Bias in genetic evaluation using random regression test-day model

Martin Lidauer, Esa A. Mäntysaari

MTT Agrifood Research Finland, Biotechnology and Food Research, Biometrical Genetics,
FI-31600 Jokioinen, Finland, e-mail: martin.lidauer@mtt.fi

The effect of an upgraded Finnish evaluation model on bias in estimated breeding values for protein yield was investigated. Evaluations based on repeatability animal model and on random regression test-day model without and with heterogeneous variance adjustment were compared. Comparisons were based on the average difference between pedigree indices and the future estimated breeding values, based on own or on daughter performance records. This was defined as empirical bias. The pedigree indices were computed from reduced data sets where four years of the most recent data were excluded. Results showed an upward bias in the protein yield pedigree indices for Ayrshire young sires of 2.2 kg, 2.5 kg and 1.8 kg from the repeatability animal model, random regression test-day model and random regression test-day model with heterogeneous variance adjustment, respectively. Pedigree indices for daughters of young sires were upward biased, whereas pedigree indices for daughters of proven sires were slightly underestimated when heterogeneous variance was not accounted. Inclusion of test-day yields from the fourth lactation onwards increased the bias. Moving from repeatability animal model to random regression test-day model did not reduce the bias, whereas adjustment of heterogeneous variance reduced bias.

Key-words: dairy cattle, genetic evaluation, test-day model, heterogeneous variance, empirical bias

Introduction

Genetic evaluation for production traits in Finnish dairy cattle has always relied on the most advanced methods available. In 1990, the sire model evaluation was replaced by a single trait repeatability animal model evaluation (Strandén and Mäntysaari 1992). Ten years later, a multiple-trait random regression (RR) test-day (TD) model (Lidauer et al. 2000) was implemented and this model was replaced by the joint Nordic test-day model in 2006 (Mäntysaari et al. 2006, Lidauer et al. 2006). The TD model was adopted because it was considered advantageous

AGRICULTURAL AND FOOD SCIENCE

Lidauer M. et al. Bias in genetic evaluation

over the repeatability animal model (RPAM) for several reasons: first and later lactations are considered as two different traits and the breeding value for each trait is presented by a RR function, which accommodates breeding values for part lactations, 305-day yield, and persistency. Further, it accounts for the stage of pregnancy and allows better modelling of the herd environment. The joint Nordic test-day model added at least two more improvements to the Finnish dairy cattle yield evaluation. It incorporates additional sire information through modelling their daughter's performances in the neighbouring countries and it accounts for heterogeneous variance. Better modelling of the environmental effects in the TD model and adjustment for heterogeneous variance should reduce a possible bias in the breeding values.

Uimari and Mäntysaari (1993) studied the reliability of the Finnish repeatability animal model evaluation by comparing pedigree indices (PI), i.e., the average of the parents estimated breeding values (EBV), with final proofs and found a potential bias in PI for young sires. They explained that the bias in PI was caused by a bias in EBV for bull dams. Mäntysaari and Sillanpää (1993) showed that a redefinition of the herd effect in the model to be a herd × parity interaction was the most efficient way of reducing the bias. This was not practical because of the small herd sizes. About 20% of the cows would have been alone in their cow comparison group. Consequently, the herd effect was modelled with two components: a fixed herd × period of five calving years × parity group effect, and a random herd × year × parity group effect, where parity group classes were two, one for first and another for second and third lactation. Lidauer and Mäntysaari (1996) reported that this redefinition reduced the bias considerably.

The use of TD yields instead of 305-d lactation yields increases the amount of observations about tenfold. However, this does not ease the modelling of the herd environment when herd size is small. In typical TD models the herd environment is preferably modelled by a herd × TD × parity interaction effect, but this is not possible for small herds. Modelling the herd effect without a parity interaction is one possibility. It increases the accuracy of EBV for cows from small herds as found by Emmerling (2001), but it may cause potential bias as found by

Uimari and Mäntysaari (1993). Hence, the same concept of random subclasses within fixed main effects as was used for modelling the herd environment in the Finnish RPAM was adopted for the Finnish TD model.

The objective of the study was to quantify bias in PI's when firstly upgrading the Finnish RPAM to a reduced rank RR TD model (RRM) and secondly upgrading the RRM to a RRM that accounts also for heterogeneous variance. PI's of young sires, daughters of young sires, and daughters of proven sires were investigated.

Material and methods

Data

The TD and 305-d lactation yield records from the three dairy cattle breeds Ayrshire, Holstein-Friesian, and Finncattle were obtained from the national milk recording of Finland. The TD data consisted of records from all lactations of cows that calved the first time within the period January 1988 through June 2000. A TD record comprised of milk, protein and fat yield, where the milk yield was mandatory. The TD yields recorded before days in milk (DIM) 8 and after 365 were excluded. All cows with observations were required to have the first lactation records to avoid bias in the genetic trend. There were a total of 1,113,629 cows, which had on average 9.4 [14.3], 4.4 [6.7], and 4.4 [6.7] TD yields on milk, protein, and fat yield in the first [later] lactations, respectively, giving a total of 26,399,237 TD records (Table 1). The 305-d data set with protein yield records from the first three lactations contained 1,956,007 observations of all those cows which had at least four TD observations on milk yield in the TD data set. The pedigree data comprised of 1,558,850 cows and 11,825 bulls of the three breeds Ayrshire (79%), Holstein-Friesian (20%), and Finncattle (1%). The genetic differences between unknown parents were described by phantom parent groups categorized by breed, time period, and selection path.

Table 1. Number of observations, mean (kg), standard deviation (kg), number of cows with observations, and number of herd-test-day (HTD) classes in test-day (TD) yield data and 305-d lactation yield data by different traits: first lactation milk (M_F), protein (P_F) and fat (F_F); later lactations' milk (M_L), protein (P_L) and fat (F_L); and 305-d protein yield (P_{305d}).

	Earlier evaluation (1996)					Recent evaluation (2000)				
	Observations	Mean	Standard deviation	Cows with observations	HTD-classes	Observations	Mean	Standard deviation	Cows with observations	HTD-classes
TD-data										
M_F	6,743,319	18.82	5.13	724,546	1,911,229	10,496,572	19.58	5.37	1,111,697	2,727,338
P_F	3,164,804	0.61	0.15	714,088	896,591	4,947,516	0.65	0.16	1,095,729	1,284,520
F_F	3,164,874	0.83	0.22	714,089	896,639	4,947,607	0.86	0.23	1,095,730	1,284,580
M_L	9,075,776	20.88	6.87	493,743	1,651,926	15,902,665	21.48	7.18	793,987	2,485,678
P_L	4,267,545	0.69	0.2	486,909	776,732	7,505,427	0.72	0.21	784,171	1,173,241
F_L	4,267,604	0.92	0.3	486,909	776,766	7,505,502	0.94	0.32	784,171	1,173,287
305-d data										
P_{305d}	1,270,543	202.26	43.56	669,118	-	1,956,007	211.6	47.07	1,024,169	-

Models

Repeatability animal model

The 305-d protein yields were modelled by a single-trait RPAM including the first three lactations. The model from the national dairy cattle evaluation used until 1999 (Strandén and Mäntysaari 1992) was used, and was the reference model in this work:

$$y_{ijklmnop} = CADP_i + CYS_j + H5YP_{klm} + hyp_{knm} + \alpha_o + \pi_o + \varepsilon_{ijklmnop},$$

where y is a 305-d protein yield. The fixed effects were calving age × days open × parity *(CADP)*, calving year × calving season (*CYS*), and herd × period of five years of calving × parity group (*H5YP*). The random effects were a random herd × year of calving × parity group (*hyp*), a genetic animal effect (α), a permanent environmental effect (π) and a measurement error (ε). There were 6 calving age and 6 days open classes per parity, and 6 calving season classes per year (February-March, April-May…). Parity group classes were two, one for the first lactation and another for the second and third lactation. The heritability was 0.3, the repeatability 0.5, and the variance ratio between residual variance and herd-year variance was 1.82.

Random regression test-day model

The TD yields (*y*) were described by a multiple-trait multiple-lactation reduced rank RRM as was used until 2006 in the national dairy cattle evaluation (Lidauer et al. 2000). A TD yield, made on a DIM *d*, was modelled for the first lactation as

$$y_{Fijklmnoq} = CAP_{Fi} + DCCP_{Fj} + YM_{Fkl} + HY_{Fmk}$$
$$+ \sum_{r=1}^{5} \varphi(d)_{Fr} \, CYSP_{Fnr} + htd_{Fmkl}$$
$$+ \sum_{r=1}^{6} s(d)_{Fr} \, a_{or} + \sum_{r=1}^{6} t(d)_{Fr} \, p_{or}$$
$$+ e_{Fijklmnoq},$$

AGRICULTURAL AND FOOD SCIENCE

Lidauer M. et al. Bias in genetic evaluation

and for all the later lactations as,

$$y_{Lijklmnopq} = CAP_{Li} + DCCP_{Lj} + YM_{Lkl} + HY_{Lmk}$$

$$+ \sum_{r=1}^{5} \varphi(d)_{Lr} CYSP_{Lmr} + htd_{Lmkl}$$

$$+ \sum_{r=1}^{6} s(d)_{Lr} a_{o(r+6)} + \sum_{r=1}^{6} t(d)_{Lr} p_{o(r+6)}$$

$$+ \sum_{r=1}^{6} t(d)_{Lr} w_{opr} + e_{Lijklmnop},$$

where trait F is the first lactation milk, protein or fat yield and trait L is the later lactation milk, protein, or fat yield. Thus, there were 6 traits in the statistical model. The model for later lactation TD yields described observations from different lactations as repeated observations.

The fixed effects were calving age × parity (CAP), days carried calf × parity group ($DCCP$), production year × month (YM), herd × production year (HY) and the stage of lactation. The latter was described by a regression function of DIM d, nested within calving year × calving season × parity group ($CYSP$). Corresponding covariables for DIM d were $\varphi(d)' = [\ c_1\ c_2\ c_3\ c_4\ c_5]$, where c_1, c_2, and c3 represent a second order Legendre polynomial at DIM d, and c_4 and c_5 are exponential terms exp($-p_1$d) and exp($-p_2$d), where p_1 was 0.05 for milk yield and 0.10 for protein and fat yield, and p_2 was 0.06, 0.01, and 0.35 [0.04, 0.20, and 0.35] for milk, protein, and fat yield of first [later] lactation, respectively. There were 9 calving age classes within each of the first four parities and one calving age class for each of the remaining parities to account for the differences in the intercepts. Further, there were 5 parity group classes: first, second, third, fourth and fifth plus later parities; 5 days carried calf classes; and 3 calving season classes: November to February, March to June, and July to October.

The random effects were herd × TD (htd), RR function for additive genetic animal effect (a), RR function for animal environmental effects (p and w) and measurement error (e). The RR functions were defined across the traits and within the lactations.

The daily breeding value of a cow o on a DIM d in the first lactation was

$$\sum_{r=1}^{6} s(d)_{Fr} a_{or},$$

where the covariables in $s(d)_F$ were defined within the trait F. Similarly, the animal environmental RR function within the first lactation was

$$\sum_{r=1}^{6} t(d)_{Fr} p_{or}.$$

The breeding values and animal environmental effects for the later lactations had corresponding covariables and coefficients of their own. In addition, each later lactation had a within lactation specific RR function

$$\sum_{r=1}^{6} t(d)_{Lr} w_{op(r)},$$

which modelled the lactation specific animal environmental effects. The covariables in $s(d)$ and $t(d)$ were derived from the eigenfunctions representing the dominant eigenvalues in full fit covariance functions (CF) (Kirkpatrick et al. 1990) applied to all traits. Thus the genetic value of an animal was described by a vector of 12 RR coefficients and the animal environmental effects across all traits and within lactations were described by a vector of 12 RR coefficients plus 6 RR coefficients for each later lactation.

Let c be the vector of all htd effects, a the vector of all additive genetic animal effects, p the vector of all animal environmental effects across traits and within lactations, w the vector of all animal environmental effects within each later lactation, and e the vector of the measurement errors. The covariance matrix of these effects was assumed to be:

$$\text{var}\begin{bmatrix} c \\ a \\ p \\ w \\ e \end{bmatrix} = \begin{bmatrix} I \otimes C & 0 & 0 & 0 & 0 \\ 0 & A \otimes D_a & 0 & 0 & 0 \\ 0 & 0 & I \otimes D_p & 0 & 0 \\ 0 & 0 & 0 & I \otimes D_w & 0 \\ 0 & 0 & 0 & 0 & R \end{bmatrix},$$

where A is the numerator relationship matrix. The (co)variance matrices D_a, D_p, D_w and R, as well

as the covariables $s(d)_F$, $s(d)_L$, $t(d)_F$ and $t(d)_L$ were derived using a two step approach (van der Werf et al. 1998) that derived CF based on multiple-trait estimates. For the multiple-trait variance component estimation TD observations from five different DIM windows along the course of the first and of the second lactation were defined as ten different traits. The DIM windows within a lactation were: DIM 5–20, 31–60, 121–150, 211–240, and 301–330. The applied method was the same as explained for the first lactation traits by Mäntysaari (1999) and Lidauer et al. (2003). An extension of the method to all lactations is described by Emmerling et al. (2002). The used variances for the random *htd* and residual effects are presented in Table 2. Daily heritabilities, and genetic and phenotypic correlations across lactations and traits for a sample of DIM are given in Table 3. Heritabilities for 305-d lactation yields constructed from daily variances (DIM = 15, 45, …, 285) were 0.42, 0.28, and 0.29 [0.34, 0.27, and 0.30] for milk, protein, and fat yield of first [later] lactations, respectively. The "combined heritability" for the average of ten TD observations from the first plus ten TD observations from the second lactation was 0.44, 0.33, and 0.35 for milk, protein, and fat, respectively.

Random regression test-day model using data from first three lactations only

In contrast to the RPAM all lactations are described by the RRM. TD observations from the fourth lactation onwards were excluded in an additional evaluation (RRM-3) to quantify how much the TD observations from the fourth lactation onwards affect bias.

Multiplicative random regression test-day model

To assess the effect of heterogeneous variance adjustment on bias in EBV the same method as used in the current Nordic TD model evaluation was implemented. The method is based on the multiplicative mixed model approach of Meuwissen et al. (1996) and was adapted for the given RRM as described by Lidauer and Mäntysaari (2001). Effects in the multiplicative random regression test-day model (M-RRM) were identical with those in the RRM. TD observations of the same trait were stratified by

Table 2. Variance in kg² (diagonal) and correlation between traits (upper triangle) for first lactation milk (M_F), protein (P_F) and fat (F_F), and later lactations' milk (M_L), protein (P_L) and fat (F_L), for random herd × test-day and random residual effects used in the test-day model.

| | Random herd × test-day effect | | | | | | Random residual effect | | | | | |
	M_F	P_F	F_F	M_L	P_L	F_L	M_F	P_F	F_F	M_L	P_L	F_L
M_F	1.7587	0.9	0.65	0.9	0.84	0.6	4.1491	0.71	0.59	0	0	0
P_F		0.0025	0.65	0.85	0.9	0.54		0.0044	0.6	0	0	0
F_F			0.0043	0.63	0.62	0.9			0.0102	0	0	0
M_L				1.9874	0.9	0.6				9.6594	0.51	0.43
P_L					0.0028	0.6					0.0089	0.52
F_L						0.0059						0.023

Lidauer M. et al. Bias in genetic evaluation

Table 3. Heritabilities (on diagonal), genetic (above diagonal) and phenotypic (below diagonal) correlations from reduced rank covariance function for first lactation milk (M_F), protein (P_F) and fat (F_F), and later lactations' milk (M_L), protein (P_L) and fat (F_L) test-day yields by a sample of days in milk (DIM).

	DIM	M_F					P_F					F_F					M_L					P_L					F_L				
		10	85	160	235	310	10	85	160	235	310	10	85	160	235	310	10	85	160	235	310	10	85	160	235	310	10	85	160	235	310
M_F	10	.18	.77	.64	.57	.49	.86	.32	.28	.24	.20	.35	.44	.63	.59	.52	.66	.58	.45	.30	.14	.60	.50	.34	.17	.02	.43	.36	.25	.13	.00
	85	.43	.29	.97	.88	.62	.60	.77	.48	.40	.29	.28	.66	.41	.32	.18	.49	.78	.72	.61	.48	.37	.60	.52	.40	.28	.24	.46	.43	.35	.27
	160	.37	.64	.33	.97	.76	.48	.79	.82	.78	.59	.25	.41	.63	.37	.29	.34	.76	.79	.72	.63	.24	.57	.59	.51	.41	.16	.45	.47	.42	.36
	235	.31	.54	.62	.29	.89	.42	.70	.78	.82	.74	.22	.33	.37	.64	.38	.23	.70	.81	.79	.73	.16	.52	.62	.58	.50	.12	.41	.48	.46	.41
	310	.24	.38	.50	.61	.21	.38	.59	.52	.74	.85	.20	.22	.31	.38	.68	.08	.51	.71	.75	.72	.08	.39	.57	.58	.54	.10	.29	.40	.41	.39
P_F	10	.86	.32	.28	.24	.19	.76	.52	.41	.37	.35	.22	.30	.25	.19	.13	.55	.34	.21	.10	.00	.64	.56	.41	.27	.15	.55	.42	.28	.15	.04
	85	.38	.77	.47	.39	.27	.52	.72	.69	.61	.41	.30	.69	.58	.48	.32	.39	.51	.43	.35	.29	.42	.68	.62	.51	.42	.34	.54	.49	.41	.34
	160	.32	.48	.77	.47	.39	.41	.69	.72	.69	.52	.25	.42	.67	.54	.44	.25	.50	.50	.46	.43	.28	.66	.68	.62	.56	.23	.53	.54	.50	.46
	235	.26	.40	.47	.78	.50	.37	.61	.68	.71	.64	.19	.33	.39	.68	.42	.14	.44	.52	.52	.51	.19	.61	.72	.70	.66	.17	.48	.56	.55	.54
	310	.20	.29	.40	.50	.82	.35	.41	.52	.64	.74	.13	.21	.33	.42	.72	.03	.29	.45	.50	.51	.11	.47	.66	.69	.66	.12	.35	.49	.51	.51
F_F	10	.75	.28	.25	.22	.20	.22	.30	.25	.19	.13	.79	.17	.58	.48	.32	.35	.24	.19	.13	.07	.53	.38	.25	.13	.03	.69	.58	.41	.24	.09
	85	.34	.66	.41	.33	.18	.44	.69	.42	.33	.21	.17	.96	.20	.18	.19	.24	.39	.37	.32	.27	.33	.48	.41	.32	.24	.47	.73	.67	.55	.44
	160	.28	.41	.63	.39	.25	.39	.58	.67	.39	.33	.58	.20	.20	.18	.16	.16	.39	.41	.38	.34	.20	.47	.47	.41	.35	.31	.72	.75	.68	.60
	235	.21	.32	.37	.64	.33	.33	.48	.54	.68	.42	.48	.18	.18	.42	.40	.10	.35	.42	.40	.37	.12	.43	.51	.48	.42	.20	.67	.78	.75	.69
	310	.13	.18	.29	.38	.68	.24	.32	.44	.42	.72	.32	.19	.16	.40	.15	.05	.23	.34	.35	.34	.06	.32	.47	.47	.42	.09	.52	.70	.71	.68
M_L	10	.33	.17	.13	.08	.03	.28	.10	.08	.07	.04	.13	.24	.17	.13	.08	.11	.49	.31	.11	-.16	.90	.46	.28	.07	-.17	.56	.25	.19	.07	-.11
	85	.20	.34	.32	.27	.18	.10	.20	.17	.13	.12	.24	.34	.33	.28	.22	.30	.16	.85	.70	.60	.31	.78	.62	.47	.37	.15	.55	.43	.34	.29
	160	.17	.33	.35	.34	.28	.08	.17	.18	.16	.14	.17	.34	.33	.38	.34	.23	.47	.22	.97	.88	.21	.60	.72	.67	.57	.13	.42	.51	.50	.44
	235	.12	.28	.34	.36	.33	.07	.13	.16	.18	.18	.13	.28	.33	.36	.34	.15	.38	.47	.27	.96	.06	.47	.70	.72	.66	.07	.34	.49	.51	.48
	310	.07	.20	.29	.34	.34	.04	.08	.14	.18	.19	.08	.22	.23	.28	.34	.05	.28	.40	.46	.24	-.15	.43	.67	.73	.74	-.04	.31	.44	.47	.49
P_L	10	.34	.16	.12	.09	.06	.31	.13	.11	.08	.04	.37	.18	.15	.11	.06	.67	.21	.19	.09	.00	.67	.22	.17	.12	-.10	.84	.35	.23	.06	-.15
	85	.18	.27	.24	.20	.13	.13	.22	.19	.15	.10	.26	.28	.26	.21	.13	.25	.61	.30	.23	.17	.33	.61	.85	.70	.61	.35	.74	.60	.49	.44
	160	.14	.24	.26	.26	.22	.11	.19	.20	.20	.17	.20	.26	.20	.19	.17	.18	.31	.59	.31	.27	.27	.43	.72	.97	.89	.20	.59	.71	.71	.69
	235	.09	.19	.25	.28	.27	.08	.15	.19	.22	.22	.15	.21	.20	.22	.22	.10	.24	.31	.60	.34	.19	.37	.70	.71	.97	.05	.46	.67	.72	.74
	310	.04	.13	.21	.27	.30	.04	.10	.17	.22	.25	.10	.13	.19	.22	.30	.00	.17	.27	.33	.62	.10	.31	.67	.51	.63	-.11	.39	.58	.66	.73
F_L	10	.31	.13	.11	.09	.08	.18	.16	.12	.10	.08	.25	.13	.22	.12	.02	.69	.22	.19	.15	.10	.67	.22	.17	.11	.04	.25	.51	.29	.08	-.14
	85	.15	.22	.20	.15	.09	.15	.28	.26	.21	.14	.13	.74	.59	.46	.39	.30	.60	.32	.24	.17	.22	.61	.30	.24	.18	.29	.13	.86	.70	.61
	160	.11	.19	.20	.19	.15	.11	.26	.27	.26	.18	.22	.28	.18	.43	.21	.23	.31	.53	.28	.22	.16	.29	.59	.30	.27	.21	.44	.18	.96	.89
	235	.06	.14	.18	.20	.18	.06	.21	.26	.28	.22	.12	.21	.27	.22	.28	.14	.21	.26	.50	.26	.08	.22	.30	.60	.33	.12	.35	.43	.22	.97
	310	-.00	.07	.14	.18	.20	.02	.14	.22	.28	.24	.02	.14	.21	.28	.30	.03	.11	.19	.24	.50	-.02	.15	.26	.33	.63	.02	.25	.37	.44	.21

the production year × month × parity class *i* and by herd *j*. The M-RRM with different model lines for first (*F*) and later (*L*) lactation traits was:

$$\mathbf{y}_{Fij}\lambda_{Fij} = \mathbf{X}_{Fij}\mathbf{b}_F + \mathbf{C}_{Fij}\mathbf{c}_F + \mathbf{Z}_{Fij}\mathbf{a}_F$$
$$+ \mathbf{W}_{Fij}\mathbf{p}_F + \mathbf{e}_{Fij} \quad , \text{ and}$$

$$\mathbf{y}_{Lij}\lambda_{Lij} = \mathbf{X}_{Lij}\mathbf{b}_L + \mathbf{C}_{Lij}\mathbf{c}_L + \mathbf{Z}_{Lij}\mathbf{a}_L$$
$$+ \mathbf{W}_{Lij}\mathbf{p}_L + \mathbf{L}_{Lij}\mathbf{w}_L + \mathbf{e}_{Lij} ,$$

were vectors $\mathbf{b}$, $\mathbf{c}$, $\mathbf{a}$, $\mathbf{p}$, $\mathbf{w}$ and $\mathbf{e}$ contained fixed effects, random *htd* effects, additive genetic animal effects, animal environmental effects across all lactations and within lactation, animal environmental effects within later lactations, and residual effects of traits *F* or *L*, respectively. The matrices $\mathbf{X}_{.ij}$, $\mathbf{C}_{.ij}$, $\mathbf{Z}_{.ij}$, $\mathbf{W}_{.ij}$, and $\mathbf{L}_{.ij}$ were design matrices related to observations in stratum *ij*, and $\lambda_{.ij}$ was a multiplicative adjustment factor for stratum *ij*, and was calculated as

$$\lambda_{.ij} = e^{-0.5\left(\beta_{.i} + \beta_{.j}\right)} .$$

The $\beta_{.i} + \beta_{.j}$ will predict the heterogeneity in the TD data stratum, and $\beta_{.i}$ and $\beta_{.j}$ were estimated simultaneously while solving the model for breeding values. For each trait, the same model was defined:

$$s_{ij} = \beta_i + \beta_j + \varepsilon_{ij},$$

where s_{ij} was an heterogeneity observation for stratum *ij* (Meuwissen et al. 1996); β_i was a fixed production year × month × parity classification; β_j was a fixed herd classification; and ε_{ij} was the residual.

Analysis of bias

The empirical bias was defined to be the difference of earlier EBV and more recent EBV, where additional data has been accumulated. Two evaluations were carried out for each model. For the earlier evaluation, all 305-d observations from calvings after February 1996 and all TD observations after June 1996 were excluded. The 305-d yield data were cut four months earlier to make the information in both reduced data sets comparable (Table 1). The number of TD observations increased from the earlier to the recent test-day model evaluation by 56% and 75% for the first and the later lactations, respectively. The larger increase for the later lactations was because the cows with observations were required to have first lactation observations to avoid bias in genetic trend due to selection.

From the TD model evaluation, 305-d equivalent breeding values were calculated as a sum of all daily breeding values from DIM 8 to 312. The EBV from the earlier and recent evaluation were standardized to yield same mean and standard deviation for cows born in 1987. The EBV for the first and later lactations were equally weighted to get a combined breeding value. The bias was averaged over Ayrshire animals of three groups: young sires, daughters of young sires, and daughters of proven sires. The groups were defined as follows: a young sire was born within 1993 to 1995, had a Finnish herd book sire, had no daughters with observations in the earlier data set, and received an EBV with a reliability of over 0.9 in the recent evaluation; a daughter of a young sire was born within 1995 to 1997, had no observation in the earlier data but had at least four TD records in the recent data and was a daughter of a bull from the young sire group; a daughter of a proven sire was born within 1995 to 1997, had no observations in the earlier data but at least four TD records in the recent data and was a daughter of a Finnish herd book sire born within 1986 to 1990 that received an EBV with a reliability of over 0.9 in the earlier evaluation, and had at least 100 more daughters in the recent evaluation. The early EBV for target animals were based on pedigree information only, and therefore they are called PI. By definition the expected value of a PI is equal to the expected value of the final proof. From here onwards the term bias is used for the deviation of the PI mean from its final proof. The bias in PI was studied for first lactation protein yield, later lactation protein yield, and for the combined protein yield, calculated as an average of first and later lactation protein yield.

Lidauer M. et al. Bias in genetic evaluation

Results

Overall, better modelling of the environmental effects by the RRM did not reduce bias compared to the RPAM. However, the observed bias was significantly smaller when adjusting for heterogeneous variance (Table 4). PI's were upward biased for young sires and daughters of young sires, whereas PI's were downward biased for daughters of proven sires. When comparing the results from the RRM-3 evaluation with that one based on RPAM, only small differences in the bias was found. PI's of young sires were biased upwards by 2.5 kg and 2.2 kg in the RRM-3 and RPAM evaluations, respectively. The same holds for the daughters of young sires, for which bias in evaluation RRM-3 and RPAM was similar for all birth year groups. Both models (RRM-3 and RPAM) underestimated the PI of daughters of proven sires but here the differences between models were more apparent (Table 4). The inclusion of all lactations into the breeding value estimation (RRM) increased the bias in young sires about 0.2 kg. The increase was found in both, PI's of first lactation as well as in PI's of later lactations. For the daughters of proven sires the PI's were about 0.2 kg less underestimated when including later lactations.

Accounting for heterogeneous variance (M-RRM) reduced bias significantly. Bias in PI's of young sires reduced by 35% compared to the RRM evaluation and was the smallest when comparing all four evaluations. PI's of daughters of young sires were less upward biased in the oldest year group and slightly upward and downward biased in the younger year groups. PI's of daughters of proven sires showed no bias or only small bias in either one or the other direction.

Later lactation PI's of young sires and of daughters of young sires were more upward biased than their first lactation PI's. The 30% larger standard deviation in breeding values for the later lactation protein yield may explain only a part of the difference. PI's for young sires were 1.24 kg (57%) more upward biased than their PI's for first lactation. For PI's of daughters of young sires the difference was even larger (Table 4). However, using either first

Table 4. Mean difference between breeding values for protein yield of earlier evaluation (PI) and recent evaluation (EBV_{recent}) for different groups of Ayrshire by different models, traits, and data sets. Single trait repeatability animal model (RPAM) including 305-d lactation yields from first three lactations; random regression test-day model when including test-day (TD) observations from first three lactations (RRM-3); from all lactations (RRM); and multiplicative RRM including observations from all lactations (M-RRM). Traits in test-day models were: first lactation protein (P_F), later lactation protein (P_L) yield and their average (P_C).

| | Year of birth | Number of animals | Observations per animal | | RPAM | PI–EBV$_{recent}$ (kg) | | | | | | | | |
| | | | TD | 305-d | | RRM-3 | | | RRM | | | M-RRM | | |
						P_F	P_L	P_C	P_F	P_L	P_C	P_F	P_L	P_C
Young sires	'93–95	260	-	-	2.23	1.94	3.19	2.52	2.16	3.4	2.74	1.66	1.95	1.78
Young sire daughters	1995	21,495	21.8	1.8	2.1	1.34	3.05	2.17	1.58	3.25	2.38	1.52	2.53	2.01
	1996	22,647	17.2	1.4	0.33	0.34	0.55	0.48	0.16**	0.42	0.32	–0.09**	–0.67	–0.40
	1997	21,923	10.7	1	1.07	0.23	1.33	0.74	0.39	1.48	0.9	0.19	0.5	0.32
Proven sire daughters	1995	18,805	22.4	1.9	–0.56	–1.79	–1.51	–1.69	–1.57	–1.28	–1.48	–0.64	–0.18	–0.43
	1996	26,128	17.5	1.4	–0.42	–1.50	–1.61	–1.59	–1.24	–1.33	–1.32	–0.26	0.23	–0.04[ns]
	1997	23,040	10.9	1	–0.66	–1.29	–1.76	–1.56	–1.01	–1.51	–1.30	–0.06[ns]	–0.10**	–0.11**

[ns] not significant; ** $P < 0.01$; for all other values $P < 0.001$

three lactations (RRM-3) or all lactations (RRM) did not affect the difference in the bias between the PI's for first and the PI's for the later lactations. On the contrary, adjustment for heterogeneous variance resulted similar bias across lactations. Heterogeneous variance adjustment reduced the bias in first lactation young sire PI's by 23% and in later lactation young sire PI's by 42%.

The size of the bias in the PI's for daughters of young sires was best expressed in daughters with full records in the recent data. The largest bias was found for daughters born in 1995, which had on average 22 TD observations, whereas the bias was considerable smaller for daughters born in 1997, which had on average 11 TD-observations. For instance, from the RRM evaluation with all data, the combined PI's for daughters of young sires were overestimated by 2.4 kg in the birth group 1995 but only by 0.9 kg in the birth group 1997. Estimates for the birth year 1995 are closer to true bias since dams EBV from both evaluations were more reliable. Dams of the daughters had the possibility to have at least the first lactation completed in the earlier evaluation and at least the fifth lactation completed in the recent evaluation.

Discussion

The use of TD-yields in the breeding value estimation yielded about the same size of bias in PI's as when 305-d yields were used. This was contrary to our expectation that a better modelling of the herd environment by the TD model would remove some sources of the bias. Apparently, the seasonal changes in the herd environment, which were modelled in the RRM but not in the RPAM, had no effect on the magnitude of the bias, unless the effect was equalled out by other effects of the RRM. Lidauer et al. (2003) found a 4 to 9% increase in the standard deviation of cow EBV's when using a TD model rather than a 305-d yield model for the breeding value estimation. They used the same heritability for both models and argued that the increase in the standard deviation was

due to the better modelling of the environment, which increases the reliabilities of the EBV's. In our study the heritability in RPAM was higher than the one used in RRM. However, the standard deviations of young sires' EBV's were of the same magnitude in both, RRM and RPAM evaluation; 12.4 kg and 11.7 kg, respectively. Thus, the RRM must have compensated the lower heritability by a better modelling of the environment. However, this did not reduce bias. A reason might be that the herd TD was modelled as random effect in RRM, which follows the same concept as in RPAM of modelling random subclasses (herd-year) within fixed main classes (herd-5-year periods). Another reason might be that the RRM gives more weights to own performances as was shown by Mrode et al. (2006). This will put more emphasize on Mendelian sampling deviations, which increases bias in case of preferential treatment.

The bias in 305-d yield PI's found here was less than reported by Uimari and Mäntysaari (1993 and 1995). They found a 5.2 kg and 13.6 kg upward bias in PI's of Ayrshire young sires in their first and later investigation, respectively. They argued that the difference was a result of the higher heritability used in the later investigation (0.25 versus 0.30) and the different set of bulls used. Based on a follow-up study (Mäntysaari and Sillanpää 1993) the model was modified to RPAM used in this study. Lidauer and Mäntysaari (1996) used the same modified model and heritability as used here (RPAM), to investigate bias in PI. They reported an upward bias of 2.2 kg in the PI's for daughters of Ayrshire young sires, which is in agreement with our result of a 2.1 kg upward bias for the birth year group 1995 when applying RPAM. Further, they found that the PI's for daughters of proven sires were almost unbiased (0.2 kg). The corresponding value from our study for the birth year group 1995 was –0.6 kg for RPAM but was –1.5 kg for RRM. The found downward bias can have several reasons. The breeding values of proven sires from the earlier evaluation were required to have an accuracy of over 0.9 and therefore should not be biased. Also the dams of the daughter group born in 1995 received from the earlier evaluation breeding values

AGRICULTURAL AND FOOD SCIENCE

Lidauer M. et al. Bias in genetic evaluation

based on own first and later lactation information. However, in progeny testing it is common that young sires are mated with cows that perform below the herd average. If these cows and their daughters face a below average herd environment their sire's first proof might be undervalued. In the case that the progeny tested sire is selected to produce a second crop of daughters, it will be mated with cows that perform mainly above the herd average, and the daughters might enjoy a more preferable herd environment, which in turn would rise the sires breeding value. Similarly, Uimari and Mäntysaari (1993) found that the EBV's for proven sires increased by 1.4 kg when adding data with the second crop daughters. This would explain to some extent the downward bias in the PI's but it does not explain the difference in bias of 0.9 kg found between RPAM and RRM. One reason could be the better ability of RRM to model lactation in progress, which yields a faster increase in standard deviation of EBV when information accumulates (Lidauer et al. 2003).

The inclusion of TD observations from the fourth lactation onwards increased the bias in the PI's of young sires between 5 and 11%. The inclusion of the later lactations increased the amount of observations for the young sires' bull dams, but not for the young sires' daughters. Thus, the increase in bias was caused by an increased upward bias in the EBV's of bull dams. Whenever a young cow is selected to become a bull dam, it will belong to the most valuable cows of the herd. Naturally, the farmer will provide best management to this cow group. Consequently, the deviation of the milk yield of a bull dam from the herd mean is not only affected by the cow's genetic potential. This is referred as preferential treatment (e.g. Kuhn et al. 1994), which is one source of bias in EBV's. An estimation of breeding values based on observations from the first lactation only would diminish this problem but at the cost of discarding most of the available records. The better management for bull dams explains to some extent the higher bias in the later lactation PI. A bias in PI caused by preferential treatment will be more pronounced with RRM, since RRM puts more weight on yield and less on pedigree infor-

mation compared to RPAM. Mrode et al. (2006) showed that upgrading a single trait animal model to a single trait RRM increases significantly the contribution of the cow's own performance to the cow's EBV. Applying a heritability of 0.57 for both models, they found for all cows without progenies an average increase in the contribution of the yield deviation from 69% to 86%, whereas the contribution of pedigree information decreased from 31% to 14%.

A significant source of the found bias was caused by heterogeneous variance across herds. Accounting for heterogeneous variance reduced bias by 35%. Similar, Meuwissen et al. (1996) reported a 38% reduction in bias of repeatability animal model PI's when accounting for heterogeneous variance. In our study bias was larger in later lactation PI's but also reduction of bias due to heterogeneous variance adjustment was larger in later lactation PI's. This implies that there is a larger source for heterogeneous variance in the later lactations, which agrees with preferential treatment of bull dams in their later lactations. The effect of heterogeneous variance found here is in agreement with the finding by Uimari and Mäntysaari (1995) who found that PI's were highly biased for bulls coming from small herds with a high within-herd variance. Contrary to our study, Van Steenbergen et al. (2006) reported a very low benefit of heterogeneous variance adjustment to reduce bias. They found a bias in bull dam EBV's for protein yield of 18.1 kg when evaluation was based on a RR TD model without adjustment for heterogeneous variance. When adjusting for heterogeneous variance, the bias was reduced by 4% only. They suggested normalizing the distribution of the residuals and Mendelian sampling deviations by a Gaussian transformation, which yielded a reduction in bias of 33%.

Too high heritabilities may be another source of bias in EBV's was found by Uimari and Mäntysaari (1993). However, the heritabilities used in this study are moderate compared to heritabilities often reported for other RRM (De Roos et al. 2004, Thompson et al. 2005).

Conclusions

The use of TD observation instead of 305-d yield observations for the breeding value estimation did not reduce a bias in pedigree indices. The better modelling of the herd environment increased the standard deviation of the cows' EBV's, which in turn also increased the size of the bias. Including the TD observations from the fourth lactation onwards yielded a moderate increase in the bias for young sires. This was expected because later lactation yields of bull dams are often inflated due to a more preferable environment. Bias in bull dam EBV's was significantly reduced when accounting for heterogeneous variance, which in turn reduced bias in PI's of young sires.

References

De Roos, A.P.W., Habers, A.G.F. and De Jong, G. 2004. Random herd curves in a test-day model for milk, fat, and protein production of dairy cattle in The Netherlands. *Journal of Diary Science* 87: 2693–2701.

Emmerling, R. 2001. *Optimierung der Zuchtwertschätzung für Milchleistungsmerkmale unter besonderer Berücksichtigung der Umwelteinflüsse in einem Testtagsmodell*. Dissertation, Technische Universität München, Germany. 163 p.

Emmerling, R., Mäntysaari, E.A. & Lidauer, M. 2002. Reduced rank covariance functions for a multi-lactation test-day model. *Proceedings of 7th World Congress on Genetics Applied to Livestock Production,* Montpellier, France, CD-ROM: 17-03.

Kirkpatrick, M., Lofsvold, D. & Bulmer, M., 1990. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124: 979–993.

Kuhn, M.T., Boettcher, P.J. and Freeman, A.E. 1994. Potential biases in predicted transmitting abilities of females from preferential treatment. *Journal of Diary Science* 77: 2428–2437.

Lidauer, M. & Mäntysaari, E.A. 2001. Multiplicative random regression model for test-day data with heterogeneous variances. INTERBULL Open Mtg., Budapest, Hungary, August 30–31, 2001. *Proceedings of the 2001 Interbull Meeting*, Budapest, Hungary. Interbull Bulletin 27: 167-171.

Lidauer, M. & Mäntysaari, E. 1996. Detection of bias in animal model pedigree indices of heifers. *Agricultural and Food Science in Finland* 5: 387–397.

Lidauer, M., Mäntysaari, E.A. & Strandén, I. 2003. Comparison of test-day models for genetic evaluation of production traits in dairy cattle. *Livestock Production Science* 79: 73–86.

Lidauer, M., Mäntysaari, E.A., Strandén, I. & Pösö, J. 2000. Multiple-trait random regression test-day model for all lactations. *Proceedings of the 2000 Interbull Meeting*, Bled, Slovenia. *Interbull Bulletin* 25: 81–86.

Lidauer, M., Pedersen, J., Pösö, J., Mäntysaari, Strandén, I., Madsen, P., Nielsen, U.S., Eriksson, Å.-J., Johansson, K. & Aamand, G.P. 2006. Joint Nordic test day model: Evaluation Model. Proceedings of the 2006 Interbull Meeting, Kuopio, Finland. *Interbull Bulletin* 35: 103–107.

Meuwissen, T.H.E., de Jong, G., & Engel B. 1996. Joint estimation of breeding values and heterogeneous variances of large data files. *Journal of Diary Science* 79: 310–316.

Mrode, R., Coffey, M. & Jones, H. 2006. Understanding cow evaluations in a random regression mdel. *Proceedings of 8th World Congress on Genetics Applied to Livestock Production*. Belo Horizonte, MG, Brazil, CD-ROM: 01-32.

Mäntysaari, E.A. 1999. Derivation of multiple trait reduced rank random regression (RR) model for the first lactation test day records of milk, protein and fat. *Proceedings of the 50th Annual Meeting of the European Association of Animal Production*. Zurich, Switzerland. pp 26.

Mäntysaari, E.A., Lidauer, M, Pösö, J., Strandén, I., Madsen, P., Pedersen, J., Nielsen, U.S., Johansson, K., Eriksson, Å.-J. & Aamand, G.P. 2006. Joint Nordic test day model: Variance components. *Proceedings of the 2006 Interbull Meeting*, Kuopio, Finland. Interbull Bulletin 35: 97–102.

Mäntysaari, E.A. & Sillanpää, M.J. 1993. Bias in pedigree indices of dairy bulls: Should the management group effect be fixed and should we use smaller heritability? *Proceedings of the 44th Annual Meeting of the European Association of Animal Production*, Aarhus, Denmark. Vol 1: 236–237.

Strandén, I. & Mäntysaari, E.A. 1992. Animal model evaluation in Finland: Experience with two algorithms. Journal of Diary Science 75: 2017–2022.

Thompson, R., Brotherstone, S. & White, I.M.S. 2005. Estimation of quantitative genetic parameters. *Philosophical Transactions of the Royal Society*. B (2005) 360: 1469–1477.

Uimari. P. & Mäntysaari, E.A. 1995. Relationship between bull dam herd characteristics and bias in estimated breeding value of bull. *Agricultural and Food Science in Finland* 4: 463–472.

Uimari, P. & Mäntysaari, E.A. 1993. Repeatability and bias of estimated breeding values for dairy bulls and bull dams calculated from animal model evaluations. *Animal Production* 57: 175–182.

van der Werf, J.H.J., Goddard, M.E. & Meyer, K., 1998. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. *Journal of Diary Science* 81: 3300–3308.

Van Steenbergen, E.J., De Roos, A.P.W. & De Jong, G. 2006. Reduction of bias in cow breeding values from a random regression test-day model. *Proceedings of 8th World Congress on Genetics Applied to Livestock Production*. Belo Horizonte, MG, Brazil, CD-ROM: 24-05.

AGRICULTURAL AND FOOD SCIENCE

Lidauer M. et al. Bias in genetic evaluation

SELOSTUS
Koelypsymalliin perustuvan jalostusarvostelun luotettavuus

Martin Lidauer ja Esa A. Mäntysaari
MTT Biotekniikka- ja elintarviketutkimus

Tutkimuksen tavoitteena oli vertailla, kuinka muutokset Suomen lypsykarjan tuotosominaisuuksien arvostelumallissa ovat vaikuttaneet polveutumisindeksin harhaisuuteen. Vertailussa oli kolme arvostelumallia: toistuvuusmalli, satunnaisregressiomalli ja heterogeenisen varianssin huomioon ottava satunnaisregressiomalli. Vertailu perustui valkuaistuotoksen polveutumisindeksien ja tulevien jalostusarvojen välisiin keskimääräisiin erotuksiin. Tämä erotus määriteltiin harhaksi. Polveutumisindeksit laskettiin aineistosta, josta oli poistettu viimeisten neljän vuoden havainnot. Tulokset osoittivat, että ayrshirerotuisten nuorsonnien polveutumisindeksit olivat toistuvuusmallilla 2,2 kg, satunnaisregressiomallilla 2,5 kg ja heterogeenisen varianssin huomioon ottavalla satunnaisregressiomallilla 1,8 kg ylöspäin harhaisia. Nuorsonnien tyttärien polveutumisindeksit olivat ylöspäin ja valiosonnien tyttärien alaspäin vinoutuneita, kun heterogeenista varianssia ei otettu huomioon. Neljännen ja sitä seuraavien lypsykausien havaintojen sisällyttäminen arvosteluun kasvatti harhaa edelleen. Toistuvuusmallin korvaaminen satunnaisregressiomallilla ei ole pienentänyt jalostusarvojen harhaa, kun sen sijaan heterogeenisen varianssin huomioon ottaminen on.