

Previsão Bayesiana de valores genéticos de touros por meio do modelo auto-regressivo para dados em painel

[*Bayesian forecasting of sires breeding values using autoregressive panel data model*]

F.F. Silva¹, T. Sáfiadi², J.A. Muniz², L.H. Aquino², G.B. Mourão³

¹Departamento de Informática - UFV

Av. P.H. Rolfs, s/n

36570-000 – Viçosa, MG

²Departamento de Ciências Exatas - UFLA – Lavras, MG

³Departamento de Ciências Exatas - ESALQ – Piracicaba, SP

RESUMO

Inferência Bayesiana do modelo auto-regressivo foi aplicado para dados em painel a dados reais de DEP's de touros da raça Nelore publicadas entre 2000 e 2005. Na análise foram consideradas a função de verossimilhança exata e a obtenção de distribuições preditivas de dados futuros. Verificou-se ser recomendável o agrupamento dos animais em grupos homogêneos de acordo com a acurácia. Constatou-se também, que em média, a eficiência de previsão dos valores de DEP's para um ano futuro é próxima de 80%.

Palavras-chave: gado Nelore, série temporal, método MCMC, DEP

ABSTRACT

A Bayesian inference of autoregressive panel data model was applied to real data of Nelore sires Expected Progenie Difference (EPD) during a five-year period (2000-2005). The exact likelihood function and predictive distributions of future observations were considered. The results indicated the importance of sires grouping in homogeneous groups according to accuracy and showed forecast efficiency for EPD values in a future year around 80%.

Keywords: Nelore cattle, time serie, MCMC method, EPD

INTRODUÇÃO

Os estudos de dados em estrutura de painel estão essencialmente orientados para estudar a heterogeneidade relativa a diferentes indivíduos que apresentam uma trajetória longitudinal. Esse fato permite identificar determinados aspectos, como tendências e sazonalidades, que são de difícil quantificação usando apenas modelos de regressão individuais.

Uma classe de modelos de regressão bastante retratada em estudos de dados em painel é a de séries temporais. Um modelo de série temporal é utilizado quando se dispõe de uma seqüência de dados obtidos em intervalos regulares de tempo

durante um período específico. Dentre esses modelos, destaca-se o auto-regressivo, pois ele se aplica a diversas situações práticas e geralmente apresenta boa qualidade de ajuste comparado aos modelos mais complexos. De acordo com Morettin e Tolo (2004), ao se estudarem esses modelos, primeiramente há necessidade de se compreender o comportamento da série, mediante a obtenção de estimativas para os parâmetros e, depois, realizar previsões fundamentadas nessas estimativas.

Em recentes estudos envolvendo análise de modelos de regressão e de séries temporais (Barreto e Andrade, 2004; Silva et al., 2005; Silva et al., 2006) a metodologia Bayesiana foi

utilizada com sucesso, pois sua característica de considerar todos os parâmetros como variáveis aleatórias, reduziu substancialmente o número de estimativas viesadas e possibilitou a utilização de um número menor de observações, pois os conceitos probabilísticos envolvidos independem do número de dados utilizados. Outro fato que merece destaque é que na metodologia Bayesiana a estimação por intervalo é realizada de forma direta, considerando a incerteza existente sobre todos os parâmetros simultaneamente, sendo, portanto, a estimação por intervalo mais precisa em relação àquela apresentada pela metodologia freqüentista, que, em geral, utiliza variâncias assintóticas e aproximadas (O'Hagan, 1994).

A inferência Bayesiana consiste de uma informação a priori, dos dados amostrais e do cálculo da densidade a posteriori dos parâmetros.

$$P(\theta | Y) = \frac{L(Y | \theta)P(\theta)}{\int L(Y | \theta)P(\theta)d\theta}, \quad (1)$$

sendo $Y = \{y_1, y_2, \dots, y_n\}$. O denominador, chamado de constante de integração, não depende de θ . Portanto, tem-se a expressão, $P(\theta | Y) \propto L(Y | \theta)P(\theta)$, que pode ser entendida como: posteriori \propto verossimilhança x priori.

A distribuição *a posteriori* de um parâmetro contém toda a informação probabilística a seu respeito. Assim, toda a inferência sobre o parâmetro é realizada por meio desta

$$P(\theta_i | Y) = \int_{\theta \neq \theta_i} P(\theta | Y)d\theta_{\theta \neq \theta_i} \quad (2)$$

A integração da distribuição conjunta a posteriori para a obtenção das distribuições marginais geralmente não é analítica, necessitando de algoritmos iterativos especializados como o Gibbs Sampler e o Metropolis-Hastings, os quais são denominados de algoritmos Markov Chain-Monte Carlo (MCMC). Portanto, para a utilização desses algoritmos, é necessário que se obtenha, a partir da distribuição a posteriori, um conjunto de distribuições chamadas de distribuições condicionais completas.

A informação a priori é dada pela densidade de probabilidade $P(\theta)$, que expressa o conhecimento do pesquisador sobre os parâmetros a serem estimados. Os dados y_1, \dots, y_n representam amostra aleatória de uma população com densidade f , e são utilizados na análise Bayesiana por meio da função de verossimilhança $L(y_1, \dots, y_n | \theta)$, que representa a distribuição conjunta desses dados.

Portanto, a partir do momento em que se opta por uma distribuição a priori, obtém-se a função de verossimilhança, torna-se possível, por meio do Teorema de Bayes, representado pela fórmula (1), obter a distribuição densidade a posteriori de θ , de forma que qualquer conclusão a seu respeito é realizada a partir desta distribuição.

distribuição. Segundo Rosa (1998), para se inferir com relação a qualquer elemento de θ , a distribuição a posteriori conjunta dos parâmetros, $P(\theta | Y)$ deve ser integrada em relação a todos os outros elementos que a constituem. Assim, se o interesse do pesquisador concentra-se sobre determinado conjunto de θ , por exemplo θ_1 , tem-se a necessidade da obtenção da distribuição $p(\theta_1 | Y)$, denominada de distribuição marginal, que é dada por:

Ao se utilizar o modelo auto-regressivo de ordem p , que apresenta termos de defasagem temporal, verifica-se uma redução na quantidade de observações que compõem a função de verossimilhança quando esta se apresenta condicionada às p primeiras observações (Box et al., 1994), uma vez que essas não são consideradas na análise. Ao se tratar de dados em painel, esse problema é ampliado, pois a condicionalidade é inerente a todos os indivíduos, o que acarreta uma perda considerável de informações. Uma maneira de

solucionar o problema é a designação de uma função de verossimilhança exata, ou incondicional, pois apesar de proporcionar uma maior complexidade ao processo de análise, permite considerar todas as observações (Dufour e Kiviet, 1998).

O objetivo central da análise Bayesiana de Séries Temporais é gerar mecanismos robustos de previsão. Uma observação futura é descrita por uma distribuição condicionada aos dados passados, denominada distribuição preditiva (Barreto e Andrade, 2004). Quando se dispõe da distribuição conjunta das previsões futuras pode-se facilmente obter a distribuição desses valores futuros. Por exemplo, pode-se estar interessado em realizar previsões a respeito de valores assumidos por uma determinada variável ao longo do ano de 2007, baseando-se em dados observados durante o ano de 2006.

Segundo Box et al. (1994) a grande utilização de modelos auto-regressivos é justificada pelas suas diversas aplicações em situações reais. Uma possível aplicação, considerando dados em painel, está relacionada com modelagem e previsão de valores genéticos, também denominados de diferenças esperadas nas progênies (DEP). Essa aplicação é possível porque as DEP's são calculadas e publicadas anualmente em sumários, e ao verificar que um mesmo touro tem seus valores genéticos calculados ao longo de vários anos, tem-se então uma série temporal. Ao se considerar vários

touros, pode-se assumir a caracterização de um arquivo de dados estruturado em forma de painel.

Ao se promover seleção, além de ser importante ter uma idéia do que se espera com o processo, ou seja, a predição do resultado da seleção, é importante também obter uma estimativa da segurança que se terá ao se adotar determinado procedimento (Carneiro et al., 2001). Essa segurança é definida pela acurácia da seleção, e em um sumário de touros, têm-se indivíduos cujas DEP's foram obtidas sob diferentes graus de acurácia, o que exige um agrupamento desses indivíduos em classes homogêneas para que essas possam ser consideradas as populações requeridas na análise de dados em painel.

O presente trabalho teve como objetivo apresentar a inferência Bayesiana do modelo auto-regressivo de ordem p , $AR(p)$, para dados em painel, aplicada a dados de DEP's calculadas durante os anos de 2000 a 2005 para a característica ganho de peso pós-desmama de touros da raça Nelore. Tal metodologia foi utilizada com o intuito de realizar previsões desses valores para anos futuros de acordo com o grau de acurácia das DEP's.

MATERIAL E MÉTODOS

O modelo auto-regressivo, denotado por $AR(p)$, em que p é o número de parâmetros considerados, é apresentado pela fórmula (3):

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t = \sum_{j=1}^p \phi_j Y_{t-j} + e_t \quad (3)$$

em que:

Y_t é o valor atual de um processo estocástico, cujos valores já assumidos no passado são dados por Y_{t-1} , Y_{t-2} , ..., Y_{t-p} ;
 $\phi_1, \phi_2, \dots, \phi_p$ são os parâmetros do modelo, denominados de parâmetros de auto-regressão;

e_t é o resíduo associado ao modelo, também denominado de ruído branco, $e_t \sim N(0, \sigma_e^2)$.

Na presença de dados em estrutura de painel, o modelo $AR(p)$ indicado em (3) continua apresentando as mesmas características, porém a ele deve ser acrescentado um índice i referente ao indivíduo, como em (4):

$$Y_{it} = \phi_{i1} Y_{i(t-1)} + \phi_{i2} Y_{i(t-2)} + \dots + \phi_{ip} Y_{i(t-p)} + e_{it} = \sum_{j=1}^p \phi_{ij} Y_{i(t-j)} + e_{it} \quad (4)$$

em que: $i=1,2,\dots,m$; $j=1,2,\dots,p$ e $t=1,2,\dots,n_i$.

De acordo com esta notação, têm-se m indivíduos, cada um com n_i observações

longitudinais, indicando que cada indivíduo i pode apresentar um número específico de observações. Neste estudo optou-se pela utilização do modelo AR(2), ou seja $p=2$, portanto o modelo contempla dois parâmetros por indivíduo. Em relação aos resíduos, assume-se que eles são amostras aleatórias provenientes

de uma mesma distribuição, portanto $e_{it} \stackrel{iid}{\sim} N(0, \sigma_e^2)$.

A função de verossimilhança exata, de acordo com o modelo apresentado em (4), é a seguinte:

em que:

$$L(Y | \Phi, \sigma_e^2) \propto \Psi(\Phi, \sigma_e^2 | Y_p)(\sigma_e^2)^{-m\left(\frac{n-p}{2}\right)} \exp\left\{-\frac{1}{2\sigma_e^2}(Y_1 - X\Phi)'(Y_1 - X\Phi)\right\}$$

$$\Psi(\Phi, \sigma_e^2 | Y_p) = (\sigma_e^2)^{-\frac{mp}{2}} |V_p|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_e^2} Y_p' V_p Y_p\right\},$$

$$Y_p = [y_{11}, y_{12}, \dots, y_{1p}, y_{21}, y_{22}, \dots, y_{2p}, \dots, y_{m1}, y_{m2}, \dots, y_{mp}]',$$

$$Y_1 = [y_{1p+1}, y_{1p+2}, \dots, y_{1n}, y_{2p+1}, y_{2p+2}, \dots, y_{2n}, \dots, y_{mp+1}, y_{mp+2}, \dots, y_{mn}]'$$

$$X = \begin{bmatrix} X_1 & 0 & 0 & 0 \\ 0 & X_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & X_m \end{bmatrix}_{(m(n-p) \times mp)}$$

$$X_i = \begin{bmatrix} y_{ip} & \dots & y_{i1} \\ y_{ip+1} & \dots & y_{i2} \\ \vdots & \vdots & \vdots \\ y_{in-1} & \dots & y_{in-p} \end{bmatrix}_{(n-p) \times p}$$

$$\Phi = [\phi_{11}, \phi_{12}, \dots, \phi_{1p}, \phi_{21}, \phi_{22}, \dots, \phi_{2p}, \dots, \phi_{m1}, \phi_{m2}, \dots, \phi_{mp}]' \in \mathbb{R}^{mp}.$$

Os vetores Φ e Y_1 , considerados nas expressões, apresentam, respectivamente, as dimensões $mp \times m(n-p)$ e $m(n-p) \times 1$.

Optou-se pela utilização de uma priori independente t-Student multivariada – Gama Inversa, respectivamente para os parâmetros Φ e σ_e^2 , a qual é dada por:

$$P(\Phi, \sigma_e^2) = P(\Phi)P(\sigma_e^2),$$

$$\Phi \sim t - Student(\mu, P) \text{ com } m(n-1) \text{ graus de liberdade e } \sigma_e^2 \sim GI(\alpha, \beta).$$

Depois de obtidas as distribuições conjuntas a posteriori, por meio do teorema de Bayes, $P(\Phi, \sigma_e^2 | Y) \propto L(Y | \Phi, \sigma_e^2)P(\Phi, \sigma_e^2)$, torna-se necessário apresentar as distribuições

condicionais completas a posteriori para Φ e σ_e^2 , sendo estas respectivamente:

$$P(\Phi | \sigma_e^2, Y) \propto \Psi(\Phi, \sigma_e^2 | Y_p) \times \text{Normal multivariada}(\hat{\Phi}, (X'X)^{-1}) \times$$

$$t - Student \text{ multivariada}(\mu, P^{-1})$$

$$\sigma_e^2 | \Phi, Y \sim \text{Gama Inversa} \left(\frac{mn + 2\alpha}{2}, \frac{1}{2} (Y_p' V_p Y_p) + \beta + \right.$$

$$\left. \frac{1}{2} (\Phi - \hat{\Phi})' (X'X) (\Phi - \hat{\Phi}) + (Y_1 - \hat{Y}_1)' (Y_1 - \hat{Y}_1) \right)$$

Nota-se que a distribuição condicional completa para o parâmetro σ_e^2 é dada por uma distribuição gama-inversa, ou seja, ela apresenta uma forma conhecida, portanto passível ao uso do algoritmo Gibbs Sampler. O mesmo não acontece para a distribuição condicional do parâmetro Φ , que não apresenta uma forma definida, requerendo nessa situação o uso do algoritmo Metropolis-Hastings.

Os algoritmos Gibbs Sampler e Metropolis-Hastings foram implementados matricialmente no software estatístico R¹. A função *mnormt* (*multivariate Normal and t-Student distributions*) foi utilizada para a geração de números aleatórios implícita nos algoritmos MCMC. Estes algoritmos foram implementados considerando uma cadeia de 20.000 iterações, de forma que as 10.000 primeiras foram eliminadas para reduzir o efeito dos valores iniciais (*burn-in*). Para avaliar suas convergências foram utilizados os critérios de Gelman e Rubin (1992) e Raftery e Lewis (1992) mediante o pacote Bayesian Output Analysis (*BOA*) do R.

Para realizar a previsão de um dado futuro, utilizou-se a teoria de distribuição preditiva descrita por Heckman e Leamer (2001), que é dada por:

$$Y_{(n+1)}^{(q)} | Y \sim N(X\Phi^{(q)}, \sigma_e^{2(q)}I),$$

em que: I é uma matriz quadrada de ordem $mp \times mp$. Assim, o conjunto de valores gerados para esta distribuição normal multivariada, provenientes de cada q iteração dos algoritmos Metropolis-Hastings e Gibbs Sampler, constituem a distribuição preditiva para um dado futuro, cuja estimativa, $\hat{P}(Y_{(n+1)} | Y)$, é representada pela média dessa distribuição.

Neste estudo foram utilizados dados cedidos pelo Grupo de Melhoramento Animal da Faculdade de Zootecnia e Engenharia de Alimentos da USP, GMA/FZEA, com sede em Pirassununga, SP. Estes correspondem aos sumários de touros Nelore compreendidos entre os anos de 2000 e 2005. Constam no arquivo dados de DEP's para a característica ganho de peso entre a desmama (205 dias de idade) e o sobreano (550 dias de

idade) de 117 touros, divididos em três grupos, conforme o valor da acurácia referente a última observação, isto é, aos dados de 2005. Os grupos foram divididos em: acurácia baixa (0 a 40%), com 31 touros, acurácia média (41 a 60%), com 63 touros e acurácia alta (acima de 60%), com a frequência de 23 touros. Considerou-se também a análise de todos os animais, sem a separação em grupos de acordo com a acurácia. Dessa forma, utilizaram-se, portanto, quatro arquivos de dados. Em todos eles, a última observação, referente ao ano de 2005, foi suprimida visando à comparação entre os valores estimados por meio da distribuição preditiva e os verdadeiros valores observados.

RESULTADOS E DISCUSSÃO

Um resumo dos resultados, expressos como porcentagem de significância dos parâmetros ϕ_1 e ϕ_2 , obtidos no ajuste do modelo AR(2) aos dados individuais de DEP's considerando cada estrutura de dados determinada pelos valores da acurácia é apresentado na Tab. 1.

Tabela 1. Porcentagem de significância dos parâmetros do modelo AR(2) ajustado às séries individuais de DEP's

Estrutura dos dados	Significância (%)	
	Parâmetro ϕ_1	Parâmetro ϕ_2
Acurácia baixa	43,47	43,47
Acurácia média	73,01	65,07
Acurácia alta	65,21	65,21
Geral	61,53	57,26

É possível inferir que as séries referentes à população de indivíduos com baixa acurácia apresentam um comportamento diferenciado das demais, pois a porcentagem de significância dos parâmetros foi menor. Nota-se também, que o parâmetro ϕ_2 apresentou porcentagem de significância um pouco menor que aquela obtida para o parâmetro ϕ_1 nas estruturas de dados referentes à acurácia média e geral, indicando que as séries de DEP's de alguns indivíduos não apresentam um comportamento auto-regressivo de segunda ordem, podendo essas serem descritas por uma ordem inferior, AR(1), ou ainda se apresentarem como um processo estacionário, que indica a ausência de autocorrelação.

¹R Development Core Team, 2006 <<http://www.R-project.org>>.

Para complementar estes resultados, foram realizadas análises das funções de autocorrelação e autocorrelação parcial para as séries dos indivíduos cujos parâmetros ϕ_1 e ϕ_2 não se mostraram significativos. Por meio destas verificou-se que na estrutura relacionada com baixa acurácia, a função de autocorrelação parcial apresentou lag's significativos para alguns indivíduos, indicando que modelos de ordem superior, AR(3), devem ser considerados em 38% dos casos. Além disso, também foram identificados indivíduos (62%) cujas séries já são caracterizadas como um processo estacionário. A variedade de processos identificados nesta estrutura de dados não foi observada nos conjuntos relacionados com a acurácia média e alta, pois o número de indivíduos cujas séries demonstraram seguir um modelo AR(3) foram, respectivamente, 23,8% e 11,1%, observando-se, portanto, uma quase totalidade de processos estacionários.

Essa discussão evidencia o fato de a acurácia baixa proporcionar maior variabilidade entre as séries, o que justifica a separação dos indivíduos em grupos. Porém, não foram observadas evidências a respeito de diferenças na variabilidade do comportamento de séries representadas pelas estruturas de acurácia média e alta, o que pode indicar uma possível união desses indivíduos em uma só estrutura.

O ponto mais marcante em relação aos resultados Tab. 1 está relacionado com a identificação de indivíduos cujas densidades preditivas serão apresentadas, pois, não há razão em discutir o processo de previsão de valores futuros de séries individuais que não apresentam um processo auto-regressivo de primeira e segunda ordem, uma vez que a densidade preditiva está fundamentada no modelo AR(2) para dados em

painel. Assim, serão discutidas sob o aspecto preditivo apenas séries cujos parâmetros mostraram-se significativos.

Nas Tab. 2, 3 e 4 são apresentados os verdadeiros valores da última observação (y_6), correspondentes as DEP's do ano de 2005, os quais não foram considerados na análise para possibilitar a comparação entre valores observados e estimados por via das densidades preditivas.

Os resultados permitem afirmar que a metodologia utilizada para realizar previsões de dados futuros individuais com base na obtenção das distribuições preditivas foi eficiente, uma vez que a porcentagem de intervalos de credibilidade que continham os verdadeiros valores das DEP's referentes ao ano de 2005 variou entre 77,8% e 85,7%. Estes resultados também conduzem a discussões relacionadas com as classes de acurácia consideradas, pois nota-se que a capacidade preditiva do modelo é menor quando a acurácia é baixa, 77,8%, e para as acurácias médias e altas, respectivamente 83,3% e 85,7%. Este fato também é explicado mediante estimativas obtidas para a variância do erro (σ_e^2) em cada uma das estruturas de dados consideradas, as quais são mostradas na Tab. 5.

Realmente nota-se que a variância do erro foi maior na estrutura de dados com baixa acurácia, e essa revelação apenas vem a confirmar discussões anteriores relacionadas com a maior variabilidade entre o comportamento das séries dos indivíduos considerados na amostra referente a esta classe de acurácia.

Tabela 2. Verdadeiros valores da última observação (y_6) de cada indivíduo (Ind.), suas estimativas (\hat{y}_6) e intervalos de credibilidade de 95% (Li: limite inferior e Ls: limite superior): estrutura de dados com acurácia baixa

Ind.	y_6	\hat{y}_6	Li	Ls	Ind.	y_6	\hat{y}_6	Li	Ls
6	8,78	9,60	7,11	12,02	20	8,41	9,40	5,72	12,18
8	8,71	10,85	5,34	15,35	21	9,05	8,44	6,82	12,32
9	8,48	10,89	4,88	16,09	22	8,40	8,97	3,21	12,25
10	3,28	9,39	4,42	10,57	23	9,15	12,77	10,07	16,28
12	3,61	4,03	-3,99	8,14	25	9,88	9,57	1,53	14,45
13	6,31	7,90	3,76	10,66	27	9,10	7,28	1,99	12,91
14	10,41	16,97	14,98	19,95	28	7,60	4,44	-1,61	6,53
15	5,62	4,38	-4,75	7,14	29	8,17	8,73	0,99	16,57
16	7,75	5,99	-2,05	11,02	31	2,97	0,68	-2,64	4,96

Tabela 3. Verdadeiros valores da última observação (y_6) de cada indivíduo (Ind.), suas estimativas (\hat{y}_6) e intervalos de credibilidade de 95% (Li: limite inferior e Ls: limite superior): estrutura de dados com acurácia média

Ind	y_6	\hat{y}_6	Li	Ls	Ind	y_6	\hat{y}_6	Li	Ls
2	8,49	9,87	-0,78	11,02	41	12,66	11,29	7,83	16,15
5	6,55	5,25	1,09	17,74	42	8,95	8,67	2,72	14,00
6	9,11	9,57	6,93	12,89	43	10,23	10,73	6,06	15,96
8	12,32	10,04	8,79	17,72	44	6,84	5,52	2,79	13,54
10	9,06	12,75	9,28	15,24	45	8,80	8,05	2,26	17,54
11	9,77	7,18	2,28	15,86	46	6,43	9,29	7,06	12,07
16	10,99	14,20	11,03	17,84	47	9,90	7,95	2,03	18,55
17	7,21	5,20	3,81	8,75	48	9,18	13,12	10,81	17,39
21	7,66	7,05	3,29	9,59	51	11,25	14,32	2,09	19,32
22	8,92	9,42	3,96	18,98	52	5,28	5,82	2,65	9,44
24	4,94	8,58	5,26	17,31	53	12,49	13,22	7,74	18,56
25	8,22	8,63	3,05	12,12	54	8,10	13,76	3,02	18,02
26	9,39	11,35	7,48	14,68	55	8,70	6,71	3,21	11,33
28	8,55	8,55	4,96	12,94	56	9,29	7,17	3,86	15,51
30	9,67	7,86	-0,07	13,36	57	9,61	8,61	5,68	12,93
31	12,87	12,78	5,52	18,96	58	9,91	11,51	6,13	14,66
33	10,06	15,86	11,44	20,13	59	9,00	10,50	4,88	16,46
35	14,95	12,01	6,47	18,32	60	9,65	10,79	2,92	18,97
37	8,36	10,21	2,66	17,91	61	10,73	10,31	6,20	15,72
38	8,58	7,01	2,21	18,01	62	5,22	7,19	0,02	12,68
40	10,86	10,44	6,10	15,40	63	9,59	15,15	11,04	21,07

Tabela 4. Verdadeiros valores da última observação (y_6) de cada indivíduo (Ind.), suas estimativas (\hat{y}_6) e intervalos de credibilidade de 95% (Li: limite inferior e Ls: limite superior): estrutura de dados com acurácia alta

Ind	y_6	\hat{y}_6	Li	Ls	Ind	y_6	\hat{y}_6	Li	Ls
1	9,20	10,38	3,10	12,01	12	14,10	14,24	2,39	23,83
3	9,06	10,94	7,12	12,08	14	10,84	14,06	6,40	16,90
6	10,55	17,46	12,30	23,07	15	8,38	6,69	1,40	13,95
7	10,03	17,62	11,99	25,09	17	9,71	10,82	4,12	15,62
8	12,55	11,25	2,22	25,13	19	13,91	13,64	4,05	23,69
9	9,12	7,51	3,19	16,31	20	11,41	12,55	2,98	19,16
10	10,85	12,93	4,90	18,65	22	11,79	12,91	4,06	17,92

Tabela 5. Estimativas da variância do erro ($\hat{\sigma}_e^2$), intervalos de credibilidade de 95% (Li: limite inferior e Ls: limite superior), fator de Gelman-Rubin ($\sqrt{\hat{R}}$) e burn-in dado pelo critério de Raftery e Lewis para cada estrutura de dados considerada

Estrutura dos dados	$\hat{\sigma}_e^2$	Li	Ls	$\sqrt{\hat{R}}$	Burn-in
Acurácia Baixa	84,32	53,52	128,32	0,9918	562
Acurácia Média	31,00	19,25	47,33	0,9994	320
Acurácia Alta	37,99	26,35	58,97	0,9999	128

CONCLUSÕES

A análise de dados de DEP's de touros Nelore, em sucessivos anos, resultou em boas estimativas para o valor de um dado futuro, obtidas a partir da distribuição preditiva. Esse processo

possibilitou a detecção de diferenças nos comportamentos de séries referentes a indivíduos pertencentes a grupos distintos de acurácia, indicando que este fator deve ser levado em consideração ao predizer DEP's em anos futuros. As diferenças relatadas não impõem barreiras

para a não adoção da metodologia em dados provenientes de animais com baixa acurácia, visto que a porcentagem de acerto na previsão de um dado futuro para essa classe foi próxima a 80%.

REFERÊNCIAS BIBLIOGRÁFICAS

- BARRETO, G.; ANDRADE, M.G. Robust Bayesian approach for AR(p) models applied to streamflow forecasting. *J. Appl. Stat. Sci.*, v.12, p.269-292, 2004.
- BOX, G.E.P.; JENKINS, G.M.; G.C. REINSEL (Eds). *Time series analysis: forecasting and control*. 3.ed. San Francisco: Holden-Day, 1994. 500p.
- CARNEIRO, A.P.S. ; TORRES, R.A. ; EUCLYDES, R.F. et al. Efeito da conexão dos dados sobre a acurácia dos testes de progênie e performance. *Rev. Bras. de Zootec.*, v.30, p.342-347, 2001.
- DUFOUR, J.M.; KIVIET, J.F. Exact inference methods for first-order autoregressive distributed lag models. *Econometrica*, v.66, p.79-104, 1998.
- GELMAN, A.; RUBIN, D.B. Inference from iterative simulation using multiple sequence. *Stat. Sci.*, v.7, p.457-511, 1992.
- HECKMAN, J.; LEAMER, E. (Eds). *Handbook of Econometrics*. Amsterdam: Elsevier, 2001. p.744.
- MORETTIN, P.A.; TOLOI, C.M.C. (Eds). *Análise de séries temporais*. São Paulo: Edgard Blucher, 2004. 535p.
- O'HAGAN, A. (Ed). *Kendall's advanced theory of statistics: Bayesian Inference*. New York: Edward Arnold Press, 1994. 402p.
- RAFTERY, A.E.; LEWIS, S. How many iterations in the Gibbs sampler? In: BERNARDO, J.M. (Ed). *Bayesian Statistics*. Oxford University, 1992. p.763-773.
- ROSA, G.J.M. *Análise Bayesiana de modelos mistos robustos via amostrador de Gibbs*. 1998. 57f. Tese (Doutorado) – Universidade de São Paulo, Piracicaba.
- SILVA, F.F.; MUNIZ, J.A.; AQUINO, L.H. et al. Abordagem Bayesiana da curva de lactação de cabras Saanen de primeira e segunda ordem de parto. *Pesq. Agropec. Brás.*, v.40, p.27-33, 2005.
- SILVA, N.M.A.; MUNIZ, J.A.; SILVA, F.F. et al. Estudo de parâmetros de crescimento de bezerros Nelore por meio de um modelo de regressão linear: uma abordagem Bayesiana. *Cien. Anim. Bras.*, v.7, p.57-65, 2006.