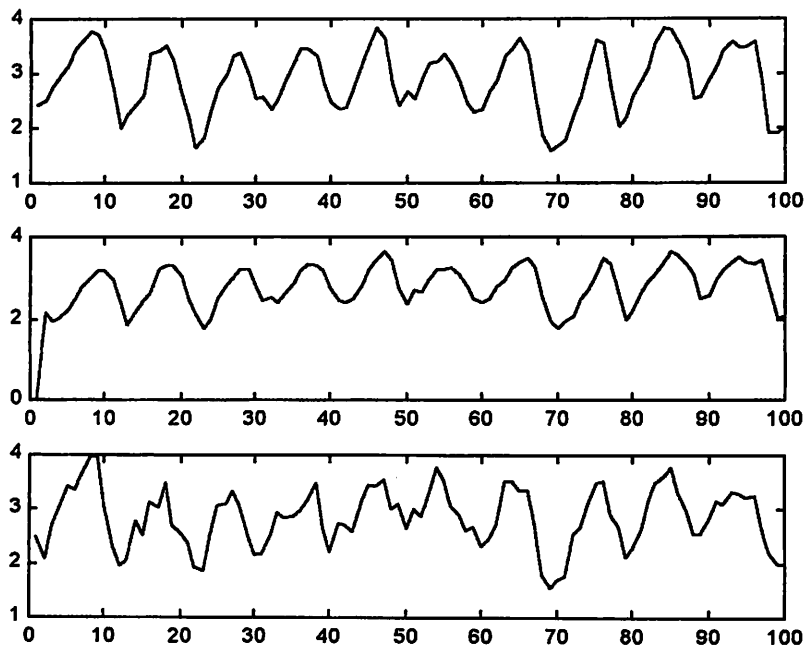


FIGURA 23: Gráfico da série original, atualizada e amostrada para trás, respectivamente, com base nas estimativas obtidas.

Pela FIGURA acima, nota-se novamente um melhor ajuste em relação à série real para a série “prevista” a partir da “amostragem para trás”. As estimativas dos erros de previsão podem ser vistas na TABELA 11.

TABELA 11 – Estimativa dos erros de previsão um passo à frente para as duas séries “previstas”.

	Estimado	DP	EQM
Série atualizada	0,0982	0,4406	0,2018
Série “amostrada para trás”	0,0264	0,2542	0,0640

Pela TABELA 11, nota-se um menor erro de previsão quando se utilizam os vetores de estados a partir da “amostragem para trás”. O gráfico dos erros de previsão para as séries pode ser visto na FIGURA 24.

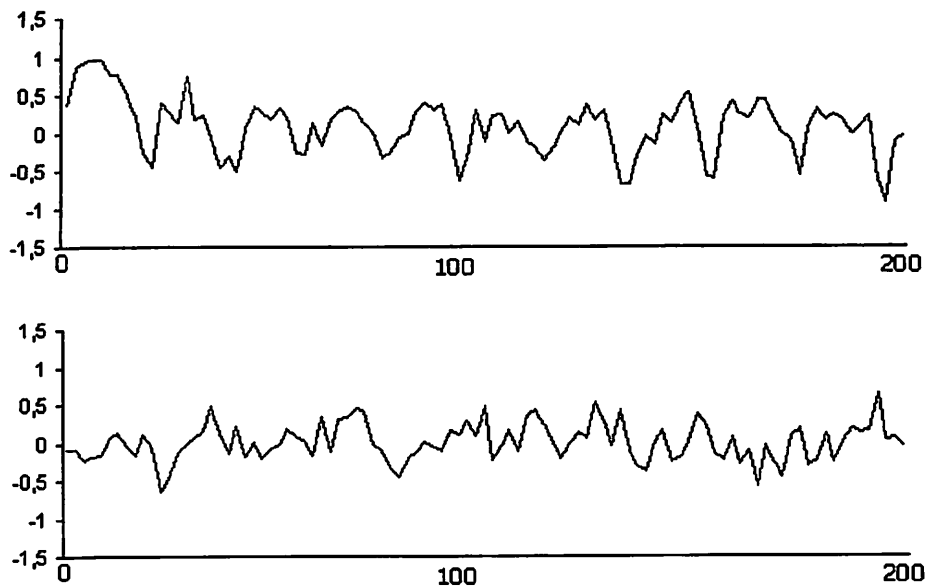


FIGURA 24 – Erros de previsão considerando a série atualizada e “amostrada para trás”, respectivamente.

Fator de desconto igual a 0,99

A FIGURA 25 apresenta o gráfico e o histograma dos 12800 valores obtidos para cada um dos parâmetros. Os gráficos estão na seguinte ordem: Série original, ϕ_1 , ϕ_2 , V e U.

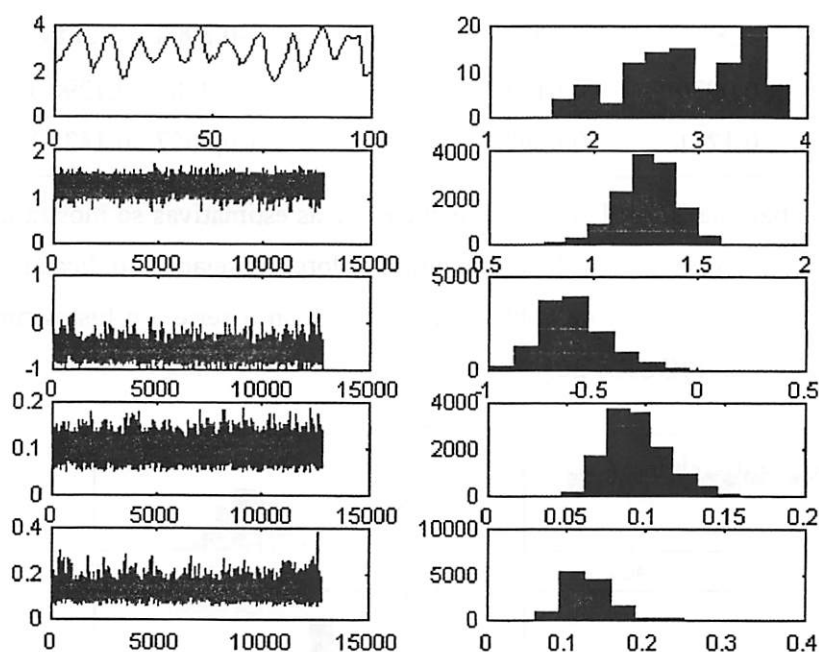


FIGURA 25: Gráfico e histograma dos valores da série $\log_{10} \text{lynx}$ e dos parâmetros ϕ_1 , ϕ_2 , V e U, respectivamente, considerando δ igual a 0,99.

Pela FIGURA 25, nota-se que a série obtida para cada um dos parâmetros parece atingir a convergência. Os histogramas para os parâmetros ϕ_1 e ϕ_2 , evidenciam a forma de uma distribuição normal, ao passo que os histogramas para V e U se assemelham à de uma Gama Inversa.

A análise descritiva dos parâmetros, considerando a série com 512 observações, pode ser observada na TABELA 12.

TABELA 12 – Resultado da inferência sobre os parâmetros, considerando 8000 observações em cada uma das duas cadeias em paralelo e $\delta = 0,99$.

Parâmetros	Estimativa	D.P	R	I.C (95%)
ϕ_1	1,2592	0,1307	1,0028	(1,0030 ; 1,5154)
ϕ_2	-0,5893	0,1536	1,0049	(-0,8904 ; -0,2882)
V	0,0939	0,0183	1,0097	(0,0580 ; 0,1298)
U	0,1241	0,0298	1,0022	(0,0657 ; 0,1825)

Com base na TABELA 12, vê-se que todas as estimativas se mostraram significativas. As estimativas de R atingiram valores desejados, indicando a convergência do amostrador de Gibbs. Os gráficos com a série e o histograma podem ser vistos na FIGURA 26.

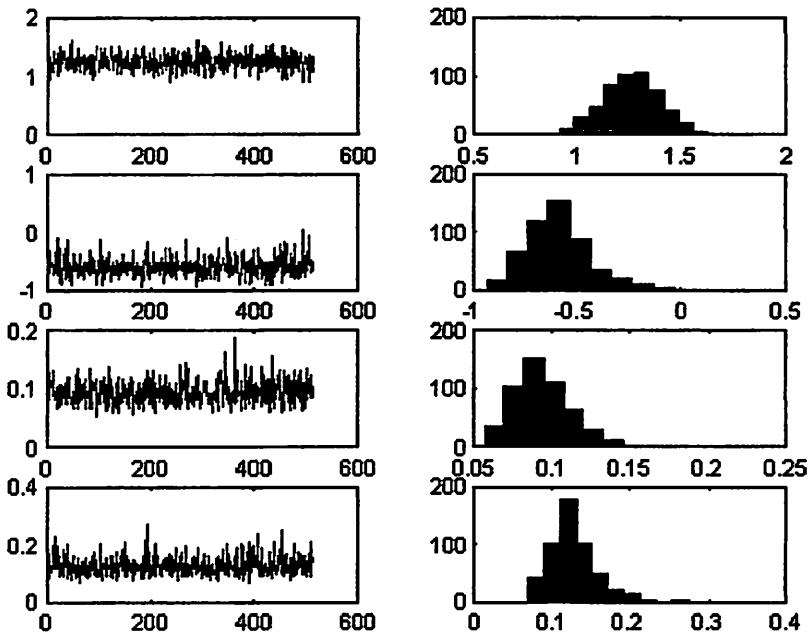


FIGURA 26: Gráfico e histograma dos valores dos parâmetros ϕ_1 , ϕ_2 , V e U, respectivamente, com base nas 512 observações.

A comparação entre as séries previstas pode ser vista na FIGURA 27.

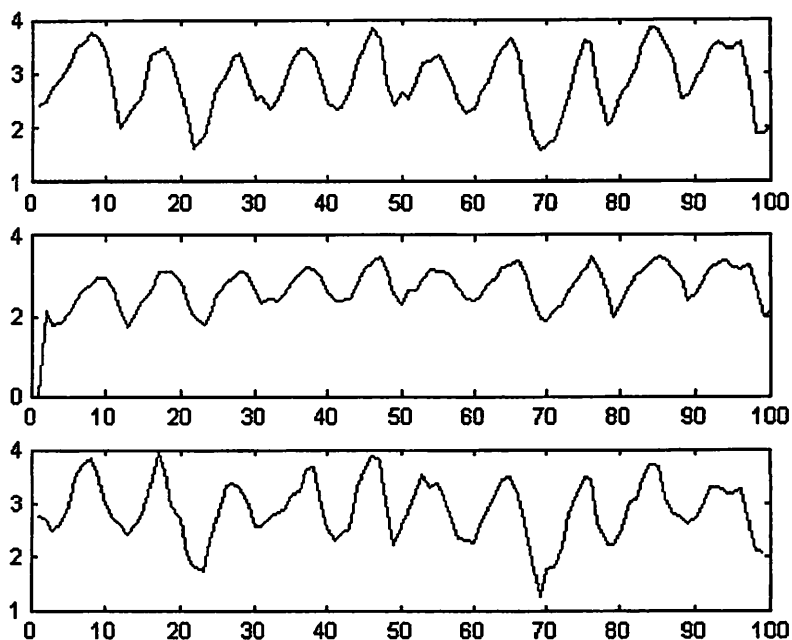


FIGURA 27: Gráfico da série original, atualizada e amostrada para trás, respectivamente, com base nas estimativas obtidas.

Pela FIGURA acima, nota-se novamente um melhor ajuste em relação à série real para a série “prevista” a partir da “amostragem para trás”. As estimativas dos erros de previsão podem ser vistas na TABELA 13.

TABELA 13 – Estimativa dos erros de previsão um passo à frente para as duas séries “previstas”.

	Estimado	DP	EQM
Série atualizada	0,1908	0,4309	0,2202
Série “amostrada para trás”	-0,0005	0,2021	0,0400

Pela TABELA 13, nota-se um menor erro de previsão quando se utilizam os vetores de estados a partir da “amostragem para trás”. O gráfico dos erros de previsão para as séries pode ser visto na FIGURA 28.

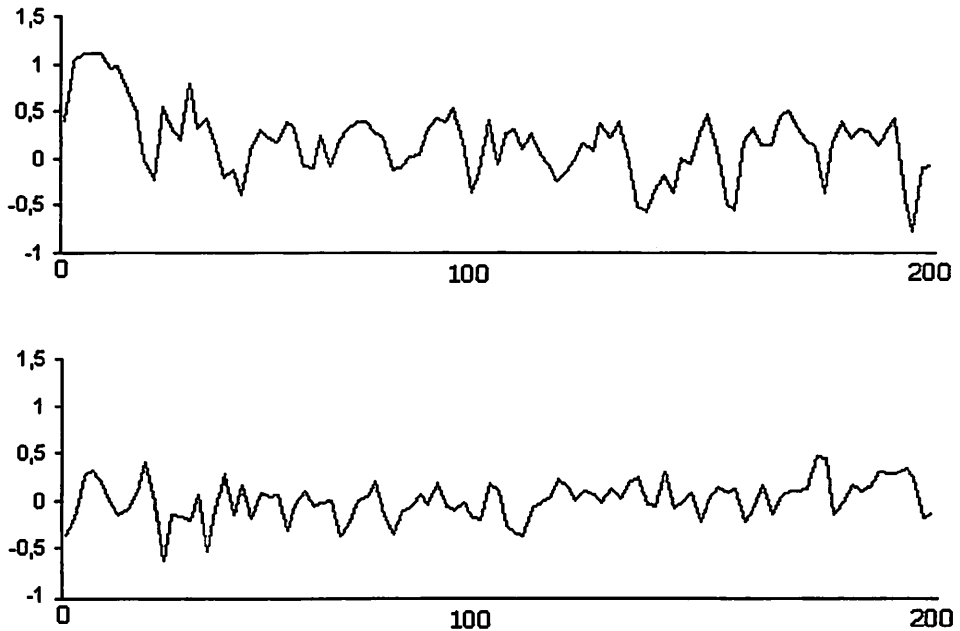


FIGURA 28 – Erros de previsão considerando a série atualizada e “amostrada para trás”, respectivamente.

O interesse se concentra, agora, no ajuste do conjunto de dados segundo um modelo AR(2) a partir da função de verossimilhança. A partir do ajuste, as estimativas dos parâmetros e os erros quadráticos médios podem ser comparadas em relação aos métodos de estimação utilizados. A TABELA 14 apresenta as estimativas dos parâmetros ϕ_1 e ϕ_2 utilizando inferência clássica e $\delta = 0,90, 0,95, 0,99$, assim como os respectivos os erros quadráticos médios.

TABELA 14 – Estimativas obtidas para ϕ_1 e ϕ_2 e erro quadrático médio, segundo a maneira de 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

Com base na TABELA 14, nota-se que o modelo utilizando δ igual a 0,99 apresentou o menor erro quadrático médio, quando comparado com $\delta = 0,90$ e 0,95, e a estimação com base na inferência clássica. Este fato implica em uma melhor performance nas previsões um passo à frente obtidas a partir dos vetores de estado, no processo de “amostragem para trás”.

Considerações Finais

A convergência, medida pelo teste de Gelman e Rubin, não apresentou problemas para nenhuma situação avaliada. Nos resultados de simulação e na aplicação com dados reais, utilizaram-se 8000 iterações em cada cadeia, sendo que 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.

Na implementação do amostrador de Gibbs no caso da aplicação com dados simulados, verificou-se sempre a estacionaridade do componente autorregressivo, a qual é garantida com base nas seguintes condições:

$$\phi_1 + \phi_2 < 1,$$

$$\phi_2 - \phi_1 < 1,$$

$$\phi_2 < 1.$$

Este procedimento foi adotado para evitar que estes parâmetros descaracterizassem a propriedade de estacionaridade.

Os altos valores apresentados pela variância W sugerem uma dificuldade na sua utilização 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 funcionou.

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

Os parâmetros utilizados na distribuição a priori para as variâncias foram α e $\beta = 2$.

5 – 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:

- (1) Foi possível obter as distribuições condicionais completas para os parâmetros avaliados e apresentou-se todo o desenvolvimento e aplicação do algoritmo FFBS para o caso de um modelo AR(2).
- (2) As séries que possuem um maior número de observações foram melhor ajustadas. As estimativas dos parâmetros se aproximaram dos valores reais, com exceção do valor de W , que assumiu valores bem maiores do que o real.
- (3) Os dados da série Canadian Lynx, utilizados na aplicação, se ajustaram muito bem com a aplicação de 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.
- (4) 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. Este fato ocorreu tanto com a série gerada quanto com a série real.
- (5) A modelagem de Séries Temporais via Modelos Lineares Dinâmicos por meio de Inferência Bayesiana se mostrou aplicável, de implementação direta e se ajustou bem tanto à série simulada quanto à série real.

6 - REFERÊNCIAS BIBLIOGRÁFICAS

BOX, G. E. P.,; TIAO, G. C. Bayesian Inference in Statistical Analysis. Massachusetts: Addison-Wesley,. 1973. 668p.

BROEMELING, L. D. Bayesian analysis of linear models. New York: Marcel Decker, 1985. 478p.

CARTER, C.K. E KOHN, R. On Gibbs Sampling for State Space Models. Biométrie, London , v.81, n.3, p. 541-553, Sept. 1994.

CASELLA, G. E GEORGE, I. E. Explaining the Gibbs Sampling. The American Statistician, Washington, v. 46, n.3, p.167 – 174, Aug.1992.

FRUHWIRTH-SCHNATTER, S.. Data Augmentation and Dynamic Linear Models. Journal of Time Series Analysis, Oxford, v.15, p.183 – 202, 1994.

FRUHWIRTH-SCHNATTER, S. Bayesian Model Discrimination and Bayes Factor for Linear Gaussian State Space Models. Journal of Royal Statistical Society B, London, v.57, p.237 – 246, 1995.

GAMERMAN, D. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian inference, London : Chapman and Hall, 1997. 264p.

GELFAND, A. E., HILLS, S. E., RACINE-POON, A. E SMITH, F. M.A. Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. Journal of the American Statistical Association, v.85, n.412, p.972 – 985, Dec.1990.

GELFAND, A. E., SMITH, A.F. M. Sampling-Based Approaches to Calculating Marginal Densities. **Journal of the American Statistical Association**, Alexandria, v.85, n.410, p.398 – 409, June. 1990.

GELMAN, A E RUBIN, D.B. Inference from iterative simulation using multiple sequence. **Statistical Science**, Hayward, v.7, n.4, p. 457-511, 1992.

GLICKMAN, M. E. 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.

JEFFREYS, H. **Theory of Probability**. 3. ed. London: Oxford University Press, 1961. 470p.

NICHOLLS, D.F. ; QUINN, B.G. **Random Coefficients Autorregressive Models: An Introduction**. Lecture notes in Statistics, New York: Springer, , 1982. v. 11.

SMITH, A. F. M.; GELFAND, A. E. (1992). Bayesian Statistics Without Tears: A Sampling – Resampling Perspective. **The American Statistician**, Washington, v. 46, n.2, p. 84 – 88, May. 1992.

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 MAXIMUN ENTROPY AND BAYESIAN METHODS**, 15, 1995. Santa Fé. **Proceedings...Novo México**,: ISDS, 1995. p.1-10.

WEST, M E HARRISON, J. **Bayesian forecasting and dynamic models**. New York: Springer, 1997. 680p.

ZELLNER, A. An Introduction to Bayesian Inference in Econometrics. New York: Wiley, 1971. 448p.

ANEXOS

ANEXO A – DISTRIBUIÇÕES DE PROBABILIDADE

Distribuição Normal

Uma variável aleatória X possui uma distribuição normal com média m e variância V se e somente se:

$$p(X) = \frac{1}{\sqrt{2V\pi}} \exp\left[-\frac{(X-m)^2}{2V}\right],$$

Cuja notação é $X \sim N[m, V]$.

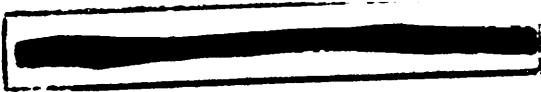
Distribuição Normal Multivariada

Um vetor aleatório n -dimensional \mathbf{X} tem uma distribuição normal multivariada se e somente se $\sum_{i=1}^n a_i X_i$ é normal para qualquer vetor de constantes $\mathbf{a} = \{a_1, \dots, a_n\}$ diferentes de zero.

Se \mathbf{X} é normal multivariado, então $E[\mathbf{X}] = \mathbf{m}$ e $V[\mathbf{X}] = \mathbf{V}$. As quantidades \mathbf{m} e \mathbf{V} definem completamente a distribuição cuja densidade é:

$$p(\mathbf{X}) = \left\{ (2\pi)^n |\mathbf{V}| \right\}^{-1/2} \exp\left[\frac{-(\mathbf{X}-\mathbf{m})' \mathbf{V}^{-1} (\mathbf{X}-\mathbf{m})}{2} \right].$$

E notação dada por $\mathbf{X} \sim N[\mathbf{m}, \mathbf{V}]$.



Tem-se, ainda, que os subvetores de X são independentes se e somente se eles forem não correlacionados. Em particular, se V é bloco diagonal, então os subvetores correspondentes de X são mutuamente independentes.

A distribuição condicional de qualquer subconjunto de X , dados outro subconjunto também possui distribuição normal dada pela seguinte maneira:

Adotando uma partição de X em X_1 e X_2 , tem-se

$$(1) (X_1|X_2) \sim N[m_1(X_2), V_{11}(X_2)], \text{ onde}$$

$$m_1(X_2) = m_1 + V_{12} V_{22}^{-1} (X_2 - m_2) e$$

$$V_{11}(X_2) = V_{11} - V_{12} V_{22}^{-1} V_{21}$$

Para obter $(X_2|X_1)$, basta inverter os subíndices.

3 – Distribuição Gama

Uma variável aleatória $\phi > 0$ possui uma distribuição Gama com parâmetros $n > 0$ e $d > 0$ se e somente se

$$p(\phi) \propto \phi^{n-1} \exp(-\phi d) \quad \phi > 0,$$

cuja notação é $\phi \sim G(n, d)$.

Inserindo a normalização, tem-se:

$$p(\phi) = \frac{d^n}{\Gamma(n)} \phi^{n-1} \exp(-\phi d), \text{ onde } \Gamma(n) \text{ é a função gama.}$$

4 – Distribuição Gama Inversa

Uma variável aleatória $\phi > 0$ possui uma distribuição Gama Inversa com parâmetros $n > 0$ e $d > 0$ se e somente se

$$p(\phi) = \frac{d^n}{\Gamma(n)} \phi^{-(n+1)} \exp\left(-\frac{d}{\phi}\right),$$

cuja notação é $\phi \sim GI(n, d)$.

ANEXO B – PROGRAMA UTILIZADA NOS PROCEDIMENTOS DE SIMULAÇÃO

O programa foi desenvolvido no Matlab 6.0. Os textos após o sinal “%” significam comentários

```
%SIMULAÇÃO - MODELO AR(2)
```

```
%-----
```

```
clear
```

```
%-----
```

```
%GERANDO A SÉRIE
```

```
%-----
```

```
am = 800;           %Tamanho da amostra a ser gerada
```

```
des = 200;          %Observações iniciais da amostra que serão descartadas
```

```
%Geração dos erros
```

```
%-----
```

```
randn('state',1);
```

```
V = 0.5;
```

```
v = normrnd(0,sqrt(V),(am+des+1),1);
```

```
randn('state',2);
```

```
W = 0.9;
```

```
w = normrnd(0,sqrt(W),(am+des+1),1);
```

```
randn('state',3);
```

```
U = 0.9;
```

```
u = normrnd(0,sqrt(U),(am+des+1),1);
```

```
%Valores dos parâmetros
```

```
%-----
```

```
fi1 = 0.6;
```

```
fi2 = -0.4;
```

```
mi(2,1) = 10;
```

```
x(1,1) = 0;
```

```
x(2,1) = 0;
```

%Construção da série

%-----

for i = 3:(am+des+1)

mi(i,1) = mi(i-1,1) + w(i,1);

x(i,1) = fi1*x(i-1,1) + fi2*x(i-2,1) + u(i,1);

yg(i,1) = mi(i,1) + x(i,1) + v(i,1);

end

for i = 1:(am + 1)

y(i,1) = yg((i+des),1);

end

%-----

% MCMC

%-----

it = 2;

nit = 16000 %Número de iterações

l = 0;

F = [1;1;0];

T = [0.1 -0.5; %Valores Iniciais

 -0.2 0.3;

 0,9 0.1;

 0.3 0,9;

 0.6 0.8];

while it < nit + 1 %Início do amostrador de Gibbs

% Utilizado para pegar os valores iniciais da matriz T
% na iteração de número 8001 (início da segunda cadeia)

%-----

if rem(it-2,nit/2)==0 & it<nit

para1(1,1) = T(1,1+(it-2)/(nit/2));

para2(1,1) = T(2,1+(it-2)/(nit/2));

para3(1,1) = T(3,1+(it-2)/(nit/2));

para4(1,1) = T(4,1+(it-2)/(nit/2));

para5(1,1) = T(5,1+(it-2)/(nit/2));

fi1 = para1(1,1);

fi2 = para2(1,1);

V = para3(1,1);

```

W = para4(1,1);
U = para5(1,1);
par1(it-1,1)=para1(1,1);
par2(it-1,1)=para2(1,1);
par3(it-1,1)=para3(1,1);
par4(it-1,1)=para4(1,1);
par5(it-1,1)=para5(1,1);
end

G = [1 0 0;      %Definição da matriz G
      0 fi1 fi2;
      0 1 0];

% Matrizes utilizadas para armazenar as quantidades amostradas
%-----
at = [0;0;0];
At = [0;0;0];
mt = [0;0;0];
tetat = [0;0;0];
Rt = [0 0 0; 0 0 0; 0 0 0];
Ct = [0 0 0; 0 0 0; 0 0 0];

%-----
% FILTRANDO PARA FRENTE
%-----

%Especificação da média e variância a priori para Teta 0
%-----
m = [0;0;0];
C = [0.3 0 0; 0 0.8 0; 0 0 0.7];
WA = [W 0 0; 0 U 0; 0 0 0]; % Matriz W*
teta = mvnrnd(m,C,1)'; % Funciona como a informação inicial

%Início da amostragem para frente
%-----
for i = 1:(am+1)

    %Atualização
    %-----
    a = G*m;
    at = [at a];

```



```

R = G*C*G' + WA;
Rt = [Rt R];

%Previsão um passo à frente
%-----
f(i,1) = F'*a;
Q(i,1) = F'*R*F + V;

%Atualização
%-----
A = (R*F)/Q(i,1);
At = [At A];
e(i,1) = y(i,1) - f(i,1);
m = a + A*e(i,1);
mt = [mt m];
C = R - A*Q(i,1)*A';
Ct = [Ct C];

%Obtenção de teta|D1, ..., tetan|Dn
%-----
teta = mvnrm(m,C,1)';
tetat = [tetat teta];
end %Final da atualização

%-----
% AMOSTRANDO PARA TRÁS
%-----

%Eliminação das primeiras e últimas informações de cada quantidade
%-----
at200 = at(:,2:(am+1));
Rt600 = Rt(:,4:3*(am+1));
f200 = f(1:am,1);
Q200 = Q(1:am,1);
e200 = e(1:am,1);
mt200 = mt(:,2:(am+1));
Ct600 = Ct(:,4:3*(am+1));
tetat200 = tetat(:,2:(am+1));

%Amostragem do vetor tetan|Dn

```

```
%-----
tetabt(1:3,am) = mvnrnd(mt200(1:3,am),Ct600(1:3,(3*am-2):(3*am)),1)';
```

```
%Início da amostragem para trás
```

```
%-----
for j = (am-1):-1:1
    tetab(1:3,j) = mvnrnd(mt200(1:3,j),Ct600(1:3,(3*j-2):(3*j)),1)';
```

```
%Definição das matrizes
```

```
%-----
sigma11 = [Ct600(1,3*j-2) Ct600(1,3*j);
           Ct600(3,3*j-2) Ct600(3,3*j)];
sigma12 = [Ct600(1,3*j-1);
           Ct600(2,3*j)];
sigma21 = sigma12';
sigma22 = Ct600(2,3*j-1);
```

```
%Definição da matriz teta t menos
```

```
%-----
tetamen = [mt200(1,j);
           mt200(3,j)];

mi2 = tetamen + sigma12*inv(sigma22)*[(tetabt(3,j+1) - mt200(2,j))];
sigma = sigma11 - sigma12*inv(sigma22)*sigma21;
```

```
mitetat = [tetabt(1,j+1);
           (tetabt(2,j+1) - fi1*tetabt(3,j+1))/fi2];
sigmatetat = [W 0;
              0 U*(fi2)^(-2)];
```

```
mimenos = inv((inv(sigma) + inv(sigmatetat)))*(inv(sigma)*mi2 +
           inv(sigmatetat)*mitetat);
sigmamenos = inv(inv(sigma) + inv(sigmatetat));
```

```
%Obtenção do vetor teta t menos
```

```
%-----
tetamenos = mvnrnd(mimenos, sigmamenos, 1)';
```

```
%Obtenção do vetor tetat
```

```

%-----
tetabt(1:3,j) = [tetamenos(1,1); tetabt(3,j+1); tetamenos(2,1)];
end          %Fim da amostragem para trás
%Obtenção dos valores de Xt e Mit
%-----
xist(1,1) = tetabt(2,1);

for i = 2:am
    mit(i,1) = tetabt(1,i);
    xist(i,1) = tetabt(2,i);
end

% Cálculo das quantidades wt, et e vt para serem utilizadas
% nas distribuições marginais a posteriori
%-----
for i = 2:am
    wt(i,1) = mit(i,1) - mit((i-1),1);
end

for i = 3:am
    et(i,1) = xist(i,1) - fi1*xist((i-1),1) - fi2*xist(i-2,1);
end

for i = 1:am
    vt(i,1) = y(i,1) - F'*tetabt(1:3,i);
end

%Cálculo do valores de  $X_{t-1}$  e  $X_{t-2}$ 
%-----
for i = 2:am
    xistm1(i,1) = xist(i-1,1);
    if i < 3
        xistm2(i,1) = 0;
    else
        xistm2(i,1) = xist(i-2,1);
    end
end

somaxtm12 = xistm1'*xistm1;
somaxtm22 = xistm2'*xistm2;

```

```

somaxtxtm1 = xist'*xistm1;
somaxtxtm2 = xist'*xistm2;
somaxtm1xtm2 = xistm1'*xistm2;

%Amostragem de fi1 e fi2 com base na distribuição marginal a posteriori
%-----
Afi = [somaxtm12 somaxtm1xtm2;
       somaxtm1xtm2 somaxtm22];
Bfi = [somaxtxtm1;
       somaxtxtm2];
par1 = mvnmd(inv(Afi)*Bfi, inv(Afi)*U, 1);

par1(it,1) = par1(1,1);
par2(it,1) = par1(2,1);

fi1 = par1(1,1); % Atualização dos valores para
fi2 = par1(2,1); % serem utilizados no amostrador

p1 = fi1;      % Utilizado para verificar a estacionariedade
p2 = fi2;

%-----
% CONDICIONAL PARA O V, W e U
%-----

% Cálculo das quantidades necessárias para a distribuição
% marginal a posteriori para as variâncias
%-----
wt2 = wt'*wt;
et2 = et'*et;
vt2 = vt'*vt;

% Valores dos parâmetros da gama inversa a priori
%-----
a = 2;
b = 2;
n = am;

% Definição dos parâmetros
%-----

a2 = (n+2*a)/2;

```

```

bv = (vt2+2*b)/2;
bu = (et2+2*b)/2;
bw = (wt2+2*b)/2;

```

```

%Cálculo dos valores da Gama Inversa para cada uma das variâncias
%-----

```

```

par3(it,1) = 1/gamrnd(a2,1/bv);
V = par3(it,1); % Atualização do valor

```

```

par4(it,1) = 1/gamrnd(a2,1/bw);
W = par4(it,1); % Atualização do valor

```

```

par5(it,1) = 1/gamrnd(a2,1/bu);
U = par5(it,1); % Atualização do valor

```

```

%Checagem da condição de estacionaridade
%-----

```

```

if (((p1 + p2)<1)&((p2-p1)<1)& (-1<p2) & (p2<1))
    it=it +1

```

```

else it=it

```

```

    f1 = par1(it,1);

```

```

    f2 = par2(it,1);

```

```

    V = par3(it,1);

```

```

    W = par4(it,1);

```

```

    U = par5(it,1);

```

```

    l = l+1

```

```

end

```

```

end    %Fim do amostrado de Gibbs

```

```

%Eliminação das 20% primeiras observações em cada cadeia
%-----

```

```

k=1;

```

```

per = 0.2;

```

```

perc = 0.2*(nit/2);

```

```

for j=(perc+1):(nit/2):nit;

```

```

    for i=j:j+((nit/2)-(perc+1));

```

```

        fiest1x(k,1)=par1(i,1);

```

```

        fiest2x(k,1)=par2(i,1);

```

```

        vvx(k,1)=par3(i,1);

```

```

        wwvx(k,1)=par4(i,1);

```

```

    uux(k,1)=par5(i,1);
    k=k+1;
end
end

```

```

%Estatísticas descritivas

```

```

%-----
medias1 = [mean(fiest1x) mean(fiest2x) mean(vvx) mean(wwx) mean(uux)]
desvios = [std(fiest1x) std(fiest2x) std(vvx) std(wwx) std(uux)]

```

```

% Utilização de 1 observação a cada 25

```

```

%-----
k=1;
i = 1;
for k = 1:(8*nit/10)
    if rem(k,25)==0
        fiest125(i,1) = fiest1x(k,1);
        fiest225(i,1) = fiest2x(k,1);
        vv25(i,1) = vv(k,1);
        ww25(i,1) = ww(k,1);
        uu25(i,1) = uux(k,1);
        i = i+1;
    end
end

```

```

%Estatísticas descritivas

```

```

%-----
medias = [mean(fiest125) mean(fiest225) mean(vv25) mean(ww25)
mean(uu25)]
desvios = [std(fiest125) std(fiest225) std(vv25) std(ww25) std(uu25)]

```

```

%Implementação do teste de Gelman e Rubin

```

```

%-----
nc=2;
d=size(fiest125,1)/2;
n1=1;
n2=d;

```

%Separação das observações em duas cadeias

%-----
pfi1 = [fiest125(n1:n2,1) fiest125(n1+d:n2+d,1)];
pfi2 = [fiest225(n1:n2,1) fiest225(n1+d:n2+d,1)];
pV = [vv25(n1:n2,1) vv25(n1+d:n2+d,1)];
pW = [ww25(n1:n2,1) ww25(n1+d:n2+d,1)];
pU = [uu25(n1:n2,1) uu25(n1+d:n2+d,1)];

%Cálculo da estatística R para os parâmetros baseado no programa gr

%-----
Rfi1=gr(pfi1)
Rfi2=gr(pfi2)
RV=gr(pV)
RW=gr(pW)
RU=gr(pU)

%Utilização das estimativas para o cálculo

%das previsões um passo à frente

%-----
fi1 = mean(fiest125);
fi2 = mean(fiest225);
V = mean(vv25);
W = mean(ww25);
U = mean(uu25);

F = [1;1;0];
G = [1 0 0;
0 fi1 fi2;
0 1 0];

at = [0;0;0];
At = [0;0;0];
mt = [0;0;0];
tetat = [0;0;0];
Rt = [0 0 0; 0 0 0; 0 0 0];
Ct = [0 0 0; 0 0 0; 0 0 0];

%-----
% Equações de atualização
%-----

```

m = [0;0;0];
C = [0.3 0 0; 0 0.8 0; 0 0 0.7];

```

```

WA = [W 0 0; 0 U 0; 0 0 0];

```

```

teta = mvnrnd(m,C,1)';

```

```

for i = 1:(am+1)

```

```

    a = G*m;

```

```

    at = [at a];

```

```

    R = G*C*G' + WA;

```

```

    Rt = [Rt R];

```

```

    f(i,1) = F'*a;

```

```

    Q(i,1) = F'*R*F + V;

```

```

    A = (R*F)/Q(i,1);

```

```

    At = [At A];

```

```

    e(i,1) = y(i,1) - f(i,1);

```

```

    m = a + A*e(i,1);

```

```

    mt = [mt m];

```

```

    C = R - A*Q(i,1)*A';

```

```

    Ct = [Ct C];

```

```

    teta = mvnrnd(m,C,1)';

```

```

    tetat = [tetat teta];

```

```

end

```

```

%-----

```

```

% Amostragem para trás

```

```

%-----

```

```

at200 = at(:,2:(am+1));

```

```

Rt600 = Rt(:,4:3*(am+1));

```

```

f200 = f(1:am,1);

```

```

Q200 = Q(1:am,1);

```

```

e200 = e(1:am,1);

```

```

mt200 = mt(:,2:(am+1));

```

```

Ct600 = Ct(:,4:3*(am+1));

```

```

tetat200 = tetat(:,2:(am+1));

```

```

tetabt(1:3,am) = mvnrnd(mt200(1:3,am),Ct600(1:3,(3*am-2):(3*am)),1)';

```



```

for j = (am-1):-1:1
    tetab(1:3,j) = mvnrnd(mt200(1:3,j),Ct600(1:3,(3*j-2):(3*j)),1)';
    sigma11 = [Ct600(1,3*j-2) Ct600(1,3*j);
               Ct600(3,3*j-2) Ct600(3,3*j)];
    sigma12 = [Ct600(1,3*j-1);
               Ct600(2,3*j)];
    sigma21 = sigma12';
    sigma22 = Ct600(2,3*j-1);
    tetamen = [mt200(1,j);
               mt200(3,j)];

    mi2 = tetamen + sigma12*inv(sigma22)*[(tetabt(3,j+1) - mt200(2,j))];
    sigma = sigma11 - sigma12*inv(sigma22)*sigma21;

    mitetat = [tetabt(1,j+1);
               (tetabt(2,j+1) - fi1*tetabt(3,j+1))/fi2];
    sigmatetat = [W 0;
                  0 U*(fi2)^(-2)];

    mimenos = inv((inv(sigma) + inv(sigmatetat)))*(inv(sigma)*mi2 +
               inv(sigmatetat)*mitetat);
    sigmamenos = inv(inv(sigma) + inv(sigmatetat));
    tetamenos = mvnrnd(mimenos, sigmamenos, 1)';

    tetabt(1:3,j) = [tetamenos(1,1); tetabt(3,j+1); tetamenos(2,1)];
end

% Cálculo da série prevista e do EQM a partir da amostragem para trás
%-----
for j = 1:am
    fb(j,1) = F'*tetabt(:,j);           % série prevista
    eb(j,1) = y(j,1) - fb(j,1);         % erro de previsão
end

%Fim do Programa

```