

Simultaneous and Recursive Random Regression Models for Milk Yield and Somatic Cell Score in Canadian Holsteins

J. Jamrozik, J. Bohmanova and L.R. Schaeffer

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

* Corresponding author: jjamrozi@uoguelph.ca

Summary

Random regression models with simultaneous and recursive links between phenotypes for milk yield and somatic cell score (SCS) on the same test-day were fitted to Canadian Holstein data. Heterogeneity of structural coefficients was allowed for across (the first 3 lactations) and within (4 days in milk intervals) lactation. Model comparisons indicated superiority of simultaneous models over recursive and standard multiple-trait models. A moderate heterogeneous (both across and within lactation) negative effect of SCS on milk yield and a smaller positive reciprocal effect of SCS on milk yield were estimated in the most plausible specification. Estimates of genetic parameters on a daily basis differed while rankings of bulls and cows for 305d milk yield, average daily SCS and milk lactation persistency remained the same among models. No apparent benefits are expected due to fitting causal phenotypic relationships between milk yield and SCS in the random regression TD model for genetic evaluation purposes.

Keywords: milk yield, somatic cell score, random regressions, structural equation model

Introduction

Phenotypic relationships between milk yield and somatic cell score (SCS) can be explained by different mechanisms. An elevated level of SCS may indicate an infection of the udder that could have an adverse effect on milk production (*infection* effect). High producing cows are more prone to mastitis that causes higher level of SCS in the cow's milk (*stress* effect). On the other hand, higher milk yield may cause a decrease in somatic cell concentration at the same level of infection (*dilution* effects). The relationships between milk yield and SCS may therefore involve recursive or simultaneous effects that cannot be accounted for properly by genetic and environmental correlations only.

Structural equation models (SEM) (Gianola and Sorensen, 2004) allow for modeling causal pathways between phenotypes. In a two-trait system (traits X and Y), *simultaneous* effect refers to the presence of reciprocal direct effect between traits ($X \rightarrow Y$ and $X \leftarrow Y$, with possible different strengths of associations),

whereas *recursive* effects postulate that one trait affects the other directly ($X \rightarrow Y$) but the reciprocal link does not occur.

Applications of SEM to milk yield and SCS in dairy cattle have been limited. Research indicated a negative effect of SCS on milk yield. The negative phenotypic relationship between milk and SCS is more likely caused by an infection effect than by a dilution effect. De los Campos *et al.* (2006) investigated relationships between SCS and milk yield in the first lactation of Norwegian Red cows. The Bayesian Information Criterion favored the model with an effect of SCS on milk yield over 3 other models fitted. Wu *et al.* (2007) inferred relationships between first lactation SCS and milk yield in the same population. Results suggested a large negative effect of SCS on milk yield and a small reciprocal effect. The effects were larger in the first 60 d after calving than in the following 60 d period. Heritability estimates from SEM were similar to those from the usual mixed model, but some genetic correlations differed considerably among models.

Causal relationships between milk yield and SCS may vary among lactation. They may also show heterogeneity depending on the stage of lactations. If SEM are deemed superior over the standard multiple-trait models and if they could be applied to a routine genetic evaluation, the resulting estimated breeding values (EBV) would be more accurate than the current multiple-trait estimates. The objective was to compare several TD models with random regressions and with different specifications of recursive and simultaneous relationships between the first 3 lactation milk yield and SCS of Canadian Holstein cows.

Material and Methods

Data

Holstein TD data with calving years from 1988 to 2007 were used. A small, computationally manageable, data set was selected by random sampling of 59 herds with minimum of 50 cows each, and with minimum of 7 TD records per cow. Data were 181,386 TD milk and SCS records from the first three lactations of 10,832 cows. Days in milk (DIM) were from 5 to 305. Each cow had both milk yield and SCS recorded on a given TD. Pedigree data included 28,557 animals.

Models

The general form of equations for the SEM describing milk yield (**M**) and SCS for the *i*-th cow on the *j*-th TD in lactation *k* can be expressed as:

$$y_{ijkM} = \lambda_{M \leftarrow SCS}^{(k)} y_{ijkSCS} + \mathbf{x}'_{ijkM} \mathbf{b}_{ijkM} + \mathbf{z}'_{ijkM} \mathbf{a}_i + \mathbf{z}'_{ijkM} \mathbf{pe}_i + e_{ijkM}$$

and

$$y_{ijkSCS} = \lambda_{M \rightarrow SCS}^{(k)} y_{ijkM} + \mathbf{x}'_{ijkSCS} \mathbf{b}_{ijkSCS} + \mathbf{z}'_{ijkSCS} \mathbf{a}_i + \mathbf{z}'_{ijkSCS} \mathbf{pe}_i + e_{ijkSCS},$$

where y_{ijkM} and y_{ijkSCS} are observed phenotypes for milk yield and SCS, respectively; $\lambda_{M \leftarrow SCS}^{(k)}$ and $\lambda_{M \rightarrow SCS}^{(k)}$ are structural parameters for lactation *k*; $\mathbf{x}_{ijkM}$ and $\mathbf{x}_{ijkSCS}$ are incidence vectors for fixed effects; $\mathbf{b}_{ijkM}$ and $\mathbf{b}_{ijkSCS}$ are

the vectors of fixed effects; $\mathbf{a}_i$ and $\mathbf{pe}_i$ are the vectors of random additive genetic and permanent environmental (PE) regression coefficients for cow *i*; $\mathbf{z}_{ijkM}$ and $\mathbf{z}_{ijkSCS}$ are vectors of covariates accounting for DIM; e_{ijkM} and e_{ijkSCS} are the residuals. Fixed effects were the same for milk yield and SCS and included herd – test-day (HTD) effect and fixed regression within region-age-season of calving classes. All regressions were functions of DIM and were modeled with orthogonal Legendre polynomials of order 4. Heterogeneous residual co-variances were allowed in 4 DIM intervals: 5-45, 46-115, 116-265 and 266-305. Residuals for observations taken at different DIM were uncorrelated. Assume that elements of $\mathbf{a}_i$ and $\mathbf{p}_i$ are ordered as traits, within regression coefficients, within lactations and within animals, and let $\text{var}(\mathbf{a}_i) = \mathbf{G}$, $\text{var}(\mathbf{p}_i) = \mathbf{P}$ and $\text{var}([e_{ijkM}, e_{ijkSCS}]) = \mathbf{R}$ denote genetic, PE and residual covariance matrices, respectively.

The system of equations can be written in matrix notation as:

$$\Lambda^{(k)} \mathbf{y}_{ijk} = \mathbf{X}_{ijk} \mathbf{b} + \mathbf{Z}_{ijk} \mathbf{a}_i + \mathbf{Z}_{ijk} \mathbf{pe}_i + \mathbf{e}_{ijk}, \quad [1]$$

$$\text{where } \Lambda^{(k)} = \begin{bmatrix} 1 & -\lambda_{M \leftarrow SCS}^{(k)} \\ -\lambda_{M \rightarrow SCS}^{(k)} & 1 \end{bmatrix}.$$

Parameters of this model (**b**, **a**, **pe**, **G**, **P**, and **R**) are called ‘system’ parameters. When $\Lambda^{(k)}$ is a matrix of full rank, equation [1] can be re-written as

$$\mathbf{y}_{ijk} = \Lambda^{(k)-1} \mathbf{X}_{ijk} \mathbf{b} + \mathbf{Z}_{ijk} \Lambda^{(k)-1} \mathbf{a}_i + \mathbf{Z}_{ijk} \Lambda^{(k)-1} \mathbf{pe}_i + \Lambda^{(k)-1} \mathbf{e}_{ijk} = \mathbf{X}_i^* \mathbf{b} + \mathbf{Z}_{ijk} \mathbf{a}_i^* + \mathbf{Z}_{ijk} \mathbf{pe}_i^* + \mathbf{e}_{ijk}^*,$$

with $\text{var}(\mathbf{a}_i^*) = \mathbf{G}^* = (\mathbf{I} \otimes \Lambda^{(k)-1}) \mathbf{G} (\mathbf{I} \otimes \Lambda^{(k)-1})'$, $\text{var}(\mathbf{pe}_i^*) = \mathbf{P}^* = (\mathbf{I} \otimes \Lambda^{(k)-1}) \mathbf{P} (\mathbf{I} \otimes \Lambda^{(k)-1})'$ and $\text{var}(\mathbf{e}_{ijk}^*) = \mathbf{R}^* = \Lambda^{(k)-1} \mathbf{R} \Lambda^{(k)-1}$, and **I** is an identity matrix of order equal to number of regression coefficients times number of lactations. Genetic parameters for traits can then be calculated in a usual manner from elements **G**^{*}, **P**^{*} and **R**^{*}. Similarly, breeding values for animal *i* ($\mathbf{a}_i^*$) can be derived from the system parameter ($\mathbf{a}_i$) as $\mathbf{a}_i^* = (\mathbf{I} \otimes \Lambda^{(k)-1}) \mathbf{a}_i$.

Seven specific models were fitted in this study and they were:

M-SCS: standard multiple-trait model,
M<-> SCS: simultaneous model (both $\lambda^{(k)}_{M<-SCS}$ and $\lambda^{(k)}_{M->SCS}$ allowed to be different from 0),
M<- SCS model with recursive effect from SCS to milk ($\lambda^{(k)}_{M->SCS}=0$),
M-> SCS: model with recursive effect from milk to SCS milk ($\lambda^{(k)}_{M<-SCS}=0$),
M<-> SCS (het): heterogeneous simultaneous model,
M<- SCS (het): model with heterogeneous recursive effect from SCS to milk,
M-> SCS (het): model with heterogeneous recursive effect from milk to SCS milk.

For heterogeneous (het) models, possible different values of both $\lambda^{(k)}_{M<-SCS}$ and $\lambda^{(k)}_{M->SCS}$ were postulated in four DIM intervals within each lactation: 5-45, 46-115, 116-265 and 266-305 DIM (the same as used for definition of heterogeneous residual co-variances). Heterogeneous matrices of structural coefficients generate heterogeneous co-variance matrices ($\mathbf{G}^*$, $\mathbf{P}^*$ and $\mathbf{R}^*$) and heterogeneous location parameters of the model.

Methods

All models were fitted using Markov Chain Monte Carlo methods and the algorithms for SEM followed the procedures of Gianola and Sorensen (2004) and Wu *et al.* (2007). Restrictions were imposed on SEM to achieve identification of parameters. In the simultaneous models, residual co-variances were set to 0 and residual variances for SCS were assumed constant. In addition, the effects of fixed regressions for milk yield and HTD effects for SCS were removed from the respective models. Data for these models were pre-corrected for the effects that were removed from the model using estimates of fixed regression coefficients for milk and HTD effect for SCS from the multiple-trait model (M-SCS). Restrictions for recursive models required setting to 0 all residual co-variances between milk and SCS.

A total of 120,000 samples were generated for each model and 100,000 were used to estimate posterior means and standard

deviations for model parameters. Models were compared by Bayes Factors defined in Kass and Raftery (1995). The log marginal likelihood (LML) was estimated by the harmonic mean of the likelihood values from the Gibbs chain. Population co-variance components and selected genetic parameters were derived from the estimates of system co-variances. Estimated breeding values of bulls with at least 20 daughters with the data (265 sires) and cows with at least one TD record (10,832 cows) were calculated for each lactation for the 305d milk yield, average daily SCS and persistency of lactation using posterior samples of genetic regression coefficients. Persistency of lactation was defined as a difference in milk yield between days 280 and 60 of lactation. Combined EBV were derived as averages of within lactation EBV for those 3 traits. Estimated breeding values from different models were compared in terms of means, SD and product-moment correlations.

Results

Model comparison criterion indicated superiority of simultaneous models for milk yield and SCS over their competitors. The model with heterogeneous within-lactation structural coefficients was the most plausible specification (LML=-648,973); the standard multiple-trait model (LML=-657,427) outperformed both recursive models with a single structural effect per lactation (LML=-657,524 for M<SCS and LML=-657,520 for M->SCS).

Simultaneous models indicated a negative effect of SCS on milk yield and a smaller positive effect of milk yield on SCS (Table 1). Slightly smaller direct effects of SCS on milk yield was estimated with the recursive model M<-SCS, and a negative effect of milk yield on SCS was observed in M->SCS models. All structural coefficients for all models were significantly different from 0 (the 99% credible interval did not include 0). Absolute values of all structural coefficients increased with lactation number. Heterogeneity of structural coefficients was documented for the majority of both across and within lactation comparisons.

Genetic, PE and residual variances, and heritabilities for the constant and linear terms of lactation curve did not differ much between models and showed very little heterogeneity in DIM intervals within lactations. Genetic and PE correlations between milk yield and SCS for the first two regression coefficients were also very similar between models with slightly more heterogeneity of estimates between DIM intervals. Relatively larger differences among models were observed for estimates of daily genetic parameters. The pattern of genetic correlations was similar to those reported in other studies: from mostly positive values in lactation 1 (Figure 1) to all negative correlations in lactation 3. Paired t-test indicated significant differences ($P < 0.001$) between MT and all SEM models. PE correlations followed similar pattern of differences among models. Phenotypic correlations between milk and SCS on the same DIM for lactation 1 are in Figure 2. All correlations for all lactations were negative; there was a slight decrease in estimates with lactation number. Absolute differences among models for these correlations were smaller than for the corresponding daily genetic and PE parameters. SEM estimates were significantly different from the M<-SCS model parameters, with the exception of model M<-SCS (het) for the first two lactations, and model M<-SCS for the third lactation. Daily heritability curves showed similar shapes between models (results not shown). Differences between SEM models and the multiple-trait model were significant except for M<-SCS and M<-SCS (het) comparisons.

Correlations between EBV from different models for within lactation and combined 305d milk yield, daily SCS and lactation persistency were all larger than 0.99 for both bulls and cows. Table 2 gives estimates of average EBV and their standard deviations for bulls and cows for combined traits.

Discussion

Simultaneous and recursive models for the first 3 lactation TD data of milk and SCS were fitted accounting for the longitudinal nature of both traits. Overall results, in terms of model plausibility, confirmed findings of Wu *et al.*

(2007). Phenotypic relationships between milk and SCS in all three lactations seem to be driven by a combination of two mechanisms. An infection effect caused a decrease in milk yield for cows with a higher SCS concentration, and the stress effect resulted in an increase of SCS for higher producing cows. The absolute values of structural coefficient estimates were smaller than results of other authors. Wu *et al.* (2007) reported a decrease in milk yield of about 1.8 kg for every unit of SCS, and an increase of about 0.04 points for SCS for every kilogram of milk yield during the first 120 d of lactation. Estimates of the effect of SCS on milk yield in our study, however, were in good agreement with phenotypic estimates of milk losses caused by SCS in Canadian Holsteins (Dürr *et al.*, 2008). They reported milk losses per unit increase from 0.33 to 0.55 kg/day in the first lactation, and from 0.77 to 1.78 kg/day in later lactations. There was an evident heterogeneity of all structural coefficients across and within lactation. The intensity of structural relationships between milk yield and SCS increased with lactation number. The largest impact of either trait on the other occurred at DIM around the peak of lactation. Wu *et al.* (2007) reported larger values for direct effects in the first 60 d of lactation compared with the subsequent 60 d period. Modeling the heterogeneity of within lactation structural coefficients was attempted in our study by fitting them separately in four DIM intervals within each lactation. The selection of those intervals followed the pattern of changes in residual co-variances for milk yield. Definition of intervals could be different, but drastic changes in the pattern of structural coefficients for other scenarios are not expected.

Modeling causal relationships between TD milk yield and SCS did not cause dramatic changes in estimated co-variance components and respective genetic parameters for regression coefficients. Larger impact, however, with statistically significant differences between models were observed for variances, heritabilities and correlations expressed on a daily basis. Results from other studies indicated changes in genetic (Wu *et al.*, 2007) and phenotypic (de los Campos *et al.*, 2006) correlations between milk yield and SCS for simultaneous and recursive repeatability

models when compared with multiple-trait estimates. Heterogeneity of structural parameters resulted in heterogeneity of genetic parameters. Co-variance matrices from the heterogeneous models may take different values (up to 12 possibilities in our study) depending on the stage of lactation. This also applies to genetic parameters derived from those matrices.

No changes in rankings of bulls and cows for milk yield, SCS and lactation persistency would be expected, when causal relationships between milk yield and SCS were accounted for in the evaluation model. This suggested that no practical benefits would be achieved when using models with causal relationships between milk and SCS for genetic evaluation purposes. Variability of EBV from different models was not influenced by the specific parameterization. An average genetic level for selected groups of bulls and cows increased (decreased) for milk yield (SCS). This could indicate possible changes in estimates of genetic trends for these traits, when compared with estimates from the standard models.

Conclusions

A random regression model with heterogeneous simultaneous relationships between TD milk yield and SCS for the first three lactations of Canadian Holstein cows was superior over a standard multiple-trait model and over models with recursive specifications between traits. Estimates of genetic parameters on a daily basis differed among models while rankings of bulls and cows for 305d milk yield, average daily SCS and milk lactation persistency remained the same.

Acknowledgements

The authors are grateful to DairyGen and the Natural Science and Engineering Research Council for their financial support and CDN for providing the data.

References

- Campos de los, G., Gianola, D. & Heringstad, B. 2006. A structural equation model for describing relationships between somatic cell score and milk yield in first-lactation dairy cows. *J. Dairy Sci.* 89, 4445-4555.
- Dürr, J.W., Cue, R.I., Monardes, H.G., Moreo-Mendez, J. & Wade, K. 2008. Milk losses associated with somatic cell count per breed, parity and stage of lactation in Canadian dairy cattle. *Livest. Sci.* 117, 225-232.
- Gianola, D. & Sorensen, D. 2004. Quantitative genetic models for describing simultaneous and recursive relationships between phenotypes. *Genetics* 167, 1407-1424.
- Heringstad, B. 2009. Structural equation models in animal breeding. Symposium 'Statistical Genetics of Livestock for the Post-Genomic Era', Madison, Wisconsin, May 4-6, 2009.
- Kass, R.E. & Raftery, A.E. 1995. Bayes factors. *J. Am. Stat. Assoc.* 90, 773-795.
- Schaeffer, L.R., Jamrozik, J., Kistemaker, G.J. & Van Doormaal, B.J. 2000. Experience with a test-day model. *J. Dairy Sci.* 83, 1135-1144.
- Wu, X-L., Heringstad, B., Chang, Y-M., de los Campos, G. & Gianola, D. 2007. Inferring relationships between somatic cell score and milk yield using simultaneous and recursive models. *J. Dairy Sci.* 90, 3508-3521.

Table 1. Estimates of structural coefficients (λ) between milk and SCS in 4 DIM intervals¹ for the first 3 lactations (posterior standard deviations in brackets), by model.

Model	Lactation	$\lambda_{M \leftarrow SCS}$				$\lambda_{M \rightarrow SCS}$			
		1	2	3	4	1	2	3	4
M<->SCS	1	-.38 (.022)				.011 (.003)			
	2	-.52 (.031)				.009 (.003)			
	3	-.68 (.051)				.019 (.004)			
M<-SCS	1	-.30 (.013)							
	2	-.43 (.018)							
	3	-.45 (.023)							
M->SCS	1					-.051 (.002)			
	2					-.051 (.002)			
	3					-.044 (.002)			
M<->SCS (het)	1	-.39 (.034)	-.47 (.024)	-.36 (.023)	-.37 (.028)	.008 (.003)	.015 (.003)	.013 (.003)	.015 (.003)
	2	-.48 (.043)	-.67 (.029)	-.52 (.027)	-.47 (.035)	.007 (.002)	.012 (.002)	.011 (.002)	.009 (.002)
	3	-.81 (.065)	-.77 (.049)	-.66 (.048)	-.69 (.051)	.018 (.003)	.023 (.003)	.021 (.003)	.020 (.003)
M<-SCS (het)	1	-.29 (.029)	-.37 (.016)	-.27 (.014)	-.27 (.022)				
	2	-.35 (.039)	-.55 (.022)	-.40 (.019)	-.35 (.029)				
	3	-.54 (.047)	-.51 (.028)	-.41 (.025)	-.45 (.036)				
M->SCS (het)	1					-.054 (.002)	-.049 (.002)	-.050 (.002)	-.049 (.002)
	2					-.052 (.002)	-.050 (.002)	-.051 (.002)	-.053 (.002)
	3					-.044 (.002)	-.041 (.002)	-.044 (.002)	-.045 (.003)

¹1 = 5-45 DIM, 2 = 46-115 DIM, 3 = 116-265 DIM, 4 = 266-305 DIM

Table 2. Estimates of average breeding values and their standard deviations for bulls (N=265) and cows (N=10,832) for combined milk yield, SCS and lactation persistency.

Model	Trait											
	Milk				SCS				Persistency			
	Bulls		Cows		Bulls		Cows		Bulls		Cows	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
M-SCS	654	642	522	654	.15	.35	.13	.34	-.08	2.33	-.59	2.45
M<->SCS	731	644	593	660	.12	.36	.09	.34	.65	2.32	.05	2.44
M<-SCS	682	645	547	658	.15	.35	.12	.34	.03	2.32	-.50	2.43
M->SCS	679	645	544	659	.14	.35	.11	.34	.03	2.33	-.50	2.46
M<->SCS (het)	730	646	592	662	.10	.36	.08	.34	.35	2.31	-.22	2.43
M<-SCS (het)	665	643	532	656	.20	.35	.17	.34	-.11	2.31	-.61	2.43
M->SCS (het)	667	643	534	658	.18	.35	.15	.34	-.12	2.34	-.63	2.47

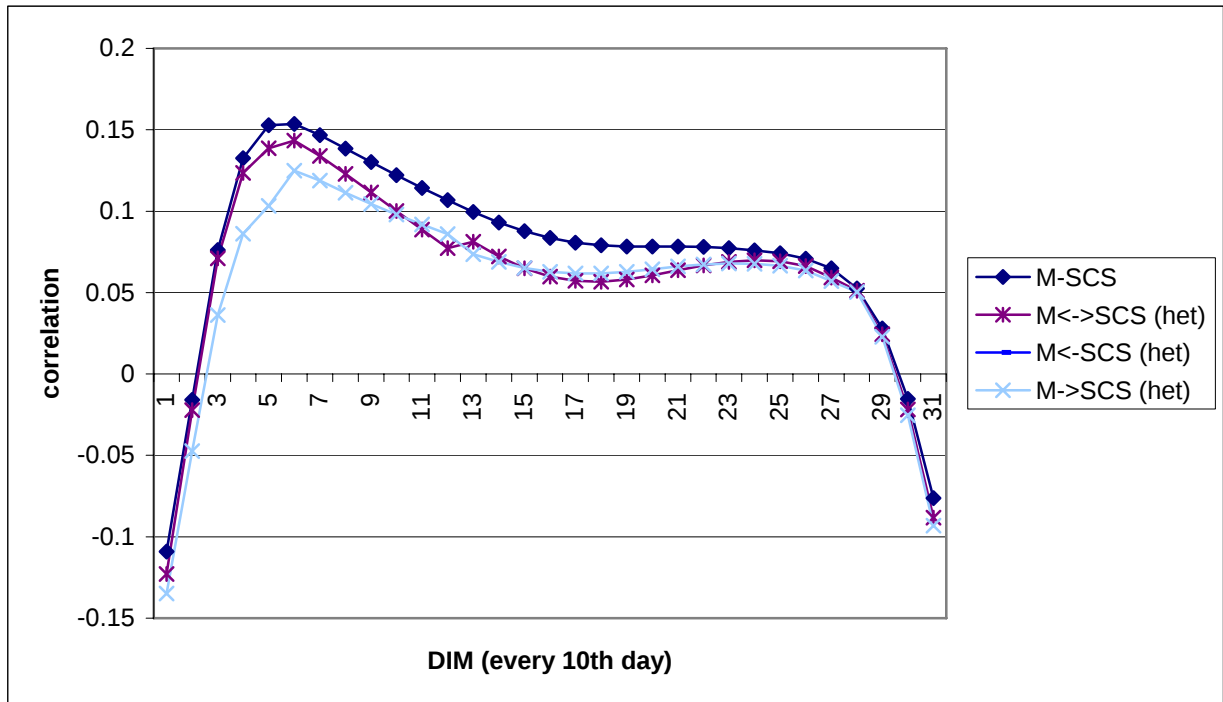


Figure 1. Daily genetic correlations between first lactation milk and SCS for multiple-trait and heterogeneous models.

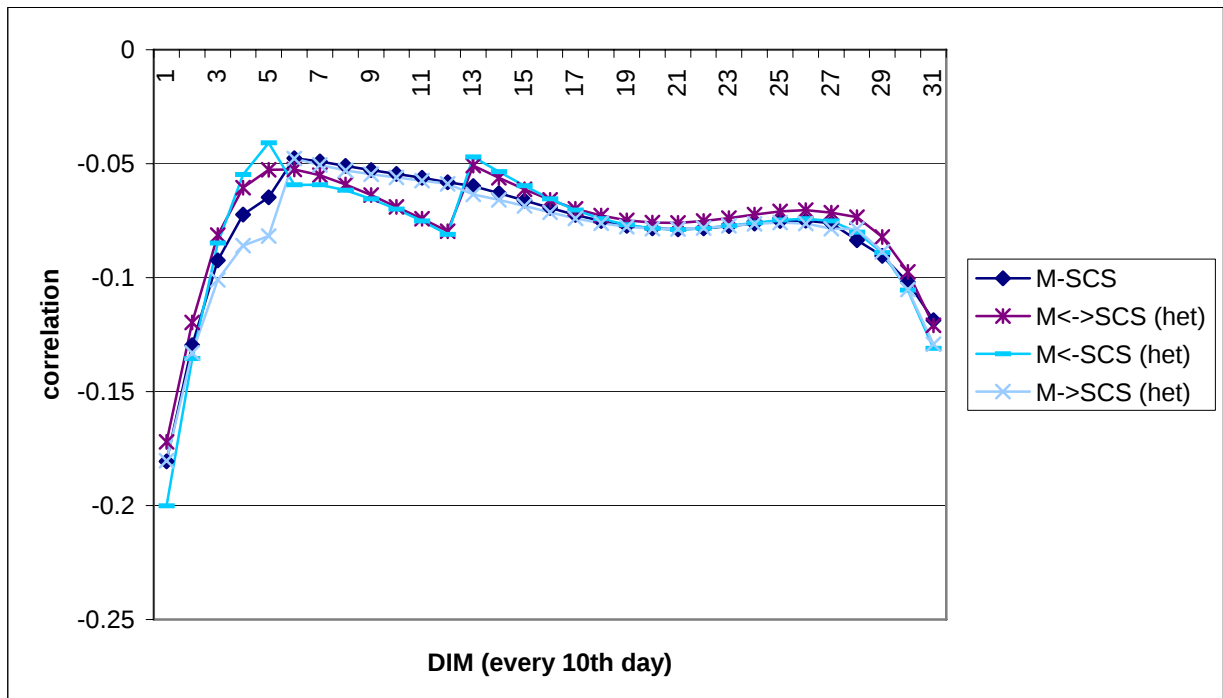


Figure 2. Daily phenotypic correlations between first lactation milk and SCS for multiple-trait and heterogeneous models.