UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

ISSN 0103-2577

Estimação Bayesiana para Processos PAR (*pm*) com a Transformação de Box-Cox

Marinho G. Andrade

Nº 77

NOTAS

Série Estatística

São Carlos – SP
Mar./2004

Estimação Bayesiana para Processos PAR(p_m) com a Transformação de Box-Cox

Marinho G. Andrade.
Departamento de Ciências da Computação e Estatística - SCE
Instituto de Ciências Matemáticas e de Computação - ICMC
Universidade de São Paulo - Campus de São Carlos
Cx. Postal 668 - 13.560-970 - São Carlos - SP

5 de Abril de 2004

Resumo

Este trabalho considera o problema de inferência e previsão dos modelos auto-regressivos com correlação periódica PAR(p_m). Os problemas de inferência e previsão são tratados no contexto clássico calculando-se estimadores de máxima verossimilhança e no contexto Bayesiano considerando-se densidades a priori não informativas e informativas conjugadas. As estimativas de máxima verossimilhança dos parâmetros desses modelos e das previsões de valores futuros assumem a normalidade da distribuição conjunta da série. Essa hipótese, no entanto, quase sempre não se verifica na prática. Neste trabalho os modelos PAR(p_m) são ajustados em duas situações distintas: 1) assumindo a hipótese de normalidade conjunta e homocedásticidade periódica das séries e 2) fazendo uso da transformação de Box-Cox para assegurar um distribuição conjunta aproximadamente normal e estabilizar a variância periódica da série transformada. No caso em que a trasformação de Box-Cox é utilizada, as previsões de mínimo erro quadrático médio da série original são estimadas com uma aproximação de segunda ordem.

1 Introdução

Muitas séries temporais revelam um comportamento periódico ou sazonal os quais podem ser modelados por processos periódicos. Os processos periódicos foram propostos originalmente em [12] e várias aplicações desses modelos nas áreas de hidrologia e econometria são apresentadas em [19], [22], [29], [30], [4], [16] e [21]. Séries com autocovariância periódicas não devem ser modeladas como processos Sazonais Autoregressivos de Medias Móveis (SARMA) proposto em [8], por serem estes modelos na realidade estacionários, veja [15] para uma discussão matemática deste tópico.

Uma classe de processos periódicos muito pesquisada é a dos processos periódicos auto-regressivos (PAR), originalmente propostos por Thomas e Fiering [24] e amplamente analisados em [17], que apresenta as propriedades assintóticas destes modelos assumindo normalidade do processo, e [26] que mostra as condições necessárias e suficientes para estacionariedade periódica.

Os modelos periódicos auto-regressivos de médias moveis (PARMA) são uma extensão do modelo PAR que inclui a parte de médias móveis. Esses modelos foram desenvolvidos em [29] e estendidos em, [28], [6], [15], as propriedades assintóticas dos modelos PARMA são apresentadas em [2] e [5] assumindo normalidade do processo. Porém, algumas dificuldades de uso desses modelos são destacados em [16].

No entanto, além da periodicidade, muitas séries temporais não são gaussianas. Assim, uma abordagem mais geral seria considerar modelos não-gaussianos [32], [14], [1]. Essa abordagem pode torna-se pouco prática devido ao grande número de modelos a serem testados visando a escolha do melhor modelo para uma série temporal. Além do mais, a escolha do melhor modelo para uma dada série pode não ser a mais adequada para outra o que torna o processo de escolha do melhor modelo ainda mais dispendioso.

Uma abordagem alternativa é transformar a série temporal original em uma série com distribuição aproximadamente normal. Outro objetivo da transformação de uma série é reduzir a correlação temporal dos resíduos do modelo ajustado, pois essa correlação implica em um erro sistemático nas previsões. Entretanto existem métodos não paramétricos tais como o método da esperança condicional [9] e o método aditivo estabilizador de variância [25] que não tratam o problema da correlação serial. O método não-paramétrico auto-regressivo [31] considera esse problema mas envolve a estimação de um grande número de parâmetros exigindo a solução de um problema de otimização com restrição de não-negatividade que pode resultar em um grande esforço computacional.

Neste trabalho estamos propondo o uso da transformação de Box-Cox que tem como objetivo transformar a série original assegurando, de forma aproximada, a normalidade, a homocedásticidade e a linearidade, reduzindo assim a correlação temporal dos resíduos [7]. Além disto esta transformação exige a estimação de um (ou dois) parâmetros além dos parâmetros do modelo. Portanto a transformação de Box-Cox reúne a praticidade com as principais propriedades exigidas de uma transformação para séries temporais (linearidade, normalidade e homocedasticidade).

O objetivo principal deste trabalho é propor uma abordagem Bayesiana para estimação dos parâmetros de um modelo PAR(p_m) com $m = 1, \ldots, S$, ($S = 12$ em séries mensais). A análise Bayesiana é feita considerando densidades de probabilidade a priori não informativas e densidades a priori informativas conjugadas. Quando a transformação de Box-Cox é aplicada nos referimos a esses modelos como BCPAR(p_m) e a análise Bayesiana é feita fazendo uso de técnicas de simulação de Monte Carlo em Cadeia de Markov (MCMC). Os estimadores clássicos de máxima verossimilhança também são apresentados para fins de comparação com a abordagem Bayesiana proposta.

Uma dificuldade encontrada no uso da transformação de Box-Cox em modelos de séries temporais é a estimativa de previsões de mínimo erro quadrático médio (MMSE) para valores futuros da série original, a partir das previsões MMSE feitas para a série transformada, porque estas previsões só podem ser feitas de forma aproximada [13] e [18]. Uma aproximação de primeira ordem é apresentada em [27], para modelos estacionários auto-regressivos AR(1). Neste trabalho mostramos que essa aproximação é equivalente às estimativas de previsões de mínimo erro absoluto (ou percentual) médio (MMPE) e estendemos essa aproximação para uma aproximação de segunda ordem para modelos PAR.

Neste trabalho as ordens p_m dos modelos auto-regressivos são selecionadas usando a decomposição do critério BIC [23] para modelos PAR(p_m) sugerida em [16]. Quando adotamos esse procedimento para as séries originais e transformadas, juntamente com a avaliação dos erros de previsão, observamos as vantagens do uso da transformação de Box-Cox tanto na diminuição dos erros de previsão como na escolha de modelos mais parcimoniosos.

Esse trabalho visa aplicações às séries de vazões naturais, afluentes em reservatórios do sistema hidroelétrico brasileiro. Estas séries históricas correspondem a um período de janeiro de 1931 a dezembro de 1998. Para comparar os modelos ajustados, consideramos três situações hidrológicas distintas, escolhendo de cada série, os anos mais secos, os anos típicos (médios) e os anos mais úmidos. As observações que precedem esses anos são usadas para ajustar os modelos e as observações de cada um desses períodos servem como teste para avaliar as previsões obtidas com cada modelo nas três estações: seca, média e úmida.

Este trabalho está organizado da seguinte forma: Nas seções 2 e 3 apresentamos os estimadores de máxima verossimilhança para os modelos PAR e BCPAR. Nas seções 4 e 5 apresentamos a abordagem Bayesiana para ambos os modelos utilizando técnicas de simulação MCMC. Na seção 6 apresentamos os estimadores da previsão para ambos os modelos no contexto clássico. Na seção 7 os estimadores Bayesianos para as previsões são apresentados usando técnicas de simulação de Monte Carlo. Na seção 8 apresentamos algumas conclusões deste trabalho.

2 Modelos PAR(p_m)

Vamos considerar a série temporal periódica, com período S, $\{z_{t(r,m)}, r = 1, \dots n, m = 1, \dots, S\}$ com $t(r,m) = S(r-1) + m$. No caso de observações mensais, $S = 12$, m representa os meses e r representa os anos. Se

$$\mu_m = E\{z_{t(r,m)}\} \tag{1}$$

$$c_{m,k} = cov\{z_{t(r,m)} z_{t(r,m)-k}\} \tag{2}$$

existem e dependem somente de k e m, $z_{t(r,m)}$ é dita ser periodicamente correlacionada ou periodicamente estacionária [16]. O modelo periódico auto-regressivo

PAR(p_m) com $m = 1, \ldots, S$, é dado por:

$$z_{t(r,m)} = \mu_m + \sum_{j=1}^{p_m} \phi_{m,j}(z_{t(r,m)-j} - \mu_{m-j}) + a_{t(r,m)} \tag{3}$$

com $p_m < S$ e $a_{t(r,m)}$ é um processo independente com distribuição normal $N(0, \omega_m^2)$ (vamos usar a notação $\tau_m = 1/\omega_m^2 > 0$). A variância da série $z_{t(r,m)}$ é dada por:

$$\sigma_m^2 = \omega_m^2 + \sum_{j=1}^{p_m} \phi_{m,j} c_{m,j} \tag{4}$$

Sob a hipótese de normalidade, um estimador assintótico eficiente de $c_{m,j}$ [17], é dado por:

$$\widehat{c}_{m,j} = \frac{1}{n} \sum_{r=1}^{n} (z_{t(r,m)} - \mu_m)(z_{t(r,m)-j} - \mu_{m-j}) \tag{5}$$

Segundo [17] e [16], estimadores assintóticos eficientes para os parâmetros $\phi_m = (\phi_{m,1}, \ldots, \phi_{m,p_m})$ podem ser encontrados resolvendo-se a equação e Yule-Walker para o modelo (3) e a variância da série $z_{t(r,m)}$ pode ser estimada substituindo-se as estimativas calculadas com (5) em (4)

Neste trabalho vamos adotar o modelo PAR(p_m) escrito de forma equivalente dada por:

$$z_{t(r,m)} = c_m + \sum_{j=1}^{p_m} \phi_{m,j} z_{t(r,m)-j} + a_{t(r,m)} \tag{6}$$

A relação entre as médias μ_m e as constantes c_m, é dada por:

$$c_m = \mu_m - \sum_{j=1}^{p_m} \phi_{m,j} \mu_{m-j} \tag{7}$$

A equação (7) pode ser escrita na forma matricial como:

$$\mathbf{M}\mu = \mathbf{c} \tag{8}$$

com $\mu = (\mu_1, \ldots, \mu_S)'$, $\mathbf{c} = (c_1, \ldots, c_S)'$ e $\mathbf{M}$ uma matriz $S \times S$ definida como:

$$\mathbf{M} = \begin{pmatrix} 1 & -a_{1,S-1} & -a_{1,S-2} & \cdots & -a_{1,1} \\ -a_{2,1} & 1 & -a_{2,S-1} & \cdots & -a_{2,2} \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ -a_{S-1,1} & -a_{S-1,2} & -a_{S-1,3} & \cdots & -a_{S-1,S-1} \\ -a_{S,1} & -a_{S,S-1} & -a_{S,S-2} & \cdots & 1 \end{pmatrix} \tag{9}$$

Com $a_{m,j} = -\phi_{m,j}$ para $j \leq p_m$ e $a_{m,j} = 0$ para $p_m < j < S$. Note que $\{\mu_m, m = 1, \ldots, S\}$ é a sazonalidade da séries, portanto estamos considerando que $\mu_0 = \mu_S$, $\mu_{-m} = \mu_{S-m}$ e $\mu_m = \mu_{S+km}$ para todo inteiro $k > 0$.

Neste trabalho, visando a abordagem Bayesiana vamos deduzir a função de verossimilhança e calcular os MLE do modelo (6).

2.1 MLE do modelo PAR(p_m)

A equação (6) não pode ser escrita diretamente para as observações do primeiro ano ($r = 1$) quando $p_m > m$. Para generalizar a equação (6) vamos definir $r = r_m, \ldots, n$ com $r_m = 1$ se $m > p_m$ e $r_m = 2$ se $m \leq p_m$. Assim a densidade de probabilidade condicional de $\mathbf{Z} = \{z_{t(r,m)}, r = r_m, \ldots, n,\ m = 1, \ldots, S\}$, é dada por:

$$f(z_{t(r,m)}|z_{t(r,m)-1}, \ldots, z_{t(r,m)-p_m}) \propto$$

$$\propto \tau_m^{1/2} \exp\left\{-\frac{\tau_m}{2}(z_{t(r,m)} - c_m - \sum_{j=1}^{p_m} \phi_{m,j} z_{t(r,m)-j})^2\right\} \tag{10}$$

Denotando por $\mathbf{Z}_m = (z_{t(r_m,m)}, \ldots, z_{t(n,m)})'$ o vetor das n observações do mês m e por $\mathbf{Z}_0 = (z_{t(r_m,m)-1}, \ldots, z_{t(r_m,m)-p_m})'$ as p_m observações imediatamente anteriores à $\mathbf{Z}_m$. A função de verossimilhança para $\mathbf{Z}_m$ condicionada as observações $\mathbf{Z}_0$, é dada por:

$$p_m(\mathbf{Z}_m|\phi_m, \tau_m, \mathbf{Z}_0) \propto$$

$$\propto \prod_{r=r_m}^{n} \tau_m^{1/2} \exp\{-\frac{\tau_m}{2}(z_{t(r,m)} - c_m - \sum_{j=1}^{p_m} \phi_{m,j} z_{t(r,m)-j})^2\} \tag{11}$$

Denotando por $\theta_m = [c_m, \phi_m']'$ e por $\mathbf{a}_m = (a_{t(r_m,m)}, \ldots, a_{t(n,m)})'$ o vetor dos resíduos associados às observações $\mathbf{Z}_m$, então a equação (6) pode ser escrita na forma matricial como:

$$\mathbf{Z}_m = \mathbf{X}_m \theta_m + \mathbf{a}_m \tag{12}$$

onde $\mathbf{X}_m$ é uma matriz $(n - r_m + 1) \times p_m$ dada por:

$$\mathbf{X}_m = \begin{pmatrix} 1 & z_{t(r_m,m)-1} & z_{t(r_m,m)-2} & \cdots & z_{t(r_m,m)-p_m} \\ 1 & z_{t(r_m+1,m)-1} & z_{t(r_m+1,m)-2} & \cdots & z_{t(r_m+1,m)-p_m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & z_{t(n,m)-1} & z_{t(n,m)-2} & \cdots & z_{t(n,m)-p_m} \end{pmatrix} \tag{13}$$

Substituindo a equação (12) na (11) a função de verossimilhança para as observações $\mathbf{Z}_m$ é escrita como:

$$p_m(\mathbf{Z}_m|\theta_m, \tau_m, \mathbf{Z}_0) \propto \tau_m^{(n-r_m)/2} \times \tag{14}$$
$$\times \exp\left\{-\frac{\tau_m}{2}(\mathbf{Z}_m - \mathbf{X}_m\theta_m)'(\mathbf{Z}_m - \mathbf{X}_m\theta_m)\right\}$$

A função de verossimilhança (14) pode ser escrita como:

$$p_m(\mathbf{Z}_m|\theta_m, \tau_m, \mathbf{Z}_0) \propto \tau_m^{(n-r_m)/2} \exp\left\{-\frac{\tau_m}{2}[(\theta_m - \widehat{\theta}_m)'(\mathbf{X}_m'\mathbf{X}_m)(\theta_m - \widehat{\theta}_m) + \right.$$

$$\left. + (\mathbf{Z}_m - \widehat{\mathbf{Z}}_m)'(\mathbf{Z}_m - \widehat{\mathbf{Z}}_m)]\right\} \tag{15}$$

$\widehat{\theta}_m = (\mathbf{X}'_m\mathbf{X}_m)^{-1}(\mathbf{X}'_m\mathbf{Z}_m)$ é o estimador de máxima verossimilhança para θ_m e $\widehat{\mathbf{Z}}_m = \mathbf{X}_m\widehat{\theta}_m$. O estimador de máxima verossimilhança para τ_m^{-1} é dado por:

$$\widehat{\tau_m^{-1}} = \frac{1}{n - r_m}(\mathbf{Z}_m - \mathbf{X}_m\widehat{\theta}_m)'(\mathbf{Z}_m - \mathbf{X}_m\widehat{\theta}_m) \tag{16}$$

A função de verossimilhança para $\mathbf{Z}$ condicionada à $\mathbf{Z}_0$ é dada por

$$p(\mathbf{Z}|\Theta, \Gamma, \mathbf{Z}_0) = \prod_{m=1}^{S} p_m(\mathbf{Z}_m|\theta_m, \tau_m, \mathbf{Z}_0) \tag{17}$$

Na equação (17), estamos denotando por $\mathbf{\Theta} = (\theta_1, \cdots, \theta_S)$ e $\Gamma = (\tau_1^{-1}, \ldots, \tau_S^{-1})$.

As propriedades assintóticas dos estimadores dos parâmetros $\widehat{\phi}_m$ são apresentadas em [5].

3 Modelos BCPAR(q_m)

Considerando a série original $\mathbf{Z} = \{z_{t(r,m)}, r = r_m, \ldots, n,\ m = 1, \ldots, S\}$, com $z_{t(r,m)} > 0$, podemos calcular a séries transformada $\mathbf{Z}^{(\lambda)} = \{z^{(\lambda)}_{t(r,m)}, r = r_m, \ldots, n,$ $m = 1, \ldots, S\}$ por:

$$z^{(\lambda)}_{t(r,m)} = \begin{cases} (z^{\lambda}_{t(r,m)} - 1)/\lambda & \lambda \neq 0 \\ \ln(z_{t(r,m)}) & \lambda = 0 \end{cases} \tag{18}$$

Algumas variações da transformação (18) são propostas em [7], mas (18) é preferível por ser contínua em $\lambda = 0$ e ter um único parâmetro (λ) a ser estimado junto com os parâmetros do modelo da série transformada.

Assumindo que a série temporal transformada $\mathbf{Z}^{(\lambda)}$ é um processo PAR(q_m), dado por:

$$z^{(\lambda)}_{t(r,m)} = d_m + \sum_{j=1}^{q_m} \alpha_{m,j} z^{(\lambda)}_{t(r,m)-j} + \epsilon_{t(r,m)} \tag{19}$$

Estamos considerando agora $r_m = 1$ se $m > q_m$ e $r_m = 2$ se $m \leq q_m$, com $q_m < S$ e $\epsilon_{t(r,m)}$ um processo com distribuição normal $N(0, \xi_m^2)$ com $E(\epsilon_{t(r,m)}\epsilon_{(r,m)+k}) = 0$, para todo $k \neq 0$ (vamos usar a notação $\gamma_m = 1/\xi_m^2 > 0$).

Procedendo da mesma forma que foi feito para o modelo PAR(p_m), temos a relação entre as médias $\mu^{(\lambda)} = (\mu_1^{(\lambda)}, \ldots, \mu_S^{(\lambda)})'$ e as constantes $\mathbf{d} = (d_1, \ldots, d_S)$ dada pela relação matricial:

$$\mathbf{M}\mu^{(\lambda)} = \mathbf{d} \tag{20}$$

A matriz $\mathbf{M}$, dada em (9), agora tem seus elementos definidos pelos parâmetros do modelo BCPAR(q_m), como $a_{m,j} = -\alpha_{m,j}$ para $j \leq q_m$ e $a_{m,j} = 0$ para $q_m < j < S$. Se $z^{(\lambda)}_{t(r,m)}$ tem distribuição aproximadamente normal (essa é uma das propriedades da transformação de Box-Cox), a variância $V(z^{(\lambda)}_{t(r,m)}) = \sigma^{2(\lambda)}_m$ tem um estimador assintótico eficiente [17], dado por:

$$\sigma^{2(\lambda)}_m = \xi^2_m + \sum_{j=1}^{q_m} \alpha_{m,j} c^{(\lambda)}_{m,j} \tag{21}$$

as covariâncias $c^{(\lambda)}_{m,j}$ são estimadas substituindo na equação (5) $z_{t(r,m)}$ e μ_m por $z^{(\lambda)}_{t(r,m)}$ e $\mu^{(\lambda)}_m$.

Para encontrar a relação entre μ_m , σ^2_m (média e variância da série original) e $\mu^{(\lambda)}_m$, $\sigma^{2(\lambda)}_m$ (média e variância da série transformada), vamos considerar a função inversa da transformação de Box-Cox $z_{t(r,m)} = G(z^{(\lambda)}_{t(r,m)})$, dada por:

$$G(z^{(\lambda)}_{t(r,m)}) = \begin{cases} (\lambda z^{(\lambda)}_{t(r,m)} + 1)^{1/\lambda} & \lambda \neq 0 \\ \exp(z^{(\lambda)}_{t(r,m)}) & \lambda = 0 \end{cases} \tag{22}$$

e usar uma aproximação de segunda ordem (ver apêndice), dada por:

$$\mu_m \approx G(\mu^{(\lambda)}_m) \left[1 + \frac{1}{2}(1-\lambda) \left(\frac{\sigma^{(\lambda)}_m}{\lambda \mu^{(\lambda)}_m + 1} \right)^2 \right] \tag{23}$$

$$\sigma^2_m \approx (\lambda \mu^{(\lambda)}_m + 1)^{2/\lambda} \left(\frac{\sigma^{(\lambda)}_m}{\lambda \mu^{(\lambda)}_m + 1} \right)^2 \left[1 + \frac{1}{2} \left(\frac{(1-\lambda)\sigma^{(\lambda)}_m}{\lambda \mu^{(\lambda)}_m + 1} \right)^2 \right] \tag{24}$$

A aproximação de primeira ordem dada em [27] para a média e para a variância são $\mu_m \approx G(\mu^{(\lambda)}_m)$ e $\sigma^2_m \approx (\lambda \mu^{(\lambda)}_m + 1)^{2/\lambda} (\sigma^{(\lambda)}_m / \lambda \mu^{(\lambda)}_m + 1)^2$ respectivamente. Notamos portanto, que a aproximação de primeira ordem para a média equivale a fazer $\lambda = 1$ em (23). Porém, $\lambda = 1$ significa não fazer transformação. Portanto, no cálculo do valor esperado da série original é importante considerar uma aproximação maior que um.

3.1 MLE para o Modelo BCPAR(p_m)

Adotando para a série transformada o mesmo procedimento feito para a série original, denotamos as observações de um determinado mês m, por $\mathbf{Z}^{(\lambda)}_m = (z^{(\lambda)}_{t(r_m,m)}, \ldots, z^{(\lambda)}_{t(n,m)})'$, o vetor de parâmetros por $\alpha_m = (\alpha_{m,1}, \alpha_{m,2}, \ldots, \alpha_{m,p_m})'$ e $\beta_m = [d_m, \alpha_m^{'}]'$. Denotamos por $\epsilon_m = (\epsilon_{t(r_m,m)}, \ldots, \epsilon_{t(n,m)})'$ o vetor dos resíduos associado às observações $\mathbf{Z}^{(\lambda)}_m$, então a equação (19) pode ser escrita na forma vetorial como:

$$\mathbf{Z}^{(\lambda)}_m = \mathbf{X}^{(\lambda)}_m \beta_m + \epsilon_m \tag{25}$$

com $\mathbf{X}_m^{(\lambda)}$ a matriz $(n-r_m+1)\times q_m$ dada por

$$\mathbf{X}_m^{(\lambda)} = \begin{pmatrix} 1 & z^{(\lambda)}_{t(r_m,m)-1} & z^{(\lambda)}_{t(r_m,m)-2} & \cdots & z^{(\lambda)}_{t(r_m,m)-q_m} \\ 1 & z^{(\lambda)}_{t(r_m+1,m)-1} & z^{(\lambda)}_{t(r_m+1,m)-2} & \cdots & z^{(\lambda)}_{t(r_m+1,m)-q_m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & z^{(\lambda)}_{t(n,m)-1} & z^{(\lambda)}_{t(n,m)-2} & \cdots & z^{(\lambda)}_{t(n,m)-q_m} \end{pmatrix} \tag{26}$$

A função de verossimilhança para $\mathbf{Z}_m^{(\lambda)}$ condicionada às q_m observações $\mathbf{Z}_0^{(\lambda)} = \{z^{(\lambda)}_{t(r_m,m)-1}, \ldots, z^{(\lambda)}_{t(r_m,m)-q_m}\}$ é dada por:

$$\begin{aligned} p_m(\mathbf{Z}_m^{(\lambda)}|\beta_m,\gamma_m,\lambda,\mathbf{Z}_0^{(\lambda)}) &\propto \gamma_m^{(n-r_m)/2}\times \\ \times\exp\Big\{-\frac{\gamma_m}{2}[(\beta_m-\widehat{\beta}_m^{(\lambda)})'(\mathbf{X}_m^{(\lambda)\prime}\mathbf{X}_m^{(\lambda)})(\beta_m-\widehat{\beta}_m^{(\lambda)}) &+ \\ +(\mathbf{Z}_m^{(\lambda)}-\widehat{\mathbf{Z}}_m^{(\lambda)})'(\mathbf{Z}_m^{(\lambda)}-\widehat{\mathbf{Z}}_m^{(\lambda)})]\Big\} \end{aligned} \tag{27}$$

Na equação (27) temos

$$\widehat{\beta}_m^{(\lambda)} = (\mathbf{X}_m^{(\lambda)\prime}\mathbf{X}_m^{(\lambda)})^{-1}(\mathbf{X}_m^{(\lambda)\prime}\mathbf{Z}_m^{(\lambda)}) \tag{28}$$

$$\widehat{\mathbf{Z}}_m^{(\lambda)} = \mathbf{X}_m^{(\lambda)}\widehat{\beta}_m^{(\lambda)} \tag{29}$$

A função de verossimilhança para a série original $\mathbf{Z}_m$ é dada por

$$p_m(\mathbf{Z}_m|\beta_m,\gamma_m,\lambda,\mathbf{Z}_0) \propto p_m(\mathbf{Z}_m^{(\lambda)}|\beta_m,\gamma_m,\lambda,\mathbf{Z}_0^{(\lambda)})\prod_{r=r_m}^{n} z_{t(r,m)}^{\lambda-1} \tag{30}$$

em (30), $J(z_{t(r,m)}, z^{(\lambda)}_{t(r,m)}) = z^{\lambda-1}_{t(r,m)}$ é o Jacobiano da transformação entre $z_{t(r,m)}$ e $z^{(\lambda)}_{t(r,m)}$.

Com $\beta = (\beta_1,\cdots,\beta_S)$ e $\Gamma = (\gamma_1^{-1},\ldots,\gamma_S^{-1})$ a função de verossimilhança completa para a série original, $\mathbf{Z} = \{z_{t(r,m)},\ r=r_m,\ldots,n,\ m=1,\ldots,S\}$ é dada por

$$p(\mathbf{Z}|\beta,\Gamma,\lambda,\mathbf{Z}_0) = \prod_{m=1}^{S} p_m(\mathbf{Z}_m^{(\lambda)}|\beta_m,\gamma_m,\lambda,\mathbf{Z}_0^{(\lambda)})\prod_{r=r_m}^{n} z_{t(r,m)}^{\lambda-1} \tag{31}$$

O estimador de máxima verossimilhança $\widehat{\beta}_m^{(\lambda)}$ é condicionado no parâmetro λ. Esse tipo de estimador é denominado *estimador profile* de β_m. O estimador de máxima verossimilhança *profile* para $\gamma_m^{-1(\lambda)}$ é dado por:

$$\widehat{\gamma}_m^{-1(\lambda)} = \frac{1}{n-r_m}(\mathbf{Z}_m^{(\lambda)}-\mathbf{X}_m^{(\lambda)}\widehat{\beta}_m^{(\lambda)})'(\mathbf{Z}_m^{(\lambda)}-\mathbf{X}_m^{(\lambda)}\widehat{\beta}_m^{(\lambda)}) \tag{32}$$

O estimador de máxima verossimilhança para λ é calculado substituindo β_m e γ_m por $\widehat{\beta}_m^{(\lambda)}$ e $\widehat{\gamma}_m^{(\lambda)}$ no logarítmo da função de verosimilhança (31) e maximizando a função:

$$g(\lambda|\mathbf{Z}) \propto \sum_{m=1}^{S}\left\{\frac{(n-r_m)}{2}\ln(\widehat{\gamma}_m^{(\lambda)}) + (\lambda-1)\sum_{r=r_m}^{n}\ln(z_{t(r,m)})\right\} \tag{33}$$

Geralmente $g(\lambda|\mathbf{Q})$ é avaliada numericamente variando os valores de λ em um intervalo $[-\lambda_0, \lambda_0]$ com um pequeno incremento. O estimador de λ é dado por:

$$\widehat{\lambda} = \max_{\lambda \in [-\lambda_0, \lambda_0]} g(\lambda|\mathbf{Z}) \tag{34}$$

Os estimadores *profile* dos parâmetros $\beta_m^{(\lambda)}$ e $\gamma_m^{-1(\lambda)}$ são dadas por:

$$\widehat{\beta}_m^{(\widehat{\lambda})} = (\mathbf{X}_m^{(\widehat{\lambda})\prime}\mathbf{X}_m^{(\widehat{\lambda})})^{-1}(\mathbf{X}_m^{(\widehat{\lambda})\prime}\mathbf{Z}_m^{(\widehat{\lambda})}) \tag{35}$$

$$\widehat{\gamma}_m^{-1(\lambda)} = \frac{1}{n-r_m}(\mathbf{Z}_m^{(\lambda)} - \mathbf{X}_m^{(\lambda)}\widehat{\beta}_m^{(\lambda)})'(\mathbf{Z}_m^{(\lambda)} - \mathbf{X}_m^{(\lambda)}\widehat{\beta}_m^{(\lambda)}) \tag{36}$$

As propriedades assintóticas de $\widehat{\beta}_m^{(\widehat{\lambda})}$, $\widehat{\gamma}_m^{-1(\lambda)}$ e $\widehat{\lambda}$ ainda é um problema em aberto.

4 Modelo PAR(p_m) Bayesiano

Iniciamos a abordagem Bayesiana para os modelos PAR(p_m) especificando a densidade de probabilidade a priori não informativa para os parâmetros $\theta_m = [c_m, \phi_m']'$ e τ_m. Para comparar com a abordagem de máxima verossimilhança vamos assumir independência entre θ_m e τ_m, assumir que θ_m tem priori constante e usar a densidade a priori não informativa de Jeffrey para τ_m dada por $p_0(\tau_m) \propto 1/\tau_m$. Assim a priori não informativa conjunta é dada por:

$$p_0(\theta_m, \tau_m) \propto \frac{1}{\tau_m} \tag{37}$$

Um segundo modelo considera densidade a priori informativa conjugada, com as densidades a priori informativas para os θ_m e τ_m, $m = 1, \ldots, S$, dadas por:

$$p_0(\theta_m|\tau_m) \backsim Normal(\delta_0, (\tau_m\Sigma_0)^{-1}) \tag{38}$$

$$p_0(\tau_m) \backsim Gama(a_0, b_0) \tag{39}$$

Portanto $p_0(\theta_m, \tau_m) = p_0(\theta_m|\tau_m)p_0(\tau_m)$ é uma densidade a priori conjugada Normal-Gama.

4.1 Posteriori do Modelo PAR(p_m)

A densidade a posteriori para os parâmetros $[\theta_m, \tau_m]$ é dadas por:

$$p_m(\theta_m, \tau_m|\mathbf{Z}_m, \mathbf{Z}_0) \propto \tau_m^{(n-r_m)/2} \exp\left\{-\frac{\tau_m}{2}[(\theta_m - \widehat{\theta}_m)'(\mathbf{X}_m'\mathbf{X}_m)(\theta_m - \widehat{\theta}_m) + \right.$$

$$\left. + (\mathbf{Z}_m - \widehat{\mathbf{Z}}_m)'(\mathbf{Z}_m - \widehat{\mathbf{Z}}_m)]\right\} p_0(\theta_m, \tau_m) \tag{40}$$

Usando priori conjugada pode ser mostrado [3] que o estimador bayesiano, $\widehat{\theta}_m^{(B)}$ (sob uma função de perda quadrática) é uma média ponderada entre o estimador de máxima verossimilhança $\widehat{\theta}_m^{(MLE)}$ e o parâmetro de localização da priori, δ_0, dado por:

$$\widehat{\theta}_m^{(B)} = (\mathbf{X}_m'\mathbf{X}_m + \boldsymbol{\Sigma}_0)^{-1}(\mathbf{X}_m'\mathbf{X}_m\widehat{\theta}_m^{(MLE)} + \delta_0\boldsymbol{\Sigma}_0) \tag{41}$$

Este resultado afeta diretamente a estimativa da sazonalidade da série $z_{t(r,m)}$ porque o primeiro termo de $\widehat{\theta}_m^{(B)}$ é a constante c_m, que está relacionada com a média sazonal no mês m. Portanto uma informação a priori dada em δ_0, indicando uma característica específica para o mês m, é refletida diretamente em $\widehat{\theta}_m^{(B)}$. Essa pode ser uma propriedade bastante interessante quando estamos preocupados com cálculos de previsões de valores futuros da série.

A densidade a posteriori para os parâmetros do modelo PAR(p_m), $\boldsymbol{\Theta} = [\theta_1, \ldots .\theta_s]$ e $\boldsymbol{\Gamma} = (\tau_1^{-1}, \ldots, \tau_S^{-1})$, assumindo independência entre $[\theta_m, \tau_m]$ e $[\theta_{m+k}, \tau_{m+k}]$ (não esquecendo que $\theta_m = \theta_{m+kS}$ e $\tau_m = \tau_{m+kS}$ para qualquer inteiro $k > 0$), é dada por:

$$p(\Theta, \Gamma|\mathbf{Z}, \mathbf{Z}_0) = \prod_{m=1}^{S} p_m(\theta_m, \tau_m|\mathbf{Z}_m, \mathbf{Z}_0) \tag{42}$$

Na equação (42), $p_m(\theta_m, \tau_m|\mathbf{Z}_m, \mathbf{Z}_0)$, são as densidades a posteriori dadas pela equação (40).

5 Modelo BCPAR(p_m) Bayesiano

Assumindo independência entre β_m e γ_m e considerando a priori não informativa de Jeffreys para γ_m, a densidade a priori não informativa para os parâmetros $\beta_m = [d_m, \alpha_m']'$, γ_m e λ do modelo BCPAR(p_m) é dada por:

$$p_0(\beta_m, \gamma_m, \lambda) \propto \frac{p_0(\lambda)}{\gamma_m} \tag{43}$$

Uma ampla discussão sobre a escolha de $p_0(\beta_m, \gamma_m, \lambda)$ é dada em [20]. As densidades a priori informativas para os parâmetros β_m, γ_m, $m = 1, \ldots, S$. e λ são dadas por

$$p_0(\beta_m, \gamma_m, \lambda) = p_0(\beta_m|\gamma_m, \lambda)p_0(\gamma_m|\lambda)p_0(\lambda) \tag{44}$$

com

$$p_0(\beta_m|\gamma_m, \lambda) \backsim Normal(\delta_0, (\gamma_m\boldsymbol{\Sigma}_0)^{-1}) \tag{45}$$

$$p_0(\gamma_m|\lambda) \backsim Gama(v_0, \omega_0) \tag{46}$$

Para a densidade $p_0(\lambda)$ vamos considerar que $\lambda \in [-\lambda_0, \lambda_0]$, com $\lambda_0 \in R$, e vamos considerar a transformação logarítmica que mapeia o intervalo $[-\lambda_0, \lambda_0]$ nos reais, dada por:

$$\eta = \ln\left(\frac{\lambda + \lambda_0}{\lambda_0 - \lambda}\right) \tag{47}$$

de forma que para $\lambda \longrightarrow \lambda_0$ temos $\eta \longrightarrow +\infty$ e quando $\lambda \longrightarrow -\lambda_0$, temos $\eta \longrightarrow -\infty$ e vamos assumir uma densidade a priori Normal para η dada por:

$$\eta \backsim Normal(0, \sigma_0^2) \tag{48}$$

Notando portanto que $\eta = 0$ corresponde exatamente a $\lambda = 0$, o que significa adotar a transformação logarítmica $z_{t(r,m)}^{(\lambda)} = \ln(z_{t(r,m)})$, ou seja $z_{t(r,m)}$ tem distribuição log-Normal. A função inversa de (47) é dada por:

$$\lambda = \frac{\lambda_0(\exp(\eta) - 1)}{(1 + \exp(\eta))} \tag{49}$$

Usando (49) o modelo BCPAR pode ser escrito usando o parâmetro η em lugar de λ, denotamos essa reparametrizaçao por $\lambda(\eta)$ dada por (49).

5.1 Posteriori do Modelo BCPAR(p_m)

Para o modelo BCPAR(p_m) a densidade a posteriori é dada por:

$$p_m(\beta_m, \gamma_m, \eta|\mathbf{Z}_m, \mathbf{Z}_0) \propto \gamma_m^{(n-r_m)/2} \exp\{-\frac{\gamma_m}{2}[S(\beta_m, \lambda(\eta))+ \tag{50}$$

$$+R(\lambda(\eta))] + (\lambda(\eta) - 1)L(\mathbf{Z}_m)\}p_0(\beta_m, \gamma_m|\eta)p_0(\eta)$$

com $L(\mathbf{Z}_m) = \sum_{r=r_m}^{n} \ln(z_{t(r,m)})$ e:

$$S(\beta_m, \lambda(\eta)) = -\frac{\gamma_m}{2}(\beta_m - \widehat{\beta}_m^{(\lambda)})'(\mathbf{X}_m^{(\lambda)\prime}\mathbf{X}_m^{(\lambda)})(\beta_m - \widehat{\beta}_m^{(\lambda)}) \tag{51}$$

$$R(\lambda(\eta)) = (\mathbf{Z}_m^{(\lambda)} - \widehat{\mathbf{Z}}_m^{(\lambda)})'(\mathbf{Z}_m^{(\lambda)} - \widehat{\mathbf{Z}}_m^{(\lambda)}) \tag{52}$$

A densidade a posteriori para os parâmetros do modelo BCPAR(p_m), $\mathbf{\Psi} = [\beta_1, \ldots .\beta_s]$, $\mathbf{\Gamma} = (\gamma_1^{-1}, \ldots, \gamma_S^{-1})$ e λ, assumindo independência entre $[\beta_m, \gamma_m]$ e $[\beta_{m+k}, \gamma_{m+k}]$ com ($\beta_m = \beta_{m+kS}$, e $\gamma_m = \gamma_{m+kS}$, k inteiro positivo), é dada por:

$$p(\Psi, \Gamma, \eta|\mathbf{Z}, \mathbf{Z}_0) = \prod_{m=1}^{S} p_m(\beta_m, \gamma_m, \lambda(\eta)|\mathbf{Z}_m, \mathbf{Z}_0) \tag{53}$$

Na equação (53), $p_m(\beta_m, \gamma_m, \lambda|\mathbf{Z}_m, \mathbf{Z}_0)$ são as densidades a posteriori dada em (50)

5.2 Condicionais do Modelo BCPAR(p_m)

A estimativa Bayesiana dos parâmetros do modelo BCPAR(p_m) só podem ser obtidas numericamente, neste trabalho vamos calcular essas estimativas usando algritmos de simulação de Monte Carlo (MCMC). Mais especificamente, o algoritmo Gibbs sampling [10] para os parâmetros β_m, γ_m e o Metropolis-Hasting [11] para o parâmetro

λ. Para implementação destes algritmos é necessário escrever as densidades condicionais a posteriori. Usando a equação (50) as densidades condicionais são:

A densidade condicional a posteriori para β_m é dada por:

$$p_m(\beta_m|\gamma_m, \eta, \mathbf{Z}_m, \mathbf{Z}_0) \propto \gamma_m^{(n-r_m)/2} \exp\left\{-\frac{\gamma_m}{2} S(\beta_m, \lambda(\eta))\right\} p_0(\beta_m|\gamma_m) \quad (54)$$

A densidade condicional a posteriori para γ_m é dada por:

$$p_m(\gamma_m|\beta_m, \eta, \mathbf{Z}_m, \mathbf{Z}_0) \propto \gamma_m^{(n-r_m)/2} \exp\left\{-\frac{\gamma_m}{2} B(\beta_m, \lambda(\eta))\right\} p_0(\gamma_m) \quad (55)$$

com $B(\beta_m, \lambda(\eta)) = S(\beta_m, \lambda(\eta)) + R(\lambda(\eta))$.

A densidade condicional a posteriori para η é dada por:

$$p(\eta|\Psi, \Gamma, \mathbf{Z}, \mathbf{Z}_0) \propto \exp\{S(\lambda(\eta))\} p_0(\eta) \quad (56)$$

com

$$S(\lambda(\eta)) = \sum_{m=1}^{S} -\frac{\gamma_m}{2} B(\beta_m, \lambda(\eta)) + (\lambda(\eta) - 1) L(\mathbf{Z}_m) \quad (57)$$

As amostras de λ são calculadas com $\lambda(\eta)$ dado por (49) para cada η gerado da posteriori condicional (56).

6 Previsão no Contexto Clássico

No contexto da máxima verossimilhança os valores futuros da série $z_{t(n+h,m)}, h \geq 1$, $m = 1, \ldots, S$ (valores para o ano $n+h$) têm um previsor de mínimo erro quadrático médio (MMSE), não tendencioso, dado por $\widehat{z}_{t(n+h,m)} = E(z_{t(n+h,m)}|\mathbf{Z})$. Outra previsão que vamos considerar neste trabalho é a previsão de mínimo erro absoluto (ou percentual) médio (MMPE), neste caso a previsão $\widehat{z}_{t(n+h,m)}$, é dada pela mediana da distribuição $p_m(z_{t(n+h,m)}|\mathbf{Z})$. Os erros de previsão considerados neste trabalho são calculados por:

$$MMSE = \frac{1}{H} \sum_{h=1}^{H} (z_{t(n+h,m)} - \widehat{z}_{t(n+h,m)})^2 \quad (58)$$

$$MMPE = \frac{100}{H} \sum_{h=1}^{H} \frac{|z_{t(n+h,m)} - \widehat{z}_{t(n+h,m)}|}{z_{t(n+h,m)}} \quad (59)$$

Quando a densidade de probabilidade dos valores futuros $\{z_{t(n+h,m)}, h \geq 1, m = 1, \ldots, S\}$ é simétrica, como é o caso da distribuição normal, os estimadores MMSE e MMPE coincidem.

6.1 Previsão com PAR(p_m)

Considerando o modelo PAR(p_m) dado em (6), para o ano $(n+h)$, $h \geq 1$, temos:

$$z_{t(n+h,m)} = c_m + \sum_{j=1}^{p_m} \phi_{m,j} z_{t(n+h,m)-j} + a_{t(n+h,m)} \tag{60}$$

Usando a equação (60) $\widehat{z}_{t(n+h,m)} = E(z_{t(n+h,m)}|\mathbf{Z})$ é dado por:

$$\widehat{z}_{t(n+h,m)} = c_m + \sum_{j=1}^{p_m} \phi_{m,j} \widehat{z}_{t(n+h,m)-j} \tag{61}$$

onde $\widehat{z}_{t(n+h,m)-j} = z_{t(n+h,m)-j}$, se $t(n+h,m)-j < Sn$, $j = 1,\ldots,p_m$. Para encontrar uma forma recursiva de calcular as previsões $\widehat{z}_{t(n+h,m)}$, vamos definir o vetor $\mathbf{Z}_{f_h}$ com os S valores futuros da série, para um ano $(n+h)$ e por $\mathbf{a}_{f_h}$ os respectivos ruídos, como:

$$\mathbf{Z}_{f_h} = \left(z_{t(n+h,1)}, z_{t(n+h,2)}, \ldots, z_{t(n+h,S)}\right)' \tag{62}$$

$$\mathbf{a}_{f_h} = \left(a_{t(n+h,1)}, a_{t(n+h,2)}, \ldots, a_{t(n+h,S)}\right)' \tag{63}$$

Então, denotando por $\phi = \{\phi_{m,1},\ldots,\phi_{m,S}, m = 1,\ldots,S\}$ e $\mathbf{C} = (c_1,\ldots,c_m)'$, podemos escrever (60) na forma matricial, por:

$$\mathbf{A}(\phi)\mathbf{Z}_{f_h} + \mathbf{C} + \mathbf{B}(\phi)\mathbf{Z}_{f_{h-1}} = \mathbf{a}_{f_h} \tag{64}$$

com $\mathbf{C} = (c_1,\ldots,c_S)'$ e $\mathbf{A}(\phi)$ e $\mathbf{B}(\phi)$ são matrizes $S \times S$ dadas por:

$$\mathbf{A}(\phi) = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ -\phi_{1,1} & 1 & 0 & \ldots & 0 \\ -\phi_{2,1} & -\phi_{2,2} & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ -\phi_{S,S-1} & -\phi_{S,S-2} & -\phi_{S,S-3} & \ldots & 1 \end{pmatrix} \tag{65}$$

$$\mathbf{B}(\phi) = \begin{pmatrix} -\phi_{1,S} & -\phi_{1,S-1} & -\phi_{1,S-2} & \cdots & -\phi_{1,,1} \\ 0 & -\phi_{2,S} & -\phi_{2,S-1} & \cdots & -\phi_{2,2} \\ 0 & 0 & -\phi_{3,S} & \cdots & -\phi_{3,3} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \phi_{S,S+1} & \cdots & -\phi_{S,S+1} \end{pmatrix} \tag{66}$$

Assim, temos:

$$\mathbf{Z}_{f_h} = -\mathbf{A}(\phi)^{-1}(\mathbf{C} + \mathbf{B}(\phi))\mathbf{Z}_{f_{h-1}} + \mathbf{A}(\phi)^{-1}\mathbf{a}_{f_h} \tag{67}$$

Denotando por $\mathbf{Z}_{f_0}$ as observações do último ano do histórico, ou seja:

$$\mathbf{Z}_{f_0} = \left(z_{t(n,1)}, z_{t(n,2)}, \ldots, z_{t(n,S)}\right)' \tag{68}$$

e escrevendo a equação (67) para $h = 1$ temos a condição inicial para a equação recursiva (67), dada por:

$$\mathbf{Z}_{f_1} = -\mathbf{A}(\phi)^{-1}(\mathbf{C} + \mathbf{B}(\phi)\mathbf{Z}_{f_0}) + \mathbf{A}(\phi)^{-1}\mathbf{a}_{f_1} \tag{69}$$

Assim, escrevendo (67) recursivamente, temos que:

$$\mathbf{Z}_{f_h} = -\left[(\mathbf{A}(\phi)^{-1}(\mathbf{C} + \mathbf{B}(\phi))\right]^h \mathbf{Z}_{f_0} + \sum_{k=1}^{h} \mathbf{D}_k(\phi)\mathbf{a}_{f_k} \tag{70}$$

com:

$$\mathbf{D}_k(\phi) = \left[\mathbf{A}(\phi)^{-1}(\mathbf{C} + \mathbf{B}(\phi))\right]^{h-k} \mathbf{A}(\phi)^{-1} \tag{71}$$

Assumindo que $\mathbf{a}_{f_k}, k = 1, \ldots, h$ são vetores aleatórios independentes com distribuição $N(0, \mathbf{\Gamma})$, $\mathbf{\Gamma} = \text{diag}(\tau_1^{-1}, \ldots, \tau_S^{-1})$, então $\mathbf{Z}_{f_h}$ também é um vetor aleatório normalmente distribuído com média $\mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$ e matriz de covariância $\mathbf{V}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$.

No contexto clássico a previsão para o ano $(n + h)$ usando as observações do último ano n, ou seja as previsões h-passos a frente são calculadas como:

$$\mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0}) = -\left[(\mathbf{A}(\phi)^{-1}(\mathbf{C} + \mathbf{B}(\phi))\right]^h \mathbf{Z}_{f_0} \tag{72}$$

A variância de $\mathbf{Z}_{f_h}$ condicionada a $\mathbf{Z}_{f_0}$ é dada por:

$$\mathbf{V}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0}) = \sum_{k=1}^{h} \mathbf{D}_k(\phi)\Gamma\mathbf{D}_k(\phi)' \tag{73}$$

As estimativas de $\mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$ e $\mathbf{V}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$ são calculadas substituindo-se as estimativas $\widehat{\phi}$ e $\widehat{\mathbf{C}}$ nas equações (72) e (73) como:

$$\widehat{\mathbf{Z}}_{f_h} = -\left[(\mathbf{A}(\widehat{\phi})^{-1}(\widehat{\mathbf{C}} + \mathbf{B}(\widehat{\phi}))\right]^h \mathbf{Z}_{f_0} \tag{74}$$

$$\widehat{\mathbf{S}}_{f_h} = \sum_{k=1}^{h} \mathbf{D}_k(\widehat{\phi})\widehat{\Gamma}\mathbf{D}_k(\widehat{\phi})' \tag{75}$$

As previsões dos valores futuros $\mathbf{Z}_{f_h} = \{z_{t(n+h,m)}, m = 1, \ldots, S\}$, têm intervalos com $100(1 - \alpha)\%$ de confiança calculados por $\widehat{z}_{t(n+h,m)} \pm \epsilon(\alpha/2)\sqrt{\widehat{s}_m}$, sendo $\widehat{z}_{t(n+h,m)}$ a previsão de $z_{t(n+h,m)}$, calculada em (74) e $\widehat{s}_m$ é o m-ésimo elemento da diagonal da matriz $\widehat{\mathbf{S}}_{f_h}$ estimada em (75), que representa a variância da previsão $\widehat{z}_{t(n+h,m)}$.

6.2 Previsão com BCPAR(p_m)

Quando a transformação de Box-Cox foi usada para transformar a série original, então a previsão da série original é calculada em função da previsão da série transformada. Para calcular a previsão da série transformada, escrevemos a equação (70) para $\mathbf{Z}_{f_h}^{(\lambda)}$:

$$\mathbf{Z}_{f_h}^{(\lambda)} = -\left[(\mathbf{A}(\alpha)^{-1}(\mathbf{D} + \mathbf{B}(\alpha))\right]^h \mathbf{Z}_{f_0}^{(\lambda)} + \sum_{k=1}^{h} \mathbf{D}_k(\alpha)\epsilon_{f_k} \tag{76}$$

com a equação (76), a previsão para $\mathbf{Z}_{f_h}^{(\lambda)}$ é dada por:

$$\mathbf{E}(\mathbf{Z}_{f_h}^{(\lambda)}|\mathbf{Z}_{f_0}^{(\lambda)}) = -\left[(\mathbf{A}(\alpha)^{-1}(\mathbf{D}+\mathbf{B}(\alpha))\right]^h \mathbf{Z}_{f_0}^{(\lambda)} \tag{77}$$

com $\mathbf{D} = (d_1, \ldots, d_S)$ e $\alpha = \{\alpha_{m,1}, \ldots, \alpha_{m,S},\ m = 1, \ldots, S\}$. A variância de $\mathbf{Z}_{f_h}^{(\lambda)}$ é dada por:

$$\mathbf{V}(\mathbf{Z}_{f_h}^{(\lambda)}|\mathbf{Z}_{f_0}^{(\lambda)}) = \sum_{k=1}^{h} \mathbf{D}_k(\alpha)\Gamma\mathbf{D}_k(\alpha)' \tag{78}$$

Substituindo os parâmetros em (77) e (78) por suas estimativas, temos que:

$$\widehat{\mathbf{Z}}_{f_h}^{(\lambda)} = -\left[(\mathbf{A}(\widehat{\alpha})^{-1}(\widehat{\mathbf{D}}+\mathbf{B}(\widehat{\alpha}))\right]^h \mathbf{Z}_{f_0}^{(\lambda)} \tag{79}$$

$$\widehat{\mathbf{S}}_{f_h}^{(\lambda)} = \sum_{k=1}^{h} \mathbf{D}_k(\widehat{\alpha})\widehat{\Gamma}\mathbf{D}_k(\widehat{\alpha})' \tag{80}$$

Para prever os valores futuros da série original $\widehat{\mathbf{Z}}_{f_h}$, usa-se a transformação inversa de Box-Cox, $G(z_{t(r,m)}^{(\lambda)})$ dada em (22).

Vamos considerar $G(z_{t(r,m)}^{(\lambda)})$ com as seguintes propriedades: Se $\mathbf{Y} = (y_1, \ldots, y_m)$ é um vetor aleatório $(1 \times m)$, então $\mathbf{G}(\mathbf{Y}) = (G(y_1), \ldots, G(y_m))$ também é um vetor aleatório $(1 \times m)$ e o valor esperado $\mathbf{E}\{\mathbf{G}(\mathbf{Y})|\mathbf{Z}\} = (E\{G(y_1)|\mathbf{Z}\}, \ldots, E\{G(y_n)|\mathbf{Z}\})$.

Usando a transformação inversa $\mathbf{G}(\mathbf{Z}_{f_h}^{(\lambda)})$ podemos escrever:

$$\mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0}) = \mathbf{E}\{G(\mathbf{Z}_{f_h}^{(\lambda)})|\mathbf{Z}_{f_0}^{(\lambda)}\} \tag{81}$$

Uma previsão ingênua para $\mathbf{Z}_{f_h}$ é dada por:

$$\widehat{\mathbf{Z}}_{f_h}^{(N)} = G(\mathbf{E}\{\mathbf{Z}_{f_h}^{(\lambda)}|\mathbf{Z}_{f_0}^{(\lambda)}\}) \tag{82}$$

$$\widehat{\mathbf{Z}}_{f_h}^{(N)} = G(\widehat{\mathbf{Z}}_{f_h}^{(\lambda)}) \tag{83}$$

com $\widehat{\mathbf{Z}}_{f_h}^{(\lambda)}$ dada pela equação (79).

Essa previsão ingênua não é a previsão de mínimo erro quadrático médio MMSE e sim a previsão que minimiza a média do erro absoluto MMAE. Ou seja, $\widehat{\mathbf{Z}}_{f_h}^{(N)}$ é a mediana da função densidade de probabilidade condicional $p(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$. Isso ocorre porque, ao usar a transformação de Box-Cox, $\mathbf{Z}_{f_h}^{(\lambda)}|\mathbf{Z}_{f_0}^{(\lambda)}$na equação (76) têm distribuições simétrica (aproximadamente normal) e a média e a mediana de $p(\mathbf{Z}_{f_h}^{(\lambda)}|\mathbf{Z}_{f_0}^{(\lambda)})$ coincidem. Então o percentil 50% de $p(\mathbf{Z}_{f_h}^{(\lambda)}|\mathbf{Z}_{f_0}^{(\lambda)})$ (a média) é levado pela transformação inversa no percentil 50% de $p(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$ (a mediana). Quando $\lambda \neq 1$ a densidade de probabilidade condicional $p(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$ é assimétrica e a previsão ingênua $\widehat{\mathbf{Z}}_{f_h}^{(N)}$ não coincide com a previsão MMSE $\widehat{\mathbf{Z}}_{f_h}$. O cálculo da previsão MSSE para $\widehat{\mathbf{Z}}_{f_h}$ só pode ser feito de forma aproximada. Neste trabalho vamos calcular as previsões MMSE usando a aproximação de segunda ordem dada em (23) e (24). Assim denotando por $\widehat{\mathbf{Z}}_{f_h} = (\widehat{z}_{t(n+h,1)}, \widehat{z}_{t(n+h,2)}, \ldots, \widehat{z}_{t(n+h,S)})'$ e $\widehat{\mathbf{Z}}^{(\lambda)}_{f_h} = (\widehat{z}^{(\lambda)}_{t(n+h,1)}, \widehat{z}^{(\lambda)}_{t(n+h,2)}, \ldots, \widehat{z}^{(\lambda)}_{t(n+h,S)})'$ e $\widehat{s}^{(\lambda)}_m$ é o m-ésimo elemento da diagonal da matriz estimada $\widehat{\mathbf{S}}^{(\lambda)}_{f_h}$, dada em (80). Então a previsão MMSE é calculada aproximadamente por:

$$\widehat{z}_{t(n+h,1)} \approx G(\widehat{z}^{(\lambda)}_{t(n+h,m)}) \left[1 + \frac{1}{2}(1-\lambda)\left(\frac{\sqrt{\widehat{s}^{(\lambda)}_m}}{\lambda \widehat{z}^{(\lambda)}_{t(n+h,m)} + 1}\right)^2\right] \tag{84}$$

$$\begin{aligned} \widehat{s}_m \approx (\lambda \widehat{z}^{(\lambda)}_{t(n+h,m)} + 1)^{2/\lambda} \left(\frac{\sqrt{\widehat{s}^{(\lambda)}_m}}{\lambda \widehat{z}^{(\lambda)}_{t(n+h,m)} + 1}\right)^2 \times \\ \times \left[1 + \frac{1}{2}\left(\frac{(1-\lambda)\sqrt{\widehat{s}^{(\lambda)}_m}}{\lambda \widehat{z}^{(\lambda)}_{t(n+h,m)} + 1}\right)^2\right] \end{aligned} \tag{85}$$

O cálculo dos intervalos de confiança para as previsões dos valores futuros $\mathbf{Z}_{f_h} = \{z_{t(n+h,m)}, m = 1, \ldots, S\}$, pode ser feito de forma direta com a transformação inversa $G(\widehat{z}^{(\lambda)}_{t(n+h,m)} \pm \epsilon_{\alpha/2}\sqrt{\widehat{s}^{(\lambda)}_m})$.

7 Previsão Bayesiana

No contexto Bayesiano a previsão dos valores futuros $\mathbf{Z}_{f_h}$, assumindo uma função de perda quadrática é dada pela média da densidade preditiva $p(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$. A previsão $\mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$ é dada por:

$$\mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0}) = \int_{\mathbf{Z}_{f_h}} p(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0}) d\mathbf{Z}_{f_h} \tag{86}$$

A densidade preditiva $p(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$ pode ser calculada usando-se a densidade a posteriori dos parâmetros do modelo. Denotando de forma geral a densidade a posteriori por $p(\mathbf{\Phi}|\mathbf{Z}, \mathbf{Z}_0)$ (aqui $\mathbf{\Phi}$ representa todos os parâmetros do modelo PAR ou BCPAR utilizado para a série $\mathbf{Z}$), então a densidade preditiva é dada por:

$$p(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0}) = \int_{\mathbf{\Phi}} p(\mathbf{Z}_{f_h}|\mathbf{\Phi}, \mathbf{Z}) p(\mathbf{\Phi}|\mathbf{Z}, \mathbf{Z}_0) d\mathbf{\Phi} \tag{87}$$

Na equação (87) a integral $\int_{\mathbf{\Phi}}$ representa as múltiplas integrais no espaço dos parâmetros $\mathbf{\Phi}$ Com a suposição de que os valores futuros são gerados por um modelo PAR(p_m) ou BCPAR(q_m), então a densidade preditiva é condicionada somente nas ultimas observações $\mathbf{Z}_{f_0}$. Portanto, $p(\mathbf{Z}_{f_h}|\mathbf{\Phi}, \mathbf{Z}_{f_0})$ é a verossimilhança dos valores futuros $\mathbf{Z}_{f_h}$ condicionada ao conhecimento dos dados passados e dos parâmetros do modelo. Substituindo (87) em (86) e trocando a ordem de integração, a previsão é dada por:

$$\mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0}) = \int_{\mathbf{\Phi}} \mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{\Phi}, \mathbf{Z}_{f_0}) p(\mathbf{\Phi}|\mathbf{Z}, \mathbf{Z}_0) d\mathbf{\Phi} \tag{88}$$

Na equação (88) a previsão $\mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$ é a média de $\mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{\Phi}, \mathbf{Z}_{f_0})$ com relação a densidade a posteriori dos parâmetros do modelo utilizado que denotamos aqui por:

$$\mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0}) = \mathbf{E}_p\{\mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{\Phi}, \mathbf{Z}_{f_0})\} \tag{89}$$

Quando a transformação de Box-Cox é utilizada para transformar os dados, então a previsão da série original, calculada usando a transformação inversa $\mathbf{Z}_{f_h} = \mathbf{G}(\mathbf{Z}_{f_h}^{(\lambda)})$, é dada por:

$$\mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0}) = \mathbf{E}_p\{\mathbf{E}[\mathbf{G}(\mathbf{Z}_{f_h}^{(\lambda)})|\mathbf{\Phi}, \mathbf{Z}_{f_0}]\} \tag{90}$$

Uma vantagem da abordagem bayesiana é que a densidade preditiva (87) e a previsão (89) e (90) podem ser calculadas facilmente usando técnicas de Monte Carlo associada com simulação MCMC (Gibbs sampling e Metropolis-Hasting).

7.1 Previsão com PAR(p_m) Bayesiano

Para calcular a previsão para o ano $(n + h)$ usando as observações do último ano n, ou seja as previsões h-passos a frente com o modelo PAR(p_m), vamos usar que $\theta_m = (c_m, \phi_m)$, então $\Theta = [\theta_1, \ldots \theta_s]$ pode ser denotado por $\Theta = [\mathbf{C}, \phi]$. Da equação (70) vemos que $p(\mathbf{Z}_{f_h}|\Theta, \Gamma, \mathbf{Z}_{f_0})$ é uma distribuição normal com média e variância dadas por:

$$\mathbf{E}(\mathbf{Z}_{f_h}|\Theta, \Gamma, \mathbf{Z}_{f_0}) = -\left[(\mathbf{A}(\phi)^{-1}(\mathbf{C} + \mathbf{B}(\phi))\right]^h \mathbf{Z}_{f_0} \tag{91}$$

e

$$\mathbf{V}(\mathbf{Z}_{f_h}|\Theta, \Gamma, \mathbf{Z}_{f_0}) = \sum_{k=1}^{h} \mathbf{D}_k(\phi)\Gamma\mathbf{D}_k(\phi)' \tag{92}$$

A estimativa para $\mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$ pode ser calculada por simulação de Monte Carlo usando uma amostra $\{\mathbf{C}^{(j)}, \phi^{(j)}, \Gamma^{(j)}, j = 1, \ldots, M\}$ com $[(\mathbf{C}^{(j)}, \phi^{(j)}), \Gamma^{(j)}] = \{[(c_m^{(j)}, \phi_m^{(j)}), \tau_m^{(j)}], m = 1, \ldots, S\}$ gerada das densidades a posteriori $p_m(\theta_m, \tau_m|\mathbf{Z}_m, \mathbf{Z}_0)$, $m = 1, \ldots, S$, com $\theta_m = (c_m, \phi_m)$. A estimativa de Monte Carlo para $\mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$ é dada por:

$$\widehat{\mathbf{Z}}_{f_h} = -\frac{1}{M}\sum_{j=1}^{M}\left[(\mathbf{A}(\phi^{(j)})^{-1}(\mathbf{C}^{(j)} + \mathbf{B}(\phi^{(j)}))\right]^h \mathbf{Z}_{f_0} \tag{93}$$

A matriz de covariância de $\mathbf{Z}_{f_h}$ condicionada à $\mathbf{Z}_{f_0}$ é dada por:

$$\mathbf{V}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0}) = \mathbf{E}(\mathbf{Z}_{f_h}\mathbf{Z}_{f_h}'|\mathbf{Z}_{f_0}) - \mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})\mathbf{E}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})' \tag{94}$$

Considerando a relação:

$$\mathbf{E}(\mathbf{Z}_{f_h}\mathbf{Z}'_{f_h}|\Theta,\Gamma,\mathbf{Z}_{f_0}) = \mathbf{V}(\mathbf{Z}_{f_h}|\Theta,\Gamma,\mathbf{Z}_{f_0}) + \mathbf{E}(\mathbf{Z}_{f_h}|\Theta,\Gamma,\mathbf{Z}_{f_0})\mathbf{E}(\mathbf{Z}_{f_h}|\Theta,\Gamma,\mathbf{Z}_{f_0}) \quad (95)$$

e tomando-se o valor esperado com relação a $(\boldsymbol{\Theta},\boldsymbol{\Gamma})$ em (95), chega-se a relação:

$$\begin{aligned}\mathbf{E}(\mathbf{Z}_{f_h}\mathbf{Z}'_{f_h}|\mathbf{Z}_{f_0}) = &\mathbf{E}_p\{\mathbf{V}(\mathbf{Z}_{f_h}|\Theta,\Gamma,\mathbf{Z}_{f_0})\}+ \\ &+\mathbf{E}_p\{\mathbf{E}(\mathbf{Z}_{f_h}|\Theta,\Gamma,\mathbf{Z}_{f_0})\mathbf{E}(\mathbf{Z}_{f_h}|\Theta,\Gamma,\mathbf{Z}_{f_0})'\}\end{aligned} \quad (96)$$

Substituindo (96) em (94), a variância $\mathbf{V}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$ é dada por:

$$\begin{aligned}\mathbf{V}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0}) = &\mathbf{E}_p\{\mathbf{V}(\mathbf{Z}_{f_h}|\Theta,\Gamma,\mathbf{Z}_{f_0})\}+ \\ &+\mathbf{E}_p\{\mathbf{E}(\mathbf{Z}_{f_h}|\Theta,\Gamma,\mathbf{Z}_{f_0})\mathbf{E}(\mathbf{Z}_{f_h}|\Theta,\Gamma,\mathbf{Z}_{f_0})'\}+ \\ &-\mathbf{E}_p\{\mathbf{E}(\mathbf{Z}_{f_h}|\Theta,\Gamma,\mathbf{Z}_{f_0})\}\mathbf{E}_p\{\mathbf{E}(\mathbf{Z}_{f_h}|\Theta,\Gamma,\mathbf{Z}_{f_0})\}'\end{aligned} \quad (97)$$

O segundo termo em (97) é dado por:

$$\begin{aligned}\mathbf{M}_{f_h}(\mathbf{Z}_{f_h}|\Theta,\Gamma,\mathbf{Z}_{f_0}) = \mathbf{E}_p\Big\{&\left[(\mathbf{A}(\phi)^{-1}(\mathbf{C}+\mathbf{B}(\phi))\right]^h\mathbf{Z}_{f_0} \\ &\mathbf{Z}'_{f_0}\left[((\mathbf{C}'+\mathbf{B}(\phi)')\mathbf{A}'(\phi)^{-1}\right]^h\Big\}\end{aligned} \quad (98)$$

A estimativa de $\mathbf{V}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$ pode ser calculada também por simulação de Monte Carlo calculando-se as estimativas de cada um dos termos:

$$\widehat{\mathbf{V}}_{f_h}=\frac{1}{M}\sum_{j=1}^{M}\left(\sum_{k=1}^{h}\mathbf{D}_k(\phi^{(j)})\widehat{\Gamma}^{(j)}\mathbf{D}_k(\phi^{(j)})'\right) \quad (99)$$

$$\widehat{\mathbf{M}}_{f_h}=\frac{1}{M}\sum_{j=1}^{M}\mathbf{M}_{f_h}(\mathbf{Z}_{f_h}|\Theta^{(j)},\Gamma^{(j)},\mathbf{Z}_{f_0}) \quad (100)$$

e substituindo-se (99), (100) e (93) na equação (97). A estimativa de Monte Carlo para $\mathbf{V}(\mathbf{Z}_{f_h}|\mathbf{Z}_{f_0})$ é dada por:

$$\widehat{\mathbf{S}}_{f_h} = \widehat{\mathbf{V}}_{f_h} + \widehat{\mathbf{M}}_{f_h} - \widehat{\mathbf{Z}}_{f_h}\widehat{\mathbf{Z}}'_{f_h} \quad (101)$$

Os intervalos de credibilidade das previsões são calculados de forma direta por $G(\widehat{z}_{t(n+h,m)}(\alpha/2))$ e $G(\widehat{z}_{t(n+h,m)}(1-\alpha/2))$, onde $\widehat{z}_{t(n+h,m)}(p)$ é o p-percentil de $\widehat{z}_{t(n+h,m)}$.

7.2 Previsão com BCPAR(p_m) Bayesiano

No contexto Bayesiano, as previsões h-passos a frente para a série transformada, ou seja a previsão para o ano $(n+h)$, usando as observações do último ano n, são calculadas da mesma forma como foi feito para o modelo PAR(p_m), agora usando a notação $\mathbf{D} = (d_1, \ldots, d_S)$ e $\beta_m = [d_m, \alpha_m]$, então $\Psi = [\beta_1, \ldots, \beta_s]$ pode ser denotado por $\Psi = [\mathbf{D}, \alpha]$. O valor esperado e a variância da densidade condicional $p(\mathbf{Z}_{f_h}^{(\lambda)} | \Psi, \Gamma, \lambda, \mathbf{Z}_{f_0}^{(\lambda)})$ são

$$\mathbf{E}(\mathbf{Z}_{f_h}^{(\lambda)} | \Psi, \Gamma, \lambda, \mathbf{Z}_{f_0}^{(\lambda)}) = -\left[(\mathbf{A}(\alpha)^{-1}(\mathbf{D} + \mathbf{B}(\alpha))\right]^h \mathbf{Z}_{f_0}^{(\lambda)} \tag{102}$$

e

$$\mathbf{V}(\mathbf{Z}_{f_h}^{(\lambda)} | \Psi, \Gamma, \lambda, \mathbf{Z}_{f_0}^{(\lambda)}) = \sum_{k=1}^{h} \mathbf{D}_k(\alpha)\Gamma\mathbf{D}_k(\alpha)' \tag{103}$$

Considerar o termo:

$$\begin{aligned}\mathbf{M}_{f_h}^{(\lambda)}(\mathbf{Z}_{f_h}^{(\lambda)} | \Psi, \Gamma, \mathbf{Z}_{f_0}^{(\lambda)}) = \mathbf{E}_p \Big\{ & \left[(\mathbf{A}(\alpha)^{-1}(\mathbf{D} + \mathbf{B}(\alpha))\right]^h \mathbf{Z}_{f_0}^{(\lambda)} \\ & \mathbf{Z}_{f_0}^{(\lambda)\prime} \left[((\mathbf{D}' + \mathbf{B}(\alpha)')\mathbf{A}'(\alpha)^{-1}\right]^h \Big\}\end{aligned} \tag{104}$$

Procedendo da mesma forma que foi feito para calcular a previsão sem a transformação, usando uma amostra $\{\mathbf{D}^{(j)}, \alpha^{(j)}, \gamma^{(j)}, \lambda^{(j)}, j = 1, \ldots, M\}$ com $[(\mathbf{D}^{(j)}, \alpha^{(j)}), \Gamma^{(j)}, \lambda^{(j)}] = \{[(d_m^{(j)}, \alpha_m^{(j)}), \gamma_m^{(j)}, \lambda^{(j)}], m = 1, \ldots, S\}$ gerada das densidades a posteriori $p_m(\beta_m, \gamma_m, \lambda | \mathbf{Z}_m, \mathbf{Z}_0)$, $m = 1, \ldots, S$, com $\beta_m = (c_m, \alpha_m)$,podemos estimar a previsão da série transformada por:

$$\widehat{\mathbf{Z}}_{f_h}^{(\lambda)} = -\frac{1}{M} \sum_{j=1}^{M} \left[(\mathbf{A}(\alpha^{(j)})^{-1}(\mathbf{D}^{(j)} + \mathbf{B}(\alpha^{(j)}))\right]^h \mathbf{Z}_{f_0}^{(\lambda(j))} \tag{105}$$

A variância da previsão da série transformada é estimada calculando-se

$$\widehat{\mathbf{V}}_{f_h}^{(\lambda)} = \frac{1}{M} \sum_{j=1}^{M} \left(\sum_{k=1}^{h} \mathbf{D}_k(\alpha^{(j)})\widehat{\Gamma}^{(j)}\mathbf{D}_k(\alpha^{(j)})' \right) \tag{106}$$

$$\widehat{\mathbf{M}}_{f_h}^{(\lambda)} = \frac{1}{M} \sum_{j=1}^{M} \mathbf{M}_{f_h}(\mathbf{Z}_{f_h}^{(\lambda(j))} | \Theta^{(j)}, \Gamma^{(j)}, \mathbf{Z}_{f_0}^{(\lambda(j))}) \tag{107}$$

com (105), (106) e (107), a estimativa de Monte Carlo para $\mathbf{V}(\mathbf{Z}_{f_h}^{(\lambda)} | \mathbf{Z}_{f_0}^{(\lambda)})$ é dada por:

$$\widehat{\mathbf{S}}_{f_h}^{(\lambda)} = \widehat{\mathbf{V}}_{f_h}^{(\lambda)} + \widehat{\mathbf{M}}_{f_h}^{(\lambda)} - \widehat{\mathbf{Z}}_{f_h}^{(\lambda)}\widehat{\mathbf{Z}}_{f_h}^{(\lambda)\prime} \tag{108}$$

A previsão de MMAE na abordagem Bayesiana é calculada da mesma forma que na abordagem clássica, dada pela equação (83) $\widehat{\mathbf{Z}}_{f_h}^{(N)} = G(\widehat{\mathbf{Z}}_{f_h}^{(\lambda)})$, com o estimador de Monte Carlo $\widehat{\mathbf{Z}}_{f_h}^{(\lambda)}$ dado em (105). A previsão MMSE $\widehat{\mathbf{Z}}_{f_h}$ é calculada considerando a aproximação de segunda ordem (84) e (85), para cada termo da equação (105). Então denotando por

$$\widehat{\mathbf{Z}}_{f_h}^{(\lambda)}(j) = (\mathbf{A}(\alpha^{(j)})^{-1}(\mathbf{D}^{(j)} + \mathbf{B}(\alpha^{(j)}))^h \mathbf{Z}_{f_0}^{(\lambda^{(j)})} \tag{109}$$

$$\widehat{\mathbf{V}}_{f_h}^{(\lambda)}(j) = \sum_{k=1}^{h} \mathbf{D}_k(\alpha^{(j)})\widehat{\Gamma}^{(j)}\mathbf{D}_k(\alpha^{(j)})' \tag{110}$$

$$\widehat{\mathbf{M}}_{f_h}^{(\lambda)}(j) = \mathbf{M}_{f_h}(\mathbf{Z}_{f_h}^{(\lambda^{(j)})}|\Theta^{(j)}, \Gamma^{(j)}, \mathbf{Z}_{f_0}^{(\lambda^{(j)})}) \tag{111}$$

$$\widehat{\mathbf{S}}_{f_h}^{(\lambda)}(j) = \widehat{\mathbf{V}}_{f_h}^{(\lambda)}(j) + \widehat{\mathbf{M}}_{f_h}^{(\lambda)}(j) - \widehat{\mathbf{Z}}_{f_h}^{(\lambda)}\widehat{\mathbf{Z}}_{f_h}^{(\lambda)\prime} \tag{112}$$

os j-ésimos termos de (105), (106), (107) e (108), a aproximação de segunda ordem para $\widehat{\mathbf{Z}}_{f_h}^{(j)} = \{\widehat{z}_{t(n+h,m)}^{(j)},\ m = 1,\ldots,S\}$ em função de $\widehat{\mathbf{Z}}_{f_h}^{(\lambda)}(j) = \{\widehat{z}_{t(n+h,m)}^{(\lambda)}(j),\ m = 1,\ldots,S\}$ e $diag\{\widehat{\mathbf{S}}_{f_h}^{(\lambda)}(j)\} = \{\widehat{s}_m^{(\lambda)}(j),\ m = 1,\ldots,S\}$ é dada por:

$$\widehat{z}_{t(n+h,m)}^{(j)} \approx G(\widehat{z}_{t(n+h,m)}^{(\lambda)}(j)) \left[1 + \frac{1}{2}(1 - \lambda^{(j)}) \left(\frac{\sqrt{\widehat{s}_m^{(\lambda)}(j)}}{\lambda \widehat{z}_{t(n+h,m)}^{(\lambda)}(j) + 1}\right)^2\right] \tag{113}$$

$$\begin{aligned} \widehat{s}_m(j) \approx (\lambda^{(j)}\widehat{z}_{t(n+h,m)}^{(\lambda)}(j) + 1)^{2/\lambda^{(j)}} \left(\frac{\sqrt{\widehat{s}_m^{(\lambda)}(j)}}{\lambda^{(j)}\widehat{z}_{t(n+h,m)}^{(\lambda)}(j) + 1}\right)^2 \times \\ \times \left[1 + \frac{1}{2}\left(\frac{(1 - \lambda^{(j)})\sqrt{\widehat{s}_m^{(\lambda)}(j)}}{\lambda^{(j)}\widehat{z}_{t(n+h,m)}^{(\lambda)}(j) + 1}\right)^2\right] \end{aligned} \tag{114}$$

Usando a equação (90) as previsão MMSE dos valores futuros $\widehat{\mathbf{Z}}_{f_h} = \{\widehat{z}_{t(n+h,m)}, m = 1,\ldots,S\}$, da série original são calculadas por:

$$\widehat{z}_{t(n+h,m)} = \frac{1}{M}\sum_{j=1}^{M} \widehat{z}_{t(n+h,m)}^{(j)} \quad \text{para } m = 1,\ldots S \tag{115}$$

Os intervalos de credibilidade das previsões são calculados de forma direta por $G(\widehat{z}_{t(n+h,m)}^{(\lambda)}(\alpha/2))$ e $G(\widehat{z}_{t(n+h,m)}^{(\lambda)}(1-\alpha/2))$ onde $\widehat{z}_{t(n+h,m)}^{(\lambda)}(p)$ é o p-percentil $\widehat{z}_{t(n+h,m)}^{(\lambda)}$.

8 Conclusão

Este trabalho apresenta uma abordagem Bayesiana para os modelos auto-regressivos periódicos (PAR) e modelos PAR com a transformação de Box-Cox (BCPAR). A transformação é adotada visando transformar séries com distribuição assimétrica em uma série com distribuição aproximadamente normal e homocedástica. Esse trabalho mostra que, na abordagem Bayesiana, as previsões com modelos PAR e BC-PAR podem ser estimadas usando algoritmos de MCMC. Quando a transformação de Box-Cox é utilizada, a previsão de mínimo erro quadrático médio (MMSE) da série original é estimada, em função da previsão da série transformada, usando uma aproximação de segunda ordem. Um problema futuro em aberto ainda é a viabilidade computacional da abordagem Bayesiana para seleção de modelos.

A Apêndice: Aproximação de Segunda Ordem da Transformação Inversa de Boc-Cox

Considerando a série original $\{z_{t(r,m)}, r = r_m, \ldots, n,\ m = 1, \ldots, S\}$, com $z_{t(r,m)} > 0$, a séries transformada $\{z^{(\lambda)}_{t(r,m)}, r = r_m, \ldots, n,\ m = 1, \ldots, S\}$ é calculada por

$$z^{(\lambda)}_{t(r,m)} = \begin{cases} (z^{\lambda}_{t(r,m)} - 1)/\lambda & \lambda \neq 0 \\ \ln(z_{t(r,m)}) & \lambda = 0 \end{cases} \tag{116}$$

Para encontrar a relação entre a média e variância da série original (μ_m , σ^2_m) e a média e variância da série transformada ($\mu^{(\lambda)}_m$, $\sigma^{2(\lambda)}_m$) vamos considerar a função inversa da transformação de Box-Cox $z_{t(r,m)} = G(z^{(\lambda)}_{t(r,m)})$, dada por:

$$G(z^{(\lambda)}_{t(r,m)}) = \begin{cases} (\lambda z^{(\lambda)}_{t(r,m)} + 1)^{1/\lambda} & \lambda \neq 0 \\ \exp(z^{(\lambda)}_{t(r,m)}) & \lambda = 0 \end{cases} \tag{117}$$

Com uma aproximação de segunda ordem para $G(z^{(\lambda)}_{t(r,m)})$, temos

$$\begin{aligned} G(z^{(\lambda)}_{t(r,m)}) \approx G(\mu^{(\lambda)}_m) + G'(\mu^{(\lambda)}_{,m})(z^{(\lambda)}_{t(r,m)} - \mu^{(\lambda)}_m) + \\ + \frac{1}{2} G''(\mu^{(\lambda)}_{,m})(z^{(\lambda)}_{t(r,m)} - \mu^{(\lambda)}_m)^2 \end{aligned} \tag{118}$$

Tomando-se o valor esperando em (118), temos

$$E(z_{t(r,m)}) \approx G(\mu^{(\lambda)}_m) + \frac{1}{2} G''(\mu^{(\lambda)}_{,m})\sigma^{2(\lambda)}_m \tag{119}$$

Usando a relação (117) em (119), temos a aproximação de segunda ordem para a média μ_m dada por

$$\mu_m \approx G(\mu^{(\lambda)}_m) \left[1 + \frac{1}{2}\lambda(\lambda - 1) \left(\frac{\sigma^{(\lambda)}_m}{(\lambda \mu^{(\lambda)}_{,m} + 1)} \right)^2 \right] \tag{120}$$

Uma aproximação de segunda ordem para σ^2_m é calculada usando as equações (118) e (119) na seguinte equação:

$$E(z_{t(r,m)} - \mu_m)^2 \approx E\Big(G'(\mu^{(\lambda)}_{,m})(z^{(\lambda)}_{t(r,m)} - \mu^{(\lambda)}_m) + \tag{121}$$
$$+\frac{1}{2}G''(\mu^{(\lambda)}_{,m})(z^{(\lambda)}_{t(r,m)} - \mu^{(\lambda)}_m)^2 - \frac{1}{2}G''(\mu^{(\lambda)}_{,m})\sigma^{2(\lambda)}_m\Big)^2$$

Desenvolvendo-se a equação (121) chega-se facilmente na aproximação de segunda ordem para σ^2_m dada por

$$\sigma^2_m \approx \frac{(\lambda\mu^{(\lambda)}_{,m} + 1)^{2/\lambda}\sigma^{2(\lambda)}_m}{(\lambda\mu^{(\lambda)}_{,m} + 1)^2}\left[1 + \frac{(1-\lambda)^2\sigma^{2(\lambda)}_m}{2(\lambda\mu^{(\lambda)}_{,m} + 1)^2}\right] \tag{122}$$

Note que fazendo $\lambda = 1$ nos termos entre colchetes nas equações (120) e (122), temos as aproximações de primeira ordem para $\mu_m \approx G(\mu^{(\lambda)}_{,m})$ e $\sigma^2_m \approx (\lambda\mu^{(\lambda)}_{,m} + 1)^{2/\lambda}\sigma^{2(\lambda)}_m/(\lambda\mu^{(\lambda)}_{,m} + 1)^2$. Uma vez que a série transformada tem distribuição aproximadamente normal (essa é uma das propriedades da transformação de Box-Cox), então a média e a mediana da série transformada coincidem. Portanto $G(\mu^{(\lambda)}_{,m})$ é na verdade a mediana da série original. Portanto, a aproximação de primeira ordem $\mu_m \approx G(\mu^{(\lambda)}_{,m})$ só é precisa se a média e a mediana da série original $z_{t(r,m)}$ forem próximas o que geralmente não ocorre em séries com distribuição assimétrica.

Referências

[1] Akkaya, A., Tiku, M.L. (2001), "Estimating parameters in Autoregressive models in non-Normal situations: symmetric Innovations", Commun. Statist.- Theory Meth., 30, 3, 517-536.

[2] Anderson, P.L., Vecchia, A.V. (1993), "Asymptotic results for periodic autoregressive moving-avarege process", J of Time Series, vol.14, 1, 1-18.

[3] Barreto G.A., Andrade, M.G. (2004), "Robust Bayesian approach for AR(p) models applied to streamflow forecasting", The Journal of Applied Statistical Sciences, vol.13, (recently accepted for publication)

[4] Bartolini, P., Salas, J.D., Obeysekera, J.T.B. (1988), "Multivariate periodic ARMA(1,1) process", Water Resour. Res., vol.25, 8, 1237-1246.

[5] Basawa, I.V., Lund, R. (2001), "Large sample properties of parameter estimates for periodic ARMA models", J. of Time Series, vol.22, 6, 651-663.

[6] Boshnakov, G.N., (1996), "Recursive Computation of the parameters of periodic autoregressive moving-average process", J. of Time Series, vol.17, 4, 333-349.

[7] Box, G., Cox, D.,(1964), "An analysis of transformations", J. R. Statis. Soc., Serie B, 26, 211-252.

[8] Box, G.E.P, Jenkins, G.M. and Reinsel, G.C. (1994), "Time series analysis: forecasting and control", 3nd edition, Holden-Day.

[9] Breiman, L., Friedman, J.H., (1985), "Estimating optimal transformations for multiple regression and correlation (with discussion)", J. Am. Statist., 80, 580-619.

[10] Casella, G. and George, E. J. (1992), "Explaining the Gibbs Sampler", The American Statistician, **46**(3), 167–174.

[11] Chib, S. and Greenberg, E. (1995). "Understanding the Metropolis-Hasting algorithm", The American Statistician, **49**(4), 327–335.

[12] Gladysev, E.G. (1961), "Periodically correlated randon sequences", Sov. Math. Dokl, vol.2, 385-388.

[13] Granger W.J., Newbold, P. (1976), "Forecasting Transformed series", J. R. Statist. Soc., Serie B, 38, 189-203.

[14] Grunwald, G.K., Hyndman, R.J., Tedesco, L., Tweedie, R.L. (2000), "Non-Gaussian linear AR(1) models", Aust. N.Z.J. Stat., 42, 4, 479-495.

[15] Lund, R., Basawa, I.V. (2000), "Recurcive prediction and likelihood evaluation for periodic ARMA models", vol.21, 1, 75-93

[16] MecLeod, A.I., (1994), "Diagnostic checking of periodic autoregression models with application, J. of Time Series, vol.15, 2, 221-233

[17] Pagano, M., (1978), "On periodic and multiple autoregressions", The Annals of Statistics, vol.6, 6, 1310-1317.

[18] Pankratz, Alan (1987), "Forecastin of Power Transformed series", Journal of Forecasting, vol.6, 239-248.

[19] Parzen, E., Pagano, M. (1979), "A approach to modeling seasonally stationary time series", Journ. of Econometrics, 9, 137-153.

[20] Pericchi, L.R. (1981), "A Bayesian approach to transformations to normality", Biometrika, vol.68, 1, 35-43.

[21] Rasmussen, P.F., Salas, J.D., Fagherazzi, L., Rassan,J.C., Bobée, B. (1996), "Estimation and validation of vontemporaneous PARMA models for streamflow simulation", Water Resour. Res., vol.32, 10, 3151-3160.

[22] Salas, J.D., Boes, D.C., Smith, R.A. (1982), "Estimation of ARMA models with seasonal parameters", Water Resour. Res., vol.18, 4, 1006-1010.

[23] Schwarz, G. (1978), "Estimating the dimension of a model", The Annals of Statistics, **6**, 2, 461–464.

[24] Thomas, H.A., Fiering, M.B. (1962), "Mathematical synthesis of stream flow sequeces for the analysis of river basins by simulation. In Design of Water Resources (eds. A. Maass, M.M. Hufschmidt, R. Dorfman, H.A. Thomas, S.A. Marglin, G.M. Fair), Cambridge, MA, Harvardy University Press.

[25] Tibshirani, R. (1988), "Estimating optmal transformations for regression via additivity and variance stabilizations", J. Am. Statist., 83, 394-405.

[26] Troutman, B.M. (1979), "Some results in periodic autoregression", Biometrika, 66, 219-228.

[27] Thyer, Mark, Kuczera, G., Wang, Q.J. (2002), "Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation", J. Hydrol, 265, 246-257.

[28] Ula, T.A. (1993), "Forecasting of multivariate periodic autoregressive moving-avarege process", J. of Time Series, vol.14, 6, 645-657.

[29] Vecchia, A.V. (1985a), "Maximum likelihood estimation for periodic autoregressive moving average models", Technometrics, vol.27, 4, 375-384

[30] Vechia A.V., (1985b), "Periodic autoregressive moving avarege (PARMA) modeling with application to water resources", Water Resopurces Bull., 21, 721-730.

[31] Xia, Y., Tong, H., Li, W.K. and Zhu Li-Xing (2000), "On the estimation of an instantaneous transformation for time series", J.R.Statist. Soc., Serie B, 62, Part 2, 383-397

[32] Yu, Gwo-H., Wen, Wen-C. (1996), "A methodology for forecasting non-Gaussian hydrological time series", Stochastic Hydraulic'96, Tickle, Gouter,Xu, Wasimi & Bouchart (eds), Balkema, Rotterdan, 507-513.

NOTAS DO ICMC

SÉRIE ESTATÍSTICA

076/2004 COSTA, E.F.; VAL, J.B.R. – Generalized LQ problem for infinite jump linear systems and the minimal solution of algebraic Riccati equations.

075/2004 LOIBEL, S.; VAL, J.B.R.; ANDRADE, M.G. – Inference for the Richards Growth model using Box and Cox transformation and Bootstrap techniques.

074/2003 AOKI, R.; BOLFARINE, H.; ACHCAR, J.A.; LEÃO P. JR., D. – Bayesian analysis of a multivariate null intercept errors-in-variables regression model.

073/2002 LEÃO P. JR., D.; AOKI, R.; SILVA, G.F.- Statistical analysis of proficiency testing results.

072/2002 ANDRADE, MARINHO G.; FERREIRA, VALÉRIA A. MARTINS; OLIVEIRA, SANDRA C. – Abordagem bayesiana com MCMC versus máxima verossimilhança para modelos ARCH(p)

071/2002 LOUZADA-NETO, F.; TOMAZELLA, V.L.D.; ANDRADE, M.G. – Bayesian analysis for reccurent lifetime data with a non homogeneous Poisson process with a frailty term.

070/2002 LEÃO JR., D.; FRAGOS, M. D.; VAL, J. B. R. DO; ANDRADE, M. G. – Separable Hansdorff measurable Radon spaces

069/2002 AOKI, R.; ACHCAR, J.A.; BOLFARINR, H.; SINGER, J.M. – Bayesian analysis of paired data in mixture of normal distributions.

068/2001 LOIBEL, S.; VAL, J.B.R.; ANDRADE, M.G. – Comparação entre as abordagens
Bayesiana e clássica do problema de inferência com modelo de crescimento populacional de Richards.

067/2001 ACHCAR, J.A; BARBOSA, V.F. – Análise bayesiana para dados de sobrevivência com funções de risco em forma de U.