Estimation of genetic parameters for production traits in Polish Simmental cattle using a random regression test-day model

Dawid Słomian[1], Klaudia Gogola[1], Monika Skarwecka[1], Kacper Żukowski[1]

[1]National Research Institute of Animal Production, Krakowska 1, 32-083 Balice, Poland

Abstract

In Poland, routine genetic evaluation of Simmental cattle for production traits is based on a multi-trait animal model using total yields for each of the first three lactations. The purpose of the study was to select a reference population for the Simmental breed, which is the second largest population of cattle in Poland, and to estimate variance components for production traits, using a single-trait, multi-lactation random regression test-day model (RR-TDM). Analyses were performed for the first three lactations within the range of 5 to 305 days in milk (DIM). Random regression was fitted using Legendre polynomials, and variance components were estimated by Gibbs sampling. The collected production traits included: milk yield in kilograms; fat, protein, and lactose expressed in kilograms and percentages; and somatic cell score. Phenotypic data from 13,966 cows that were purebred Simmentals, comprising 256,341 test-day records from the first three lactations, were used to estimate variance components. The dataset was prepared without crossbred animals and records from herds with only one cow. Based on additive genetic, permanent environmental, and residual variance components for each trait across lactation, daily heritability was calculated. The average heritability estimates were: 0.20 for milk yield, 0.18 for fat yield, 0.25 for fat content, 0.19 for protein yield, 0.37 for protein content, 0.22 for lactose yield, 0.32 for lactose content, and 0.13 for somatic cell score. The highest average heritability was observed for protein percentage, whereas the lowest was observed for somatic cell score. Depending on the trait, minimum and maximum heritability occurred at different stages of lactation. For most traits, maximum heritability occurred in the late stage of lactation. The results confirm that the RR-TDM approach is suitable for modeling production traits and for describing changes in genetic variation during lactation.

Key words: pedigree BLUP, test-day model, simmental cattle, routine evaluation

Introduction

The routine evaluation of the breeding value of Simmental cattle in Poland for production traits is currently based on a multi-trait animal model that uses aggregated production data from the first three lactations. However, advances in statistical methods and computational capabilities enable models that use test-day data, enabling a more accurate description of changes in yield and variance components throughout the entire lactation.

Test-day models with random regression allow for the treatment of performance on successive days of lactation as a continuous trait and for the modelling of individual genetic and environmental curves for animals. The advantage of this approach is that it allows for changes in genetic variance and permanent environment across days of lactation, rather than assuming a homogeneous variance structure throughout lactation (Schaeffer, 2004).

The significance of this approach for the Polish Simmental cattle population was confirmed by Otwinowska-Mindur et al. (2025), who, using a multi-lactation test-day model with random regression, demonstrated significant variability in heritability over the course of lactation and, for most of the analyzed traits, the highest heritability values during the final stage of lactation.

The objective of this study was: (1) to prepare a reference dataset comprising purebred Simmental cows, (2) to estimate variance components and daily heritability estimates for production traits using a single-trait, multi-lactation random regression test-day model, and (3) to prepare the first validation of breeding values obtained using a pedigree BLUP model.

Materials and Methods

To estimate the variance components, a dataset was prepared using animals with purebred Simmental breed composition. After data quality control, crossbreds and observations from herds with only one cow were removed. The final dataset included 13,966 cows and 256,341 test-day observations from the first three lactations. Observations between the 5th and 305th days of lactation were included.

Eight traits were analyzed: milk yield, fat yield, fat content, protein yield, protein content, lactose yield, lactose content, and SCS. For each trait, a single-trait, multi-lactation test-day model was used.

The pedigree data were checked and prepared using the Relax2 program (Strandén and Vuori, 2006). The prepared pedigree and phenotypic data were used as input to the RENUMF90 program, which renumbered the animals and created input files for further analysis in GIBBSF90 (Lourenco et al., 2022). Before running RENUMF90, the day of lactation was rescaled to the range for Legendre polynomials, and for each record, three normalized values of Legendre polynomials were calculated. These values corresponded to the three regression coefficients subsequently used to model the random additive genetic and permanent environmental effects.

The variance components were estimated using Gibbs sampling in the GIBBSF90 program. For each trait analyzed, 200,000 Markov chain iterations were performed. The first 20,000 iterations were discarded as the burn-in period. Afterwards, every tenth iteration was saved, resulting in a total of 18,000 samples from the posterior distribution used for further analysis. The posterior distributions and the results of the variance component estimates were then analyzed using the POSTGIBBSF90 program. Based on the estimated (co)variance matrices of the regression coefficients, the results and summary statistics were calculated using the R package.

The calculated variance components were then used as model parameters in the analysis of a larger dataset of production records. Breeding values were estimated using the MiXBLUP 3.0 software (Vandenplas et al., 2022). The pedigree BLUP test-day animal model can be written as:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Wf} + \boldsymbol{Vp} + \boldsymbol{Vu} + \boldsymbol{e},$$

where $\boldsymbol{y}$ is the vector of cows' test-day records for production traits from the first three lactations, $\boldsymbol{b}$ is the vector of fixed effects of herd-test-day-parity-milking frequency, $\boldsymbol{f}$ is a fixed lactation curve effect, $\boldsymbol{p}$ is a vector of permanent environmental effects expressed as random regression coefficients of the Legendre polynomials, $\boldsymbol{u}$ is a random additive genetic effect based on pedigree.

The phenotype data from 49,564 cows were used for the full evaluation, while the pedigree included 121,963 cows and 61,350 bulls.

The linear regression (LR) method proposed by Legarra and Reverter (2018) was used for the validation. This method is based on comparing breeding values obtained from two consecutive evaluations conducted with the same model but using different phenotypic information. The full data set contains all available information, while the truncated data set includes all phenotypic data except the youngest animals from the last four years. The LR method provides statistics describing bias, dispersion, and predictive performance (Legarra and Reverter, 2018; Bermann et al., 2024)

Validation cows were defined as animals whose observations were removed when preparing the truncated dataset. Validation bulls were young sires of cows in the validation group. Validation was performed separately for

6,355 cows and 160 bulls. For both groups, breeding values estimated based on the full and truncated datasets were compared using regression:

$$\boldsymbol{EBVf} = b_0 + b_1\boldsymbol{EBVt} + e,$$

where **EBVf** represents the vectors of breeding values estimated based on the full data set, while **EBVt** represent the vector of breeding values predicted by the truncated data set, b_0 represent the intercept, which indicated a systematic bias in the model's prediction, and b_1 represents the regression slope, the dispersion of prediction compared to actual results. The R^2 coefficient is one of the results of linear regression and serves as a measure of prediction accuracy; it indicates the percentage of variance in the **EBVt** explained by **EBVf**.

Results & Discussion

Variance components

The average heritability was calculated based on results from the first three lactations; the highest heritability was obtained for protein content (0.37) and lactose content (0.32), followed by fat content (0.25). For traits expressed in kilograms, the average heritability values were lower, at 0.20 for milk yield, 0.177 for fat yield, 0.19 for protein yield, and 0.22 for lactose yield. The lowest heritability was recorded for SCS (0.13). The repeatability values were higher than the heritability values for all analyzed traits. The highest repeatability values were obtained for lactose yield (0.72), milk yield (0.72), and protein yield (0.68), while the lowest average repeatability was recorded for fat content (0.38). Detailed results on heritability and repeatability are presented in Table 1.

The estimates obtained were in general agreement with those previously reported for Simmental cattle. Dong et al. (2026) have likewise recently reported similar findings for Chinese Simmental cattle, with heritabilities estimated at 0.28 for fat content, 0.52 for protein content, and 0.35 for lactose content using a random regression test-day model. The differences in estimate sizes across studies might be due to differences in population structure.

Validation results

For the validation cows, correlation between full and truncated datasets ranged from 0.54 for fat content to 0.77 for protein content. Relatively high correlation values were also obtained for lactose content (0.74), protein yield (0.73), milk yield (0.71), and lactose yield (0.71). The lowest correlation was observed for fat content (0.54) and SCS (0.59). For the

Table 1: Heritability and repeatability estimates for production traits across the firs three lactations.

Trait	Lactation	heritability	repeatability
Fat yield	1st	0.18	0.56
	2nd	0.20	0.61
	3rd	0.16	0.60
	Total	0.18	0.59
Fat content	1st	0.27	0.40
	2nd	0.23	0.37
	3rd	0.25	0.37
	Total	0.25	0.38
Lactose yield	1st	0.17	0.67
	2nd	0.24	0.73
	3rd	0.22	0.74
	Total	0.22	0.72
Lactose content	1st	0.32	0.53
	2nd	0.34	0.61
	3rd	0.31	0.62
	Total	0.32	0.59
Milk yield	1st	0.19	0.67
	2nd	0.21	0.72
	3rd	0.20	0.73
	Total	0.20	0.72
Protein yield	1st	0.17	0.63
	2nd	0.20	0.69
	3rd	0.19	0.70
	Total	0.19	0.68
Protein content	1st	0.40	0.61
	2nd	0.34	0.63
	3rd	0.38	0.63
	Total	0.37	0.62
SCS	1st	0.09	0.49
	2nd	0.14	0.54
	3rd	0.15	0.59
	Total	0.13	0.54

validation bulls, the results were lower than those for validation cows. The correlations ranged from 0.40 for SCS to 0.79 for protein content. High correlations were also observed for the lactose content (0.73) and protein yield (0.61). Detailed correlation and coefficient of determination values are shown in Figure 1.

The regression coefficient b_1 for the validation cows was very close to the expected value of 1, ranging from 0.92 (fat content) to 1.05 (protein content). The results closest to the ideal value were those for lactose content (1.01), protein yield (0.98), milk yield (0.97), and lactose yield (0.97). In the context of validation bulls, the b1 values were more variable, ranging from 0.59 for SCS to 1.03 for protein content. Only protein content and lactose content were values obtained that were relatively close to the expected values. Detailed results for the regression coefficient b1 are shown in Figure 2.

In the absence of a systematic shift and with proper centering of the breeding values, the expected value of b_0 should be close to 0. For the validation cows, the b_0 values ranged from -2.26 for protein content to 2.18 for protein yield, except for milk yield, for which b_0 was 51.67. For the validation bulls, greater variation was observed, ranging from -26.90 for SCS to 90.13 for milk yield. However, the breeding values were analysed on their original scale and were neither centered nor standardised before regression. Detailed results for the bias b_0 are shown in Figure 3.

Conclusions

The results obtained confirm the usefulness of the multi-lactation random regression test-day model for evaluating production traits in the Polish Simmental cattle population. The estimated heritability and repeatability values indicate varying degrees of genetic determination for the traits analysed, with milk composition traits generally exhibiting higher heritability than traits in kilograms. Pedigree BLUP validation showed satisfactory stability for validation cows, as indicated by moderate-

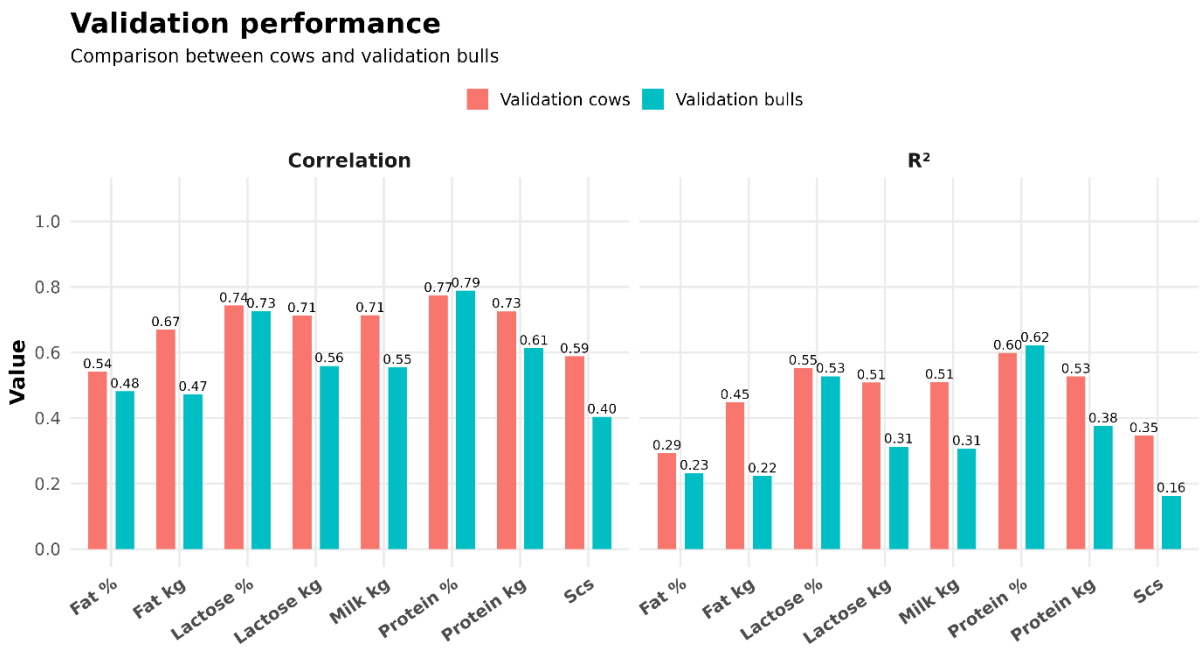


Figure 1. Correlation and coefficient of determination (R^2) between breeding values estimated using the full and truncated datasets for validation cows and bulls across production traits.

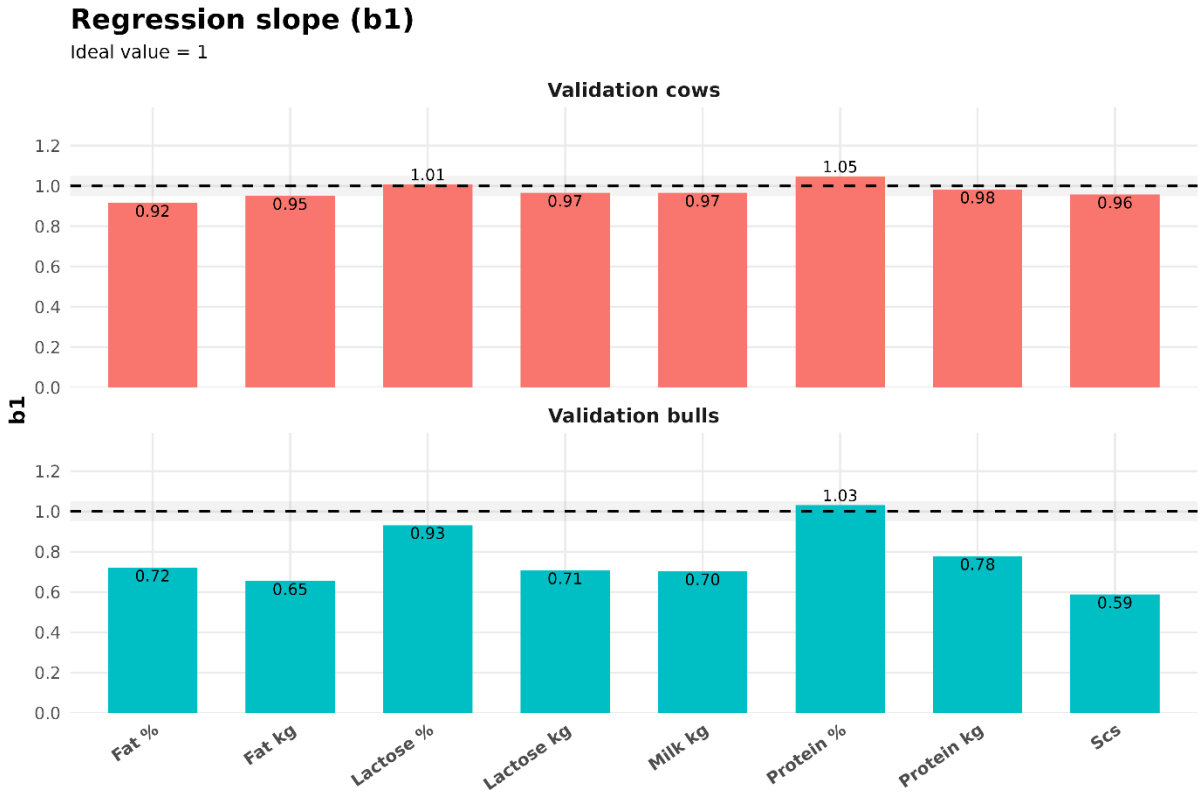


Figure 2. Regression slope (b_1) for validation cows and bulls across production traits

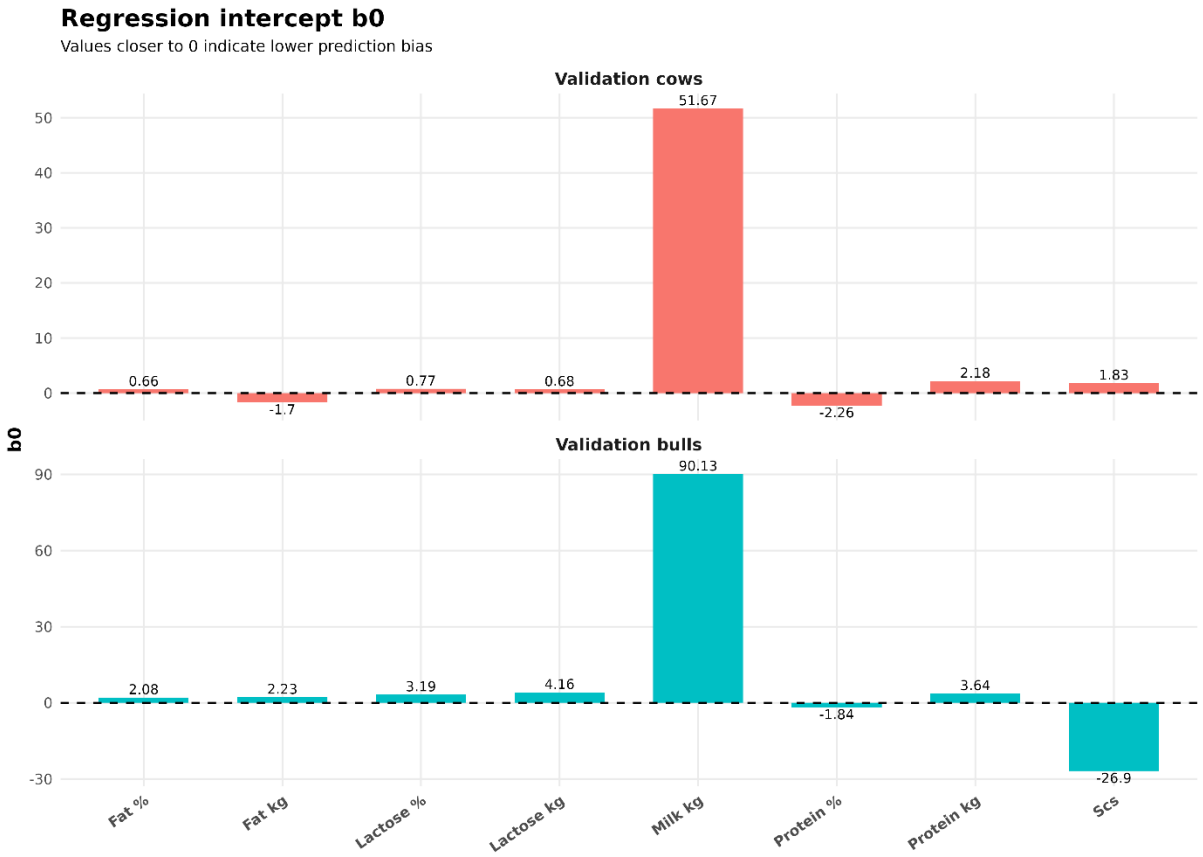


Figure 3. Regression intercept (b_0) for validation cows and bulls across production traits.

to-high correlations between full and truncated evaluations and a regression coefficient (b_1) close to 1. For the validation bulls, evaluation consistency was lower and more variable across traits, which may be related to the smaller validation group size and the limited available information. The b_0 values, however, showed greater variation due to the lack of standardization of breeding values. Overall, the results obtained provide a basis for the random regression test-day model for Polish Simmental cattle and for the further development of the system toward genomic evaluation using the single-step approach.

References

Bermann, M., Legarra, A., Alvarez Munera, A., Misztal, I., and Lourenco, D. 2024. Confidence intervals for validation statistics with data truncation in genomic prediction. *Genet. Sel. Evol.*, 56, 18. https://doi.org/10.1186/s12711-024-00883-w.

Dong, C., Ma, P., Tang, Y., Han, H., Chen, S., Yang, J., Sun, Z., Wang, W., Yu, Y., and Li, S. 2026. Genetic parameter estimation and fine-mapping of milk-production traits and somatic cell score in Chinese Simmental cattle. *J. Dairy Sci.*, 109(4), 4195–4208. https://doi.org/10.3168/jds.2025-27039.

Legarra, A., and Reverter, A. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet. Sel. Evol.*, 50, 53. https://doi.org/10.1186/s12711-018-0426-6.

Lourenco, D., Tsuruta, S., Aguilar, I., Masuda, Y., Bermann, M., Legarra, A., and Misztal, I. 2022. Recent updates in the BLUPF90 software suite. *Proceedings of the 12th World Congress on Genetics Applied to Livestock Production*, 1530–1533. https://doi.org/10.3920/978-90-8686-940-4_366.

Otwinowska-Mindur, A., Ptak, E., Jagusiak, W., and Zarnecki, A. 2025. Genetic parameters for milk production traits of Simmental cows with random regression test-day model. *Animal,* 19(2), 101395. https://doi.org/10.1016/j.animal.2024.101395.

Schaeffer, L.R. 2004. Application of random regression models in animal breeding. *Livest. Prod. Sci.*, 86(1–3), 35–45. https://doi.org/10.1016/S0301-6226(03)00151-9.

Strandén, I., and Vuori, K. 2006. RelaX2: Pedigree analysis program. *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production*, Belo Horizonte, Brazil, 27–30.

Vandenplas, J., Veerkamp, R., Calus, M., Lidauer, M., Strandén, I., Taskinen, M., Schrauf, M., & Napel, J. T. 2022. 358. MiXBLUP 3.0 – software for large genomic evaluations in animal breeding programs. *Proceedings of the 12th World Congress on Genetics Applied to Livestock Production.* https://doi.org/10.3920/978-90-8686-940-4_358