BAYESIAN COMPARISON OF RANDOM REGRESSION MODELS FOR TEST-DAY YIELDS IN DAIRY CATTLE

J. Jamrozik and L.R. Schaeffer

Centre for Genetic Improvement of Livestock, Department of Animal and Poultry Science, University of Guelph, Guelph, ON, Canada, N1G 2W1

INTRODUCTION

Several approaches can be used to compare models. The simplest method evaluates the overall goodness of fit of the model using residual variance estimates. The weakness of this approach is that it does not take into account the level of parameterization of the model. Deviance Information Criterion (**DIC**) (Spiegelhalter *et al*., 1988) is an asymptotic criterion that reflects both goodness of fit and degree of parameterization. The Bayes Factor (**BF**) (Kass and Raftery, 1995) is an indication how much more probable a model is than a competing one, assuming that the prior probabilities of the models are the same. BFs take into account all the assumptions made (prior distributions, number of parameters), they can be used to compare non-nested models, and they have an interpretation in terms of model posterior probability. Schwartz Criterion (**SC**) (Kass and Raftery, 1995) can be used to construct an approximation to the logarithm of the BF and it does not require specification of the prior distribution of the models. Many different functions can be used in random regression test-day (**TD**) models for dairy production traits. They include parametric lactation curves and orthogonal polynomials of different order. The selection of an appropriate function to model genetic and environmental lactation curves might have an impact on the accuracy and the stability of the genetic evaluation system. The objective of this study was to compare several TD models for milk yield in first lactation in terms of overall goodness of fit (RES), asymptotic DIC, BF and approximate of it (SC). Bayesian methods with Gibbs sampling were used to calculate comparison statistics. Models differed in functions representing cow's lactation curve under linear hierarchical TD model specification (Jamrozik *et al*., 2001).

MATERIAL AND METHODS

Data were first-lactation test day milk yields of Canadian Holsteins (Jamrozik *et al*., 2002). Cows calved between January 1st, 1990 and December 31st, 1997 were included. Small subset of the data including 120,446 records on 14,798 cows in 150 herds was selected for the purpose of this study. Total number of animals was 25,087, and the number of herd-test day subclasses was 6107. Nine classes of season of calving (two seasons) by age at calving (four groups) were defined. Days in milk on TD, from 5 to 305, inclusive, were divided into 29 DIM classes: $5 – 20, 21 – 30, …, 280 – 290, 291 – 305$.

Models. The overall single-trait random regression model was a two-stage linear hierarchical model that can be written as:

$\mathbf{y}_i = \mathbf{H}_i\,\mathbf{h} + \mathbf{X}_i\,\mathbf{b} + \mathbf{f}\,(\,\mathbf{w}_i\,,\,\mathbf{t}_i\,) + \boldsymbol{\varepsilon}_i\,,$

$\mathbf{w}_i = \mathbf{u}_i + \mathbf{e}_i\,,$ for i=1,2,…, N, with first and second stage conditional distributions as

$y_i \mid h, b, w_i, \gamma \sim N [H_i \, h + X_i \, b + f (w_i , t_i), R_i (\gamma)]$ and $w_i \mid u_i, E \sim N [u_i, E]$, respectively. y_i is a vector of n_i TD records of cow i (i=1,2,…,N) taken at known times t_i; h is a vector of herd-test date effect, b is a vector of age-season-days in milk classes, matrices H_i and X_i relate cow records to appropriate elements in h and b, respectively. $f (w_i, t_i)$ is an expected trajectory (lactation curve) for i-th cow, with cow-specific parameters w_i, ε_i is a vector of n_i residuals associated with y_i. The first stage covariance matrices $R_i (\gamma) = \sum^{+} I_{n_{ij}} \, \gamma_j$, for $\gamma = [\gamma_1, \gamma_{2,...}, \gamma_{29}]'$. In the second stage model, u_i is a vector of additive genetic effects of cow i on w_i , e_i represents a vector of second stage residuals with $e_i \mid E \sim N (0, E)$. Prior distributions for the parameters were:

$u \mid G \sim N(0, A \otimes G)$, where A is an additive genetic relationship matrix between individuals, and G is the additive genetic covariance matrix between elements of u_i,

$p (b) = $ const, $- b_{max} < - b_i < b_{max}$, $b_{max} = 10^3$,

$p (h) = $ const, $- h_{max} < - h_i < h_{max}$, $h_{max} = 10^3$,

$\gamma_k \mid v_k , s_k^2 \sim SIC [v_k, v_k s_k^2]$, k=1,2,…, 29,

$G \mid v_g , G_0 \sim IW [v_g, v_g G_0]$,

$E \mid v_e , E_0 \sim IW [v_e, v_e E_0]$,

where v_k and s_k^2 are parameters of independent inverted chi-square distributions, v_g (v_e) and G_0 (E_0) are hyper-parameters of the inverted Wishart distributions.

Five models with different functions for the first-stage trajectory were fitted and they were:

W – model with three-parameter Wilmink function (Wilmink, 1987) ,

AS – model with-five parameter Ali and Schaeffer function (Ali and Schaeffer, 1987),

L4 – model with Legendre polynomials of order four (five coefficients for regressions),

L3 – model with Legendre polynomials of order three (four coefficients for regressions,

L2 – model with Legendre polynomials of order two (three coefficients for regressions). Orthogonal Legendre polynomials (Abramowitz and Stegun, 1965) were scaled so that covariate for the constant term was equal to 1.

Model comparisons.
BF was defined as in Kass and Raftery (1995). Suppose, for simplicity, that two models (M_0 and M_1) are compared. Let the parameters for the model M_k be θ_k. Then the BF to contrast model M_0 against model M_1 is $B_{01} = p(y \mid M = M_0) / p(y \mid M = M_1)$, where $p(y \mid M = M_k)$ =

$\int p(y \mid \theta_k, M = M_k) \, p(\theta_k \mid M = M_k) \, d\theta_k$ is an integrated (marginal) likelihood, and $p(\theta_k \mid M = M_k)$ is the prior density under model k. If the prior probabilities of each model being true are equal, then the BF gives the ratio between the posterior probabilities of each pair of models. According to Jeffreys (1961), B_{01} value greater than 150 (or log of B_{01} greater than 10) indicates very strong evidence in favor of model M_0. Posterior probability of a given model can be easily calculated using marginal likelihoods of competing models.

An approximation to the logarithm of the BF can be computed using SC. For model M_k, it was defined as $SC_k = \log p(y \mid \hat{\theta}_k) - 0.5(d_k) \log (n)$, where $\hat{\theta}_k$ is the maximum likelihood estimate of θ_k, d_k is the dimension of θ_k and n is the number of observations. As n goes to infinity,

[$(SC_1 - SC_0) - \log (B_{10})$] / B_{10} → 0, which means that difference in SC for two models approximates BF. Even for a very large sample size, however, this approximation does not give a correct value of BF. This method requires also specification of the number of parameters of each model. On the other hand, SC does not require specification of prior distribution of model parameters. In general, a bigger value of SC indicates a better fitting model. More information on SC model comparisons can be found in Kass and Raftery (1995).

DIC was calculated as $\overline{D} + p_D$, where $\overline{D} = E_{\theta|y} [D(\theta_k)]$ is a measure of the fit of the model, and $D(\theta_k) = -2 \log p(y | \theta_k)$ (deviance). p_D measures the complexity of the model (number of effective parameters) and it is expressed as $\overline{D} - D(\overline{\theta}_k)$, where $\overline{\theta}_k$ is the posterior mean of θ_k. Smaller values of DIC are associated with better fitting models. The asymptotic properties of DIC method require that the number of observations grow at the same rate as number of parameters of the model (Spiegehalter *et al.*, 1998). This might not be the case for animal models in general, but TD models with up to ten records per cow seem to follow this assumption.

Residual variances for all models were assumed heterogeneous in 29 DIM intervals. RES method of model comparison used weighted (by number of days in interval) average of estimates of residual variances. Gibbs sampling was used to draw from posterior distribution of parameters of the models. Posterior means were estimated using 100,000 samples (after 10,000 burn-in iterations). Log- likelihoods were stored for each Gibbs cycle and maximum values of likelihoods were determined after iterations. The marginal density $p(y | M = M_k)$ was estimated by the harmonic mean of the likelihood values from the Gibbs chain as $p(y | M = M_k)= \{1/m \Sigma [p(y | \theta_k^i, M_k)]^{-1}\}^{-1}$, where θ_k^i (i =1, 2, …, m) are draws obtained from the posterior distribution under model k. Number of parameters of the model for SC comparison was defined as number of levels of the first-stage hierarchical model plus number of residual variance components.

RESULTS AND DISCUSSION

Table 1 shows estimates of residual variances, logarithms of marginal likelihoods, and values of SC and DIC criterions for compared models. Logarithms of BF were greater than 10 for all possible pairs of compared models. This gave very strong evidence in favor of the model with the larger value of log of the marginal likelihood. The magnitude of marginal likelihood estimates indicated that for a given pair of models, the posterior probability of selecting the right model is nearly 1. Approximate BF calculated from SC were different from exact BF values. Rankings of models based on different criteria are shown in Table 2. DIC and RES methods gave identical ranking of compared models. BF and SC differed in rankings of AS – L3 and L2 – W pairs. This again indicated poor performance of SC as BF approximation for the analyzed models. Model **L5** was the most plausible specification by all methods, followed by **L4** and **AS**. Model with Wilmink function (**W**) showed the poorest performance compared with other models. Similar results were reported by Lopez-Romero *et al.* (2001) who compared different test-day models using Spanish Holstein data. All comparison criteria were relatively easy to calculate under Gibbs scheme. Numerical instability of the harmonic mean estimator of marginal likelihood was not an issue for the data and the models used. Estimates of marginal likelihood, SC and DIC for model L4 were virtually the same when either 100,000 or 1,000,000 samples were used.

Table 1. Estimates of residual variance (RES), marginal likelihood (log p(y | M=M$_i$)), Schwartz Criterion (SC) and Deviance Information Criterion (DIC), by model

Model	RES	log p(y | M=M$_i$)	SC	DIC
L4	4.16	-257,931	-254,106	802,909
AS	4.46	-261,542	-259,110	819,888
L3	4.44	-260,834	-257,880	816,029
L2	4.83	-264,535	-262,595	832,695
W	4.74	-265,006	-261,689	832,396

Table 2. Ranking of models by residual variance (RES), Bayes Factor (BF), Schwartz Criterion (SC) and Deviance Information Criterion (DIC)

Model	RES	BF	SC	DIC
L4	1	1	1	1
AS	3	3	2	3
L3	2	2	3	2
L2	5	4	5	5
W	4	5	4	4

CONCLUSION

Random regression TD models with different functions used as regressions showed very strong differences in terms of BF. Different comparison methods ranked models differently. BF indicated that TD models with Legendre polynomials outperformed models with lactation curves with the same number of parameters.

REFERENCES

Abramowitz, M. and Stegun, I.A. (1965) "Handbook of Mathematical Functions". Dover, New York.
Ali, T.E. and Schaeffer, L.R. (1987) *Can. J. Anim. Sci.* **67** : 637-644.
Jamrozik, J., Gianola, D. and Schaeffer, L.R. (2001) *Livest. Prod. Sci.* **71** : 223-240.
Jamrozik, J., Schaeffer, L.R. and Weigel, K.A. (2002) *J. Dairy Sci.* (Submitted).
Jeffreys, H. (1961) "Theory of probability". Oxford University Press, New York.
Kass, R.E. and Raftery, A.E. (1995) *J. Amer. Stat. Assoc.* **90** : 773-795.
Lopez-Romero, P., Rekaya, R., Chang, J.M., Gianola, D. and Carabano, M.J. (2001*) J. Anim. Sci.*, **79**, Suppl.1 / *J. Dairy Sci.*, **84**, Suppl. 1 / *Poult. Sci.*, **80**, Suppl. 1 / 54th Annu. Rec. Meat Conf., vol. II, 112.
Spiegelhalter, D.J., Best, N.G. and Carlin, B.P. (1998) Technical Report, Medical Research Council, Biostatistics Unit, Cambridge, UK.
Wilmink, J.B.M. (1987) *Livest. Prod. Sci.* **16** : 335-348.