

Linear and logistic models for multiple-breed genetic analysis of heifer fertility in Mexican Simmental–Simbrah beef cattle

J. J. Baeza-Rodríguez , M. Montaña-Bermúdez, V. E. Vega-Murillo & M. E. Arechavaleta-Velasco

To cite this article: J. J. Baeza-Rodríguez , M. Montaña-Bermúdez, V. E. Vega-Murillo & M. E. Arechavaleta-Velasco (2017): Linear and logistic models for multiple-breed genetic analysis of heifer fertility in Mexican Simmental–Simbrah beef cattle, Journal of Applied Animal Research, DOI: [10.1080/09712119.2017.1357559](https://doi.org/10.1080/09712119.2017.1357559)

To link to this article: <http://dx.doi.org/10.1080/09712119.2017.1357559>



© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 07 Aug 2017.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Linear and logistic models for multiple-breed genetic analysis of heifer fertility in Mexican Simmental–Simbrah beef cattle

J. J. Baeza-Rodríguez ^a, M. Montaño-Bermúdez^b, V. E. Vega-Murillo^c and M. E. Arechavaleta-Velasco^b

^aCampo Experimental Mochochá, Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias, Mochochá, Mexico; ^bCentro Nacional de Investigación Disciplinaria en Fisiología y Mejoramiento Animal, Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias, Ajuchitlán, Mexico; ^cCampo Experimental La Posta, Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias, Paso del Toro, Mexico

ABSTRACT

The aim of this study was to compare alternative models for the genetic evaluation of heifer fertility in Simmental–Simbrah cattle. The analyses were conducted using a database with 37,390 female birth information recorded from 1984 to 2007, and 59,018 individuals in the pedigree. Three generalized mixed models were adjusted for a single trait in a multiracial population: linear animal, linear sire and logistic sire. The models were analysed by restricted maximum likelihood procedure with Average Information algorithm. Two strategies of cross-validation were carried out to evaluate the predict ability of the models. The heritability from linear animal and sire models had lower values than that estimated with the logistic model, 0.04 ± 0.00 , 0.05 ± 0.00 and 0.20 ± 0.03 , respectively. High Spearman and Kendall correlations were observed between the ranks of breeding values (BV) estimated from the linear and logistic sire models, 0.99 and 0.94, respectively. In contrast, these correlations were lower between the animal and sire models, 71% and 54%, respectively. The logistic sire model was the best estimating the BV with ancestors' information, while the linear animal model (LAM) was the best predicting with scattered information. In general, it was considered that best fit and prediction was produced with the LAM.

ARTICLE HISTORY

Received 16 May 2016
Accepted 14 July 2017

KEYWORDS

Heritability; cross-validation; binomial trait; genetic trend; model prediction ability

Introduction

Improving reproductive performance means greater benefit in productive efficiency than what can be obtained with growth traits (Dickerson 1970); Van Eenennaam (2013), using economic principles, reported that improvement in reproduction can be up to four times more important than improving traits of final products in cow-calf systems which sale calves at weaning.



The genetic improvement of beef cattle in Mexico has carried out by selecting in traits of growth and carcass quality; few beef cattle breeds' associations have included genetic evaluations of traits related to the sexual precocity and fertility in their improvement programs. These are complex characteristics that are not commonly used for selection purposes but have significant influence on the biological and economic efficiency of the cow-calf system. Progress in its implementation as a genetic performance traits has been slowed down for several reasons: the long time that it takes to obtain relevant information, the difficulty in measuring such traits, the low heritability of reproductive characteristics and statistical problems associated with the application of linear models to binomial traits (i.e. estimators outside the parametric ranges (0, 1), inefficient estimators due to the presence of heteroscedasticity, association between mean and variance, etc. and biased BV).

A management tactic commonly recommended to increase the production efficiency is culling females that do not get pregnant within the deadlines set in production systems

(Dziuk and Bellows 1983; Azzam and Azzam 1991; MacNeil and Vukasinovic 2011). Eliminating heifers that were not pregnant at an early age is equivalent to removing the herd subfertile females. According to a study conducted in Montana, USA, when a heifer fails to become pregnant in the first breeding season, calving percentage on her productive life will be approximately 55% (Selk 2015). Given the little inheritable nature of this trait (<10%) (Cammack et al. 2009), using it as improving factor does not guarantee genetic gain; the phenotype of a reproductive trait is not a good indicator of its genetic merit. Selecting for this performance, animals of any genetic potential would be eliminated, lower or higher. Expected progeny difference (EPD) is the best option to improve reproductive characteristics; this estimator gives a clear picture of the genetic status of any individual within the population.

Currently, herd fertility, heifer fertility and stayability are main traits that researchers are exploring. Heifer fertility is a trait included in the breeding objectives as an indicator of sexual maturity (Shiotsuki et al. 2009); comprises puberty, fertilization and the successful maintenance of pregnancy. The heifer's ability to calve at early age is regarded as a profitable performance indicator.

Heifer fertility is a dichotomous variable, it is distributed as a Bernoulli random variable; in this kind of variables there is dependence between mean and variance and the mean takes

CONTACT M. Montaño-Bermúdez  montano.moises@inifap.gob.mx  Centro Nacional de Investigación Disciplinaria en Fisiología y Mejoramiento Animal, Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias, K1 Carretera a Colón. Ajuchitlán, Colón, Querétaro CP 76280, Mexico

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

values in the range 0–1. Therefore, random effects, as the additive genetic, do not have a normal distribution as continuous variables do; this is an assumption in the standard procedures of estimating BV and its failure to comply could result in bias. Gianola (1982) published that variables exhibiting discontinuous distributions are analysed better if a continuous underlying distribution is postulated. One way to solve this problem is by using nonlinear models, as logistics, where logits would be within the limits 0 to 1, and the random effects have a normal distribution.

The aim of this study was to compare alternative models for the genetic evaluation of heifer fertility in Simmental–Simbrah cattle, allowing using as much information as possible and having less bias. For this purpose, goodness of fit, prediction ability and rank of BV of three models were compared; an animal linear model, a sire linear model and a sire logistic model.

Material and methods

Population of study

Information was provided by The Mexican Simmental–Simbrah Association; consisting of 37,390 useful records of heifers that were taken from 1984 to 2007. Table 1 shows the structure and descriptive statistics of the database.

Data editing

Basic depuration of data included eliminating individuals without parents' identification, coming from embryo transfer technique or females which did not have production or progeny records. Genetic analysis was conducted using a database of multibreed heifers and 59,018 individuals in the pedigree.

Trait definition

To define fertility (FERT), records were coded as '1' when heifer's first calving occurred before 1270 days old; otherwise, fertility was coded as '0'.

Statistical analysis

To prepare data for genetic analysis, preliminary linear models were fitted to identify important fixed effects for the expression

Table 1. Pedigree and data structure: mean, standard deviation (SD), range, minimum (Min) and maximum (Max) for heifer fertility of Simmental–Simbrah beef cattle.

Pedigree structure	Animal model	Sire model			
Number of animals in pedigree	59,018	40,820			
Number of animals with productive records	37,390	37,390			
Number of sires in pedigree	3430	3430			
Number of sires with progeny in the data	3430	3430			
Number of dams in pedigree	26,101	0			
Number of dams with progeny and own productive record	1524	1524			
Statistical average for both models	Mean	SD	Range	Min	Max
Calves/sire	10.81	17.64	313	1	314
Sires/year/herd	4.38	3.77	29	1	30
Active years/sire	4.08	3.13	17	1	18
Dam/sire	10.82	17.64	313	1	314
Calves/dam	1.42	0.78	8	1	9

of FERT (Ghorbani et al. 2013), with the GLM procedure of SAS (2010).

Models for univariate analyses

Three generalized mixed models were adjusted for the genetic evaluation of heifer fertility as a single trait, in a multiracial population of Simmental–Simbrah cattle: (1) linear animal model (LAM), (2) linear sire model (LSM) and (3) logistic sire model (BSM). The logistic animal model was not included because it did not converge; Jamrozik et al. (2012) and Abdullahpour et al. (2006) mentioned that in using animal models with threshold procedures it is frequent that convergence is not achieved, due to the frequency of extreme subclasses, in which all the observations are either zero or one (Misztal et al. 1989).

The fixed effects considered for all models were: contemporary group (CG) and covariates: age at calving of cow's dam, linear (MA) and quadratic (MA2), Simmental breed genes proportion (GP), heterozygosis (Het) and recombination loss (RL). CG was defined as an effect determined by the herd, year and season of heifer's birth. Herd identification was assigned taking into account the number designated to the heifer's owner at its birth and who remained as its owner at weaning. CGs with less than four records were eliminated. Four birth seasons were created: (1) January–March, (2) April–June, (3) July–September and (4) October–December. Age of the dam at the heifer birth was calculated in days.

The fixed effects GP, Het and RL were included in the model with the objective of performing the multiracial evaluation, in this way it was possible to include more animals, improve the evaluation of purebred individuals and simultaneously compare Simmental and Simbrah cattle. These effects were estimated for each animal as:

$$GP = \frac{\text{Sire}_i + \text{Dam}_i}{2},$$

$$\text{Het} = (\text{Sire}_i \times (1 - \text{Dam}_i)) + (\text{Dam}_i \times (1 - \text{Sire}_i)),$$

$$\text{RL} = (\text{Sire}_i \times (1 - \text{Sire}_i)) + (\text{Dam}_i \times (1 - \text{Dam}_i)),$$

in which Sire_i and Dam_i were the proportion of Simmental in the sire and dam, respectively.

For the generalized linear mixed models analysis of LAM and LSM, an identity link function was used to connect expected value of the random variable FERT with a linear function of explanatory variables; on the other hand, in the case of BSM the logit link function was utilized. For all models, the linear predictors were

$$\eta = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a},$$

where the conditional distribution for $y|a$ was binomial, with parameter π for probability of calving; and the $\eta = g(\pi)$ link functions; $\boldsymbol{\beta}$ was a $p \times 1$ vector for fixed effects; \mathbf{a} was a $q \times 1$ vector of direct genetic random effect (animal or sire, depending on the corresponding model); \mathbf{X} and \mathbf{Z} were incidence matrices that associate the data with the related effects. It was assumed that random effects followed a multivariate normal distribution. The mixed model normal equations were as follows:

$$\begin{bmatrix} \mathbf{X}'\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\mathbf{X} & \mathbf{X}'\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\mathbf{Z} \\ \mathbf{Z}'\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\mathbf{X} & \mathbf{Z}'\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{H}'\mathbf{R}^{-1}\mathbf{y}^* \\ \mathbf{Z}'\mathbf{H}'\mathbf{R}^{-1}\mathbf{y}^* \end{bmatrix},$$

where

$$H = \frac{\partial \mu}{\partial \eta_j} = \frac{\partial \frac{e^{\eta_j}}{1 + e^{\eta_j}}}{\partial \eta_j},$$

$$R = \text{var}(y|u)$$

and

$$y^* = y - \mu + H\eta.$$

Random effects' means were equal to zero and the structure of (co)variance used was

$$\begin{pmatrix} a \\ e \end{pmatrix} \sim N \left[0, \begin{pmatrix} \mathbf{A}\sigma_a^2 & 0 \\ 0 & \mathbf{I}_n\sigma_e^2 \end{pmatrix} \right]$$

where **A** is the additive numerator relationship matrix (59,018 × 59,018 order for the animal model and order of 40,820 × 40,820 for the sire models), σ_a^2 is the direct additive genetic variance, σ_e^2 is the residual variance and \mathbf{I}_n is the identity matrix with order equal as number of records (37,390). It was assumed that genetic effects, animal or sire, were independent of residuals in the model. The LAM and LSM were run with responses binomially distributed and with no adjustment to residual variance. For the logit function of logistic sire model, errors had a standard logistic distribution with θ mean and variance was set to $\pi^2/3$ (Guerra et al. 2006; Williams 2011).

All models were analysed with the ASREML software (Gilmour et al. 2009). Genetic and phenotypic parameters were estimated by restricted maximum likelihood using an Average Information (AI) algorithm. The convergence for maximization of the likelihood function was presumed when the REML log-likelihood last changes less than $0.002 \times$ iteration number and the individual variance parameter estimate changes less than 1%.

Estimates of heritability (h^2) for linear models were evaluated in the observable scale, while the logistic model was performed on the logistic scale. Formulas used for calculating h^2 will be described next for the LAM, LSM or BSM, respectively.

$$\widehat{h}_A^2 = \frac{\widehat{\sigma}_a^2}{\widehat{\sigma}_a^2 + \widehat{\sigma}_e^2}, \quad \widehat{h}_S^2 = \frac{4\widehat{\sigma}_a^2}{\widehat{\sigma}_a^2 + \widehat{\sigma}_e^2}, \quad \widehat{h}_L^2 = \frac{4\widehat{\sigma}_a^2}{\widehat{\sigma}_a^2 + \pi^2/3},$$

The h^2 was set to the same scale for comparison; estimators obtained from linear models were converted to the underlying scale using the formula (Van Vleck 1972; Roff 2001; Martinez et al. 2005; Guerra et al. 2006).

$$h_u^2 = \frac{h_o^2 p(1 - p)}{z^2},$$

where h_u^2 is the underlying scale heritability, h_o^2 is the observed scale heritability, p is the proportion of individuals in the population who calved and z is the standard normal curve ordinate at threshold point where it cuts an area equals to p .

Comparison of models

The models were compared based on their goodness of fit; two information criteria were used to estimate overall fit of each model, Akaike (AIC) and Bayesian (BIC); the following formulas

were used:

$$\text{AIC} = -2 \log(\mathbf{L}) + 2\mathbf{p},$$

$$\text{BIC} = -2 \log(\mathbf{L}) + \mathbf{p} \times \log(\mathbf{n}),$$

where **L** is the likelihood, **p** is the number of parameters in the model and **n** is the number of observations. Objective of the AIC was to identify the best model generated by data and the BIC was to find the best prediction model (Crawley 2002). The models with the lowest AIC or BIC were considered the best models.

Models were also compared through the animals' rank obtained on the basis of BV estimated with the complete data set; CORR procedure of SAS (2010) software was used to calculate two nonparametric measures of association, rank correlation coefficient of Spearman (CORS) and the tau-b of Kendall (CORK); both statistics were used to evaluate the existing correspondence among BV obtained from the three models.

The ability of models to predict BV was assessed by k-fold cross-validation approach; two strategies of cross-validation were conducted. In the first approach, all data set was divided into two groups, one with animals born before 2004 who formed the training group with approximately 75% of the records, and the remaining individuals formed the test group; it was a temporary strategy (generational) that evaluates how well each model predicts BV, when the values to predict were from young animals through the phenotype of their ancestors and relatives relationship information. In the second assay, the entire data set was randomly divided into four subgroups of equal size, for this, random sampling without replacement was applied; each of these subgroups served as test group while the other three came together to form the corresponding training group; it was used as a strategy to evaluate the predictive ability of the models when 25% of the information was not available. Afterwards, in both appraisals, the training group served to estimate all fixed and random effects; which they were used to predict observations in the testing group.

Two criteria were used to compare the predictive ability of the models; the mean-square error of prediction (MSEP) and the Pearson correlation (CORR). The MSEP for each iteration was calculated by the formula:

$$\text{MSEP}_i = \frac{1}{n} \sum_{i=1}^n (y_i - \widehat{y}_i)^2,$$

where y_i is the observed response, \widehat{y}_i is the predicted response and n is the number of data in the test subset. The \widehat{y} for individual i within a test group was derived as the sum of predicted effects over all effects estimated in the training set. In each training analysis, the data excluded one group to train on the remaining groups to estimate independent effects, which were then used to predict response of individuals from the omitted group (validation set). The MSEP for each model was the arithmetic mean of its k-MSEP_i to obtain a single result; model with the lowest MSEP value was the one with better predictive ability.

$$\overline{\text{MSEP}} = \frac{1}{k} \sum_{i=1}^k \text{MSEP}_i.$$

Pearson's correlation was used to estimate the existing association between the BV observed and those predicted for each subgroup in every model; test data served to obtain observed BV, and training data for predicted ones. Similarly, to MSEF, arithmetic mean of the CORR of k subgroups served as single value to compare the models.

$$\rho_{y,\hat{y}} = \frac{1}{n} \sum_{i=1}^n \frac{\text{cov}(y, \hat{y})}{\sigma_y \sigma_{\hat{y}}},$$

where $\text{cov}(y, \hat{y})$ is the estimated covariance between the observed BV, and those predicted, σ_y is the standard deviation of the observed BV, and $\sigma_{\hat{y}}$ is the standard deviation of predicted BV ones; n is the number of records in the test or training subgroup. Decision criterion was: the model with the highest correlation was the one who had better prediction ability.

Results and discussion

Parameter estimation and model fitting

The mean and the variance obtained for the fertility of the heifers were 0.271 and 0.198, respectively. Thirty percent of heifers calved before 1270 days old. Figure 1 shows the genetic trends of fertility in the Simmental and Simbrah

heifers based on EPD estimated with the LAM for heifers born between 1984 and 2007. No clear genetic trend for both breeds was observed.

Variance components estimated with different models could not be directly compared because they belong to different distributions. It is more relevant if the comparison is between ratios of two components, as it would be for h^2 (Abdollahi-Arpanahi et al. 2013), and is advisable to put them in the same scale (Dempster and Lerner 1950; Van Vleck and Gregory 1992).

The heritabilities calculated with the three models are presented in Table 2. Differences were found between h^2 obtained from each distribution: lower values were obtained from the linear models than with the logistic models. The estimated h^2 for LAM, LSM and BSM was 0.04, 0.05 and 0.20, respectively; the LAM and LSM estimates were converted to a linear scale. Similarly, Sun and Su (2010) reported that the heritabilities for reproductive traits estimated with logistic models were higher than those from linear models. The heritabilities' estimates obtained from the linear models in the observable scale were one-third of that obtained with BSM; when they were transformed to the underlying scale, this proportion was reduced to a quarter, all of them were less than 0.05. Guerra et al. (2006) found the same differences that we found in this study, between h^2 estimates from linear models and from

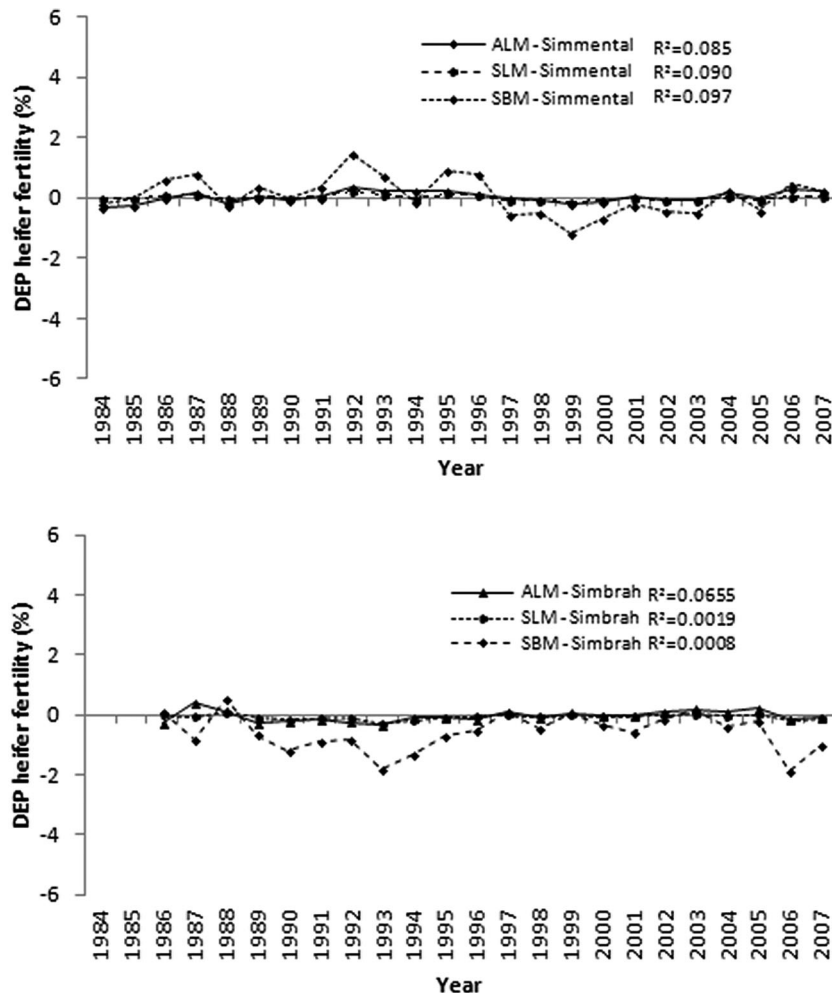


Figure 1. Genetic trends of heifer fertility based on expected progeny differences (EPD) using the animal linear (ALM), sire linear (SLM) and sire logistic (SBM) models for Simmental and Simbrah breeds.

Table 2. Variance components' estimators of restricted maximum likelihood and heritabilities of heifer's fertility for the models: linear animal (LAM), linear sire (LSM) and logistic sire (BSM).

	LAM	LSM	BSM
Additive variance	0.00948 (0.00136)	0.01211 ^a (0.00192)	0.69373 ^a (0.10699)
Error variance	0.12848 (0.00152)	0.13516 (0.00107)	3.2865 ^b
Phenotypic variance	0.13796 (0.00108)	0.13819 (0.00109)	3.4634 (0.02675)
Heritability	0.0687 (0.0097) ^c	0.0876 (0.0137) ^d	–
Underlying scale heritability	0.036 (0.0052) ^e	0.047 (0.0073) ^e	0.2003 (0.0293) ³
Heterogeneous variance factor (Deviance/DF)	–	–	0.79
Number of records	37,390	37,390	37,390
log-L	2797.61	2791.05	2834.35
AIC	84.33	100.33	116.33
BIC	68.36	68.36	68.33

Note: Models with lower values in AIC or BIC had best fit (bold font).

^a4(Sire variance).

^bError variance restricted to $\pi^2/3$.

^c(Additive variance)/(additive + error variance).

^d4(sire variance)/(sire variance + error variance).

^eTransformation: $h_b^2 = h_u^2(p(1-p)/Z^2)$, where h_u^2 is the h^2 in the underlying scale, h_b^2 is the h^2 in the binomial scale, p is the fraction of cows that calved and Z is the height to the ordered at the truncation point for a p area under the normal curve.

threshold models for multibreed beef cattle fertility; but h^2 of the LAM was higher than h^2 of sire models, indicating that h^2 estimated for these characteristics obtained with binomial data in the linear models was lower when it was converted to a linear scale, 0.07 vs. 0.04 for LAM and 0.09 vs. 0.05 for LSM; this may indicate that the additive variance estimator was confounded partially with nonadditive genetic variance (Gianola 1982). In a simulation study, Hoeschele and Tier (1995) determined that animal model produced a greater bias in the estimate of h^2 than the sire model; this in small progeny groups (<40), like in our study. Divergences in the h^2 were also found when the type of model was taking into account; h^2 obtained with the LAM was smaller than that estimated using a linear model but, as expected, the difference was lower between the underlying heritabilities. Similarly, in a simulation study, Van Vleck and Gregory (1992) found that h^2 was overestimated when it was calculated with binomial data by LAM, and then transformed to the underlying scale. An interesting point to consider in this study was the moderate h^2 obtained with the logistic model; h^2 calculated by the linear models and then transformed represented one quarter of h^2 estimated by the logistic model. As in this study, Doyle et al. (1996, 2000), Djemali et al. (1987), Buddenberg et al. (1989), Evans et al. (1999) and Snelling et al. (1996) found a moderate h^2 (> 0.20); likewise, the h^2 in the underlying scale of nonlinear models was greater than that calculated on the observed scale by linear models, and they attributed this difference to the lower prediction accuracy of the latter.

Goodness of fit

Table 2 shows the log-L values of the univariate analysis for each proposed model, and the information criteria (AIC and BIC). In our study, these two criteria were used to compare the models; this was possible because their likelihood functions were similar and because the same records were used; the results indicated that there were small differences among the adjustment of the three models.

The predictive ability of the three models was evaluated considering all the information of the database. The model that

showed greater MSEP was the LAM, followed by the BSM and LSM, with 11.12, 0.01 and 0.0001, respectively. CORR coefficients, obtained between the observed and adjusted values, indicated a different situation for the models, LAM had the best fit, followed by BSM and then LSM, with coefficients of 0.67, 0.63 and 0.62, respectively. Considering the information criteria (AIC and BIC), MSEP and CORR, the model with the best fit was LAM.

Prediction of random effects

From the point of view of genetic improvement, the differences in the rank of BV between models were relevant. The correlations CORS and CORK between predicted breeding values (PBV) obtained from the different models with all records available are presented in Table 3. A high correlation was observed between LSM and the BSM of 0.99 for CORS and of 0.94 for CORK, indicating that small differences between the rank of the PBV of the two sire models are expected. Similar results were obtained by Sun and Su (2010), they used the entire database in the genetic evaluation of success to first service, and they suggest that with a database large enough, the realignment in the ranks would not be significant. The correlation between the PBV obtained with the sire models and the generated with the animal model were in average 0.71 for CORS and 0.54 for CORK. The difference between the ranges of PBV was greater when it was obtained from the comparison between the animal model and any of the two sire models, than when this was the result of the contrast made between the sire models, even though the latter have different linking functions. Therefore, it was considered that the rank of PBV was affected more meaningfully by the relationship information than by

Table 3. Correlations of Spearman rank (above diagonal) and Kendall Tau-b (below diagonal) between the predicted breeding values obtained from the models: linear animal, linear sire and logistic sire for heifer fertility.

Model	Linear animal	Linear sire	Logistic sire
Linear animal		0.71	0.71
Linear sire	0.54		0.99
Logistic sire	0.54	0.94	

the type of distribution used in the estimation. Ramirez Valverde et al. (2001) found that when reliability of the sires changes, the rank of PBV also changed and, in their case, as the information was increased, the CORS of LAM gradually improved to the point that was similar to the nonlinear sire model. Contrary to what is stated here, the results obtained by Vazquez et al. (2009) suggest that the model type, linear or nonlinear, has an impact on the arrangement of PBV and therefore, correlations may change.

Evaluation of predictive ability with cross-validation

The results of the predictive ability of procedure 1 are shown in Table 4; and of procedure 2 are in Table 5. Using the MSEP as evaluation criterion, it was observed that the best fit was obtained with the BSM, secondly the LSM and finally LAM with 0.02, 3.19 and 27.91, respectively. With regard to the Pearson correlation for evaluating model fit, it was found that fit's differences between models were small, two percentage points in CORS, and the BSM had the best prediction according to the CORK indices, followed by LSM and then by LAM with 0.35, 0.26 and 0.23, respectively. Unlike what was observed in the assessment of the adjustment using all data set, judging with SMPE and CORR criteria, the BSM performed the best prediction of the BV with young animals and had the best association between PBV's ranks of the subpopulations of training and test, according to nonparametric rank correlations.

As in procedure 1, MSEP and CORR were used in procedure 2 to investigate the predictive ability of the models, and CORS and CORK to compare the association between the dispersion of the PBV. LSM had the best estimates of the association between the PBV predicted with all information and those predicted with the reduced information for the three models; we consider that this model made the best prediction of the PBV when a quarter of the information was not included in the evaluation, even when LAM had greater value of MSEP than LSM.

The results are shown in Tables 4 and 5. Although the greatest MSEP obtained was for LAM, we considered that it was the model that best predicted the PBV when the information was reduced; this model also had the best estimates of association between PBV's ranks. Correlation coefficients estimated for LSM and BSM were very similar and 29% lower than those obtained with LAM. The MSEP obtained from the assessments made on the four subpopulations of each linear model, LAM

Table 4. Cross-validation by using different prediction criteria for all tested models of heifer fertility in Simmental–Simbrah beef cattle (procedure 1).

	LAM ^a	LSM ^b	BSM ^c
Sub-sample size	12,475	12,475	12,475
MSEP ^d	27.91	3.19	0.02
CORR ^e	0.35	0.33	0.35
CORS ^f	0.33	0.35	0.35
CORK ^g	0.23	0.26	0.35

^aAnimal linear model.

^bSire linear model.

^cSire logistic model.

^dMean-square error of prediction.

^ePearson's correlation.

^fSpearman correlation.

^gKendall correlation.

Table 5. Cross-validation by using different prediction criteria for all tested models of heifer fertility in Simmental–Simbrah beef cattle (procedure 2).

	LAM ^a	LSM ^b	BSM ^c
<i>Sub-sample 1</i>			
N	5168	5168	5168
MSEP	3.62	0.74	81.06
CORR	0.72	0.54	0.56
CORS	0.68	0.50	0.52
CORK	0.51	0.36	0.37
<i>Sub-sample 2</i>			
N	5234	5234	5234
MSEP	10.8	1.70	83.90
CORR	0.74	0.55	0.55
CORS	0.72	0.52	0.52
CORK	0.54	0.38	0.38
<i>Sub-sample 3</i>			
N	5158	5158	5158
MSEP	6.64	1.56	130.94
CORR	0.73	0.51	0.52
CORS	0.70	0.47	0.47
CORK	0.52	0.35	0.34
<i>Sub-sample 4</i>			
N	5166	5166	5166
MSEP	7.09	2.75	248.55
CORR	0.77	0.61	0.59
CORS	0.74	0.56	0.57
CORK	0.55	0.41	0.41
<i>Sample mean</i>			
MSEP ^d	7.04	1.69	136.11
CORR ^e	0.74	0.55	0.56
CORS ^f	0.71	0.51	0.52
CORK ^g	0.53	0.38	0.38

^aAnimal linear model.

^bSire linear model.

^cSire logistic model.

^dMean-square error of prediction.

^ePearson's correlation.

^fSpearman correlation.

^gKendall correlation.

and LSM, was of small magnitude and constant within model. Contrary to the above, the MSEP for evaluations conducted with BSM was above than those obtained in the linear models, and very divergent between them, this probably caused by sampling. Opposite form what it was found, Vazquez et al. (2009) obtained similar MSEP for both sire models, linear and logistic. In their cross-validation test, Sun and Su (2010) estimated comparable CORS, but with slight advantage in stability for the logistic sire model over the LSM.

Like us, Ramirez-Valverde et al. (2001) evaluating calving difficulty found that increasing the number of records per sire improved the predictability of the animal model with respect to the nonlinear sire model, with the difference that in our case, LAM exceeded the predictability of BSM, while in their study, the performance of the models ended up being similar. Divergence between the two cross-validation procedures in this study lies primarily in the number of progeny per sire; although procedure 1 measured the contribution of recent records in the estimation of PBV, also has implicit a greater reduction of progeny per sire than it was present in procedure 2; the CORS of LAM in the latter procedure, doubled the estimate of the first, while in the case of BSM this increase was only 60%. These authors found a better predictive ability of the animal model over the sire model, as can be seen in the results obtained in procedure 2, where the animal model has better performance regardless of which of the two sire models was compared to.

Conclusions

The results showed that the h^2 of heifer fertility was better estimated with BSM than with the linear models, although this did not lead to a better prediction of PBV; the estimation of the PBV was not significantly affected by the distribution of the adjusted variable; on the other hand, the quality of information and connectivity impacted on the prediction and ranking of the PBV. Taking into account all the criteria used to compare the models, it can be concluded that the genetic evaluation of Simmental heifer fertility can be performed better with LAM, provided that the database is large enough; the interpretation of the logistic evaluation results is more complicated and, on the other hand, the number of animals with EPD was higher in the animal model than in the sire models.

Acknowledgment

The authors would like to thank the Mexican Simmental-Simbrah Association for providing data for this study.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

J. J. Baeza-Rodríguez  <http://orcid.org/0000-0001-9339-9695>

References

- Abdollahi-Arpanahi R, Peñagaricano F, Aliloo H, Ghiasi H, Urioste JI. 2013. Comparison of Poisson, probit and linear models for genetic analysis of number of inseminations to conception and success at first insemination in Iranian Holstein cows. *Livest Sci.* 153:20–26.
- Abdollahpour R, Moradi-Shahrehabak M, Mehrabani-Yeganeh H, Sayadnejad M, Eghbal A. 2006. Genetic analysis of dystocia in Holstein cattle of Iran by threshold and linear models. *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production*; Aug 13–18; Belo Horizonte, Minas Gerais, Brazil.
- Azzam SM, Azzam AM. 1991. A Markovian Decision Model for beef cattle replacement that considers spring and fall calving. *J Anim Sci.* 69:2329–2341.
- Buddenberg BJ, Brown CJ, Johnson ZB, Dunn JE, Peterson HP. 1989. Heritability estimates of pregnancy rate in beef cows under natural mating. *J Anim Sci.* 67:2589–2594.
- Cammack KM, Tomas MG, Enns RM. 2009. Reproductive traits and their heritabilities in beef cattle. *Prof Anim Sci.* 25:517–528.
- Crawley MJ. 2002. *Statistical computing: an introduction to data analysis using S-Plus*. 1st ed. Chichester, West Sussex: Wiley; p. 772.
- Dempster ED, Lerner IM. 1950. Heritability of threshold characters. *Genetics.* 35:212–236.
- Dickerson G. 1970. Efficiency of animal production, modeling the biological components. *J Anim Sci.* 30:849–859.
- Djemali M, Berger PJ, Freeman AE. 1987. Ordered categorical sire evaluation for dystocia in Holsteins. *J Dairy Sci.* 70:2374–2384.
- Doyle SP, Golden BL, Green RD, Brinks JS. 2000. Additive genetic parameter estimates for heifer pregnancy and subsequent reproduction in Angus females. *J Anim Sci.* 78:2091–2098.
- Doyle SP, Green RD, Golden BL, Mathiewis GL, Comstock CR, LeFever DG. 1996. Genetic parameter estimates for heifer pregnancy rate and subsequent re-breeding rate in Angus cattle. *J Anim Sci.* 74(Suppl.1):117. (Abstr.)
- Dziuk RA, Bellows RA. 1983. Management of reproduction of beef cattle, sheep and pigs. *J Anim Sci.* 57(Suppl. 2):355–379.
- Evans JL, Golden BL, Bourdon RM, Long KL. 1999. Additive genetic relationships between heifer pregnancy and scrotal circumference in Hereford cattle. *J Anim Sci.* 77:2621–2628.
- Ghorbani Sh, Tahmoorespur M, Maghsoudi A, Abdollahi-Arpanahi R. 2013. Estimates of (co)variance components for production and reproduction traits with different models in Fars native fowls. *Livest Sci.* 151:115–123.
- Gianola D. 1982. Theory and analysis of threshold characters. *J Anim Sci.* 54:1079–1096.
- Gilmour AR, Gogel BJ, Cullis, BR, Thompson R. 2009. *ASReml User Guide*, Release 3.0. VSN International Ltd; Hemel Hempstead, HP1 1ES, UK: <http://www.vsn.co.uk>.
- Guerra JLL, Franke DE, Blouin DC. 2006. Genetic parameters for calving rate and calf survival from linear, threshold, and logistic models in a multi-breed beef cattle population. *J Anim Sci.* 84:3197–3203.
- Hoeschele I, Tier B. 1995. Estimation of variance components of threshold characters by marginal posterior modes and means via Gibbs sampling. *Genet Sel Evol.* 27:519–540.
- Jamrozik J, McGrath S, Kemp RA, Miller SP. 2012. Genetic analysis of female fertility traits in Canadian Simmentals. *Livest Sci.* 150:302–309.
- MacNeil MD, Vukasinovic N. 2011. A prototype national cattle evaluation for sustained reproductive success in Hereford cattle. *J Anim Sci.* 89:1712–1718.
- Martinez GE, Koch RM, Cundiff LV, Gregory KE, Kachman SD, Van Vleck LD. 2005. Genetic parameters for stayability, stayability at calving, and stayability at weaning to specified ages for Hereford cows. *J Anim Sci.* 83:2033–2042.
- Misztal I, Gianola D, Foulley L. 1989. Computing aspects of a nonlinear method of sire evaluation for categorical data. *J Dairy Sci.* 72:1557–1568.
- Ramirez-Valverde R, Misztal I, Bertrand JK. 2001. Comparison of threshold vs linear and animal vs sire models for predicting direct and maternal genetic effects on calving difficulty in beef cattle. *J Anim Sci.* 79:333–338.
- Roff DA. 2001. The threshold model as a general purpose normalizing transformation. *Heredity.* 86:404–411.
- SAS. 2010. The data analysis for this paper was generated using SAS/STAT software, Version 9.3 for Windows. Copyright © 2002–2010 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.
- Selk G. 2015. 3 Advantages of culling open replacement heifers. *Progressive Cattleman*. [Published on 21 Apr 2015; updated 2016 Feb 19]. <http://www.progressivecattle.com/topics/reproduction/6794-3-advantages-of-culling-open-replacement-heifers>.
- Shiotsuki L, Silva JAIV, Tonhati H, Albuquerque LG. 2009. Genetic associations of sexual precocity with growth traits and visual scores of conformation finishing and muscling in Nelore cattle. *J Anim Sci.* 87:1591–1597.
- Snelling WM, MacNeil MD, Golden BL. 1996. Application of continuous and binary trait methods to reproductive measures of Hereford cattle. *J Anim Sci.* 74(Suppl. 1):115(Abstr.).
- Sun C, Su G. 2010. Comparison on models for genetic evaluation of non-return rate and success in first insemination of the Danish Holstein cows. *Livest Sci.* 127:205–210.
- Van Eenennaam LA. 2013. Considerations related to breed and biological type. *Vet Clin North Am Food Anim Pract.* 29:493–516.
- Van Vleck LD. 1972. Estimation of heritability of threshold characters. *J Anim Sci.* 55:218–225.
- Van Vleck LD, Gregory KE. 1992. Multiple-trait restricted maximum likelihood for simulated measures of ovulation rate with underlying multivariate normal distributions. *J Anim Sci.* 70:57–61.
- Vazquez AI, Gianola D, Bates D, Weigel KA, Heringstad B. 2009. Assessment of Poisson, logit, and linear models for genetic analysis of clinical mastitis in Norwegian Red cows. *J Dairy Sci.* 92:739–748.
- Williams R. 2011. Comparing Logit and Probit coefficients between models and across groups. <https://www3.nd.edu/~rwilliam/stats/Oglm.pdf>.