

Interpretation of residual feed intake by phenotypic recursiveness in dairy cattle: A simulation study

X.-L. Wu^{1,2}, K.L. Parker Gaddis¹, H.D. Norman¹, J. Burchard¹, E.L. Nicolazzi¹, E.E. Connor³, J.B. Cole⁴ and J. Durr¹

¹Council on Dairy Cattle Breeding, Bowie, MD 20716, USA

²Department of Animal and Dairy Sciences, University of Wisconsin, Madison, WI 53706, USA

³Department of Animal and Food Sciences, University of Delaware, Newark, DE 19716, USA

⁴USDA, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD 20705-2350, USA

There has been an increasing interest in the use of residual feed intake (RFI) as a measure of net feed efficiency in recent years. RFI is estimated by the residual from a linear regression, which treats feed intake as a linear regression function of key energy sinks. Re-arranging the linear regression for RFI suggested an alternative interpretation of RFI, which postulates the RFI phenotypes as resulting from the recursive effects of energy sinks on dry matter intake. This leads to a Bayesian recursive structural equation model for a direct genetic evaluation of RFI. A simplified algorithm was proposed to compute the Bayesian recursive model. Perspectives on the recursive model were taken from a simulation study.

Abstract

Keywords: Dry matter intake, Feed efficiency, milk, SNP, structural equation model.

Feed efficiency is an important trait for dairy cattle because feed costs comprise almost half of the total costs associated with dairy production. Residual feed intake (RFI), which was initially proposed by Koch *et al.* (1963), is becoming increasingly popular as a measure of net feed efficiency in the past decade. The original idea of genetic evaluation on RFI consists of two stages (Berry and Crowley, 2013). In the first stage, dry matter intake (DMI) is fitted by single-trait linear regression (LR) encompassing energy sinks with or without relevant factors, and the residuals are taken to be the RFI phenotypes (Løvendahl *et al.*, 2018; Templeman *et al.*, 2015). In the second-stage, RFI is fitted by a mixed-effects model, which includes additive individual genetic effects for genetic evaluation. Combining both stages leads to the one-step model, eliminating the need to specifically estimate the residuals as the RFI phenotypes (Templeman *et al.*, 2015).

Introduction

Fitting phenotypes as regressor variables is criticized because standard regression models assume that regressor variables have been measured precisely or observed without error (Lu *et al.*, 2015). In reality, however, phenotypes are subject to measurement errors. Multiple-trait models have been used, which bypass the above critics, but they represent indirect methods because RFI's genetic values are obtained through a follow-up partial regression procedure based on the estimated variance-covariance components (Kennedy *et al.*, 1993; Lu *et al.*, 2015; Templeman and Lu,

2020). By re-arranging the regression models for RFI, we came with an alternative, causative interpretation of RFI. This led to a Bayesian recursive structural equation model (RSEM) which allows for directly predicting genetic values and estimating genetic parameters for RFI and all the involving traits jointly.

Bayesian recursive structural equation model

Consider a single animal, say i . Following Løvendahl *et al.* (2018), we let the energy sink model include metabolic body weight ($MBW=BW^{0.75}$), energy-corrected milk (ECM), and change in body weight (ΔBW). That is

$$y_{i1} = \mu_1 + \lambda_{12}y_{i2} + \lambda_{13}y_{i3} + \lambda_{14}y_{i4} + r_i \quad (1)$$

Here, μ_1 the overall mean and λ_{1j} qualifies the rate of the change with DMI (denoted by y_1) with respect to the three energy sink traits, denoted by y_2 (MBW), y_3 (ECM), and y_4 (ΔBW), respectively. The residual (r_i) as the RFI phenotype is then described by a mixed-effects model:

$$r_i = \mathbf{x}'_{ir}\boldsymbol{\beta}_r + \mathbf{z}'_{ir}\mathbf{a}_r + e_{ir} \quad (2)$$

where $\boldsymbol{\beta}$ is a vector of “fixed” effects of an appropriate length; \mathbf{a}_r is a vector containing the “random” additive genetic effects for all individuals; \mathbf{x}_{ir} and \mathbf{z}_{ir} are the corresponding incidence vectors relating the RFI phenotypes to the fixed and random effects, and e_{ir} is an error term. Note that “random” permanent environmental effects are not considered in model (2), but they may be relevant in real applications.

Combining equations (1) and (2) and moving all the y variables to the left-hand side leads to the following one-step model:

$$y_{i1} - \sum_{j=2,3,4} \lambda_{1j} y_{ij} = \mu_1 + \mathbf{x}'_{ir}\boldsymbol{\beta}_r + \mathbf{z}'_{ir}\mathbf{a}_r + e_{ir} \quad (3)$$

The above is recognized as a recursive structural equation between DMI and the three energy sink traits (MBW, ECM, and ΔBW) for the i th individual. It postulates that the RFI phenotypes result from the recursive effects from energy sinks to DMI, but the feedback or simultaneous effects are assumed to be non-existent. Because the model parameters in (3) pertain to RFI, not DMI, hence it can provide a direct evaluation of RFI.

To complete the recursive structural equation model, we define mixed-effect models for the energy sinks as follows

$$y_{ij} = \mu_j + \mathbf{x}'_{ij}\boldsymbol{\beta}_j + \mathbf{z}'_{ij}\mathbf{a}_j + e_{ij}, \text{ for } j = 2, 3, 4. \quad (4)$$

The recursive structural equation model that combines equations in (3) and (4) for the i th individual is the following:

$$\Lambda y_i = \mu + X_i \beta + Z_i \alpha_i + e \quad (5)$$

Here, we have

$$y_i = (y_{i1}, y_{i2}, y_{i3}, y_{i4})';$$

$$\mu = (\mu_1, \mu_2, \mu_3, \mu_4)', \text{ and}$$

$$e = (e_1, e_2, e_3, e_4)';$$

The vectors for the fixed and random effects were sorted by traits within animals. For example, we have

$$\beta = (\beta_{11} \ \beta_{12} \ \beta_{13} \ \beta_{14} \dots \beta_{k1} \ \beta_{k2} \ \beta_{k3} \ \beta_{k4})', \text{ where } k \text{ is the total number of fixed effect levels.}$$

Assuming a full loading of each fixed effect on all the four traits, we have

$$X_i = x_i' \otimes I, \text{ where } x_i' = \dots = x_{i4}' = x_i', I \text{ is a } 4 \times 4 \text{ identity matrix, and } \otimes \text{ is the Kronecker product operator.}$$

$$\text{Similarly, we have } \alpha = (\alpha_{11} \ \alpha_{12} \ \alpha_{13} \ \alpha_{14} \dots \alpha_{n1} \ \alpha_{n2} \ \alpha_{n3} \ \alpha_{n4})'$$

$$\text{where } n \text{ is the total number of animals, and } Z_i = z_i' \otimes I, \text{ where } z_i' = \dots = z_{i4}' = z_i'.$$

Finally, the structural matrix (Λ) defines the phenotypic relationships between DMI and energy sinks:

$$\Lambda = \begin{pmatrix} 1 & -\lambda_{12} & -\lambda_{13} & -\lambda_{14} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (6)$$

Bayesian modeling of structural equation models via Markov chain Monte Carlo (MCMC) implementation is described by Gianola and Sorensen (2004) and Wu et al. (2017, 2018). Here, we briefly describe a simplified algorithm to compute the recursive model for RFI. Assume zero genetic and residual covariances, and hence zero phenotypic covariances, between RFI and energy sinks, the posterior inferring of structural coefficients does not involve any unknown parameters for the energy sink traits. Assign a multivariate normal prior distribution for all unknown structural coefficients,

$$\lambda \sim MVN(\mathbf{1}\lambda_0, \mathbf{I}\sigma_p^2), \text{ where } \lambda = (\lambda_{12} \ \lambda_{13} \ \lambda_{14} \ \lambda_{15})'.$$

Then, given the sampled fixed and random effects (β , a , and p) for RFI and the residual variance-covariance matrix (R_0), the conditional posterior distribution of λ is a multivariate normal distribution:

$$\lambda | else \sim MVN(\hat{\lambda}, V_{\lambda}) \tag{7}$$

where

$$\hat{\lambda} = V_{\lambda} (\sum_{i=1}^n Y_i' R_i^{-1} (y_i - X_i \beta_r - Z_i a_r) + 1 \tau_{\lambda}^{-2} \lambda_0)$$

$$V_{\lambda} = (\sum_{i=1}^n Y_i' R_i^{-1} Y_i + \tau_{\lambda}^{-2} I)^{-1}$$

In the above, $y_i = (y_{i1} \ y_{i2} \ y_{i3} \ y_{i4})'$ and Y_i is a working 4x(4-1) matrix constructed by noting that $\Lambda y_i = y_i - Y_i \lambda$ for each individual.

That is,

$$Y_i = \begin{pmatrix} y_{i2} & y_{i3} & y_{i3} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Furthermore, the posterior inference of the model parameters for RFI, given the structural equation coefficients, does not involve any unknown parameters for the energy sinks. Thus, a simplified algorithm for computing the recursive structural equation model is proposed as follows:

The fixed effects, random effects (β_r and a_r), and variance-covariance matrices G_r and R_r for the energy sinks are estimated based on a standard multiple-trait mixed-effects model, independent of the computed RFI phenotypes. Computing the submodel for the energy sinks can be implemented through Markov chain Monte Carlo sample by iteratively sampling unknown parameters from their conditional posterior distributions, or implemented by REML.

The unknown structural coefficients were sampled iteratively from the multivariate normal distribution (7). Then, given the sampled structural coefficients and the

computed RFI phenotypes ($r_i = y_{i1} - \sum_{j=2,3,4} \lambda_{1j} y_{ij}$), the fixed and random effects and the variance components ($\sigma_{a_r}^2$, $\sigma_{p_r}^2$, and $\sigma_{e_r}^2$) for RFI are sampled from their respective conditional posterior distributions based on a standard single-trait mixed effects model.

Note that MCMC simulation was only necessary for sampling the structural coefficients and unknown parameter for RFI only. This drastically simplified the model computing when dealing with a large dataset. Also note that the covariances between RFI and energy sinks are fixed to zeros, but the covariances between DMI and energy sinks are not, which are computed as follows:

$$G_0 = A^{-1} G_e A^{-1}, \text{ and } R_0 = A^{-1} R_e A^{-1}, \text{ where } G_e = \begin{pmatrix} \sigma_{a_e}^2 & 0' \\ 0 & \sigma_{e_e}^2 \end{pmatrix}, R_e = \begin{pmatrix} \sigma_{a_e}^2 & 0' \\ 0 & \sigma_{e_e}^2 \end{pmatrix}$$

Simulation results

We simulated standardized phenotypes of DMI, MBW, ECM, and ?BW based on an actual pedigree for 908 cows. These animals were progenies derived from 125 sires and 477 dams and raised in the USDA Beltsville Agricultural Research Center (BARC) Dairy Herd (Beltsville, Maryland, USDA). For simplicity, the fixed effects included only the overall mean ($\mu_j=0$) for each trait, and the random effects included individual additive genetic effects plus the residuals. The simulated genetic and residual variance-covariance components were shown below:

$$G = \begin{pmatrix} 0.199 & 0.205 & 0.191 & 0.012 \\ 0.205 & 0.221 & 0.033 & 0.037 \\ 0.191 & 0.033 & 0.287 & -0.005 \\ 0.012 & 0.037 & -0.005 & 0.040 \end{pmatrix}; R = \begin{pmatrix} 0.581 & 0.236 & 0.311 & 0.179 \\ 0.236 & 0.334 & 0.242 & 0.144 \\ 0.311 & 0.242 & 0.671 & -0.022 \\ 0.179 & 0.144 & -0.022 & 0.96 \end{pmatrix}$$

We compared four models: (1) one-step linear regression (LR1), (2) two-stage linear regression (LR2), (3) recursive structural equation model (RSEM), and (4) multiple-trait mixed-effects model (MT). To set up an accurate bench model for comparison, the (co) variance components in the MT model were assumed to be known and took directly from the actual values. RSEM was implemented via MCMC simulation. Ten parallel chains (or replicates) were run for each model, each of 2000 iterations, after a burn-in of 1000 iterations and thinned every two draws. The means and standard deviations of parameter estimates were obtained as averages across the ten replicates.

All the MCMC chains covered well after 1,000 iterations. The estimated structural coefficients from RSEM corresponds closely to the partial regression coefficients from the one-step LR. The partial regression coefficients from two-stage LR agreed perfectly with the phenotypic partial regression from the MT model, but differed from the genetic partial coefficient based on the MT model. These results indicates that both RSEM and the single-trait linear regression approaches depicts RFI by the phenotypic relationships between DMI and energy sinks. Although there were some differences in partial regression coefficients between two-stage LR and one-step LR (or RSEM), the estimated additive genetic values are highly comparable between these models (Spearman’s correlation was 0.999). The correlations of the estimated RFI genetic were also highly comparable values between RSEM and MT, but the correlation was slightly lower than that between RSEM and a LR model. Thus, we concluded that RSEM resembled the LR model more than the MT. This was because both RSEM and the LR approaches inferred phenotype recursive relationships. For the MT model, RFI genetic values were inferred based on genetic relationships between DMI and energy sinks, which can differ from the phenotypic relationship. This, however, does not mean

Table 1. Estimated partial regression coefficients (structural coefficients) from linear regression (LR), recursive structural equation model (RSEM), and multi-trait, mixed-effects model (MT).

Energy sink	Two-stage LR	One-step LR	RSEM	MT	PMean	PSD	b.COVP	b.COVG	b.COVE
MBW	0.327	0.026	0.313	0.027	0.312	0.027	0.327	0.349	0.300
ECM	0.498	0.027	0.470	0.027	0.469	0.027	0.498	0.629	0.404
LW	0.127	0.026	0.137	0.025	0.137	0.025	0.127	0.049	0.151

SD = standard deviation; PSD = posterior standard deviation; PMean = Posterior mean. b.COVP, b.COVG, b.COVE = partial regression coefficient based on phenotypic, genetic, and residual (co)variances, respectively.

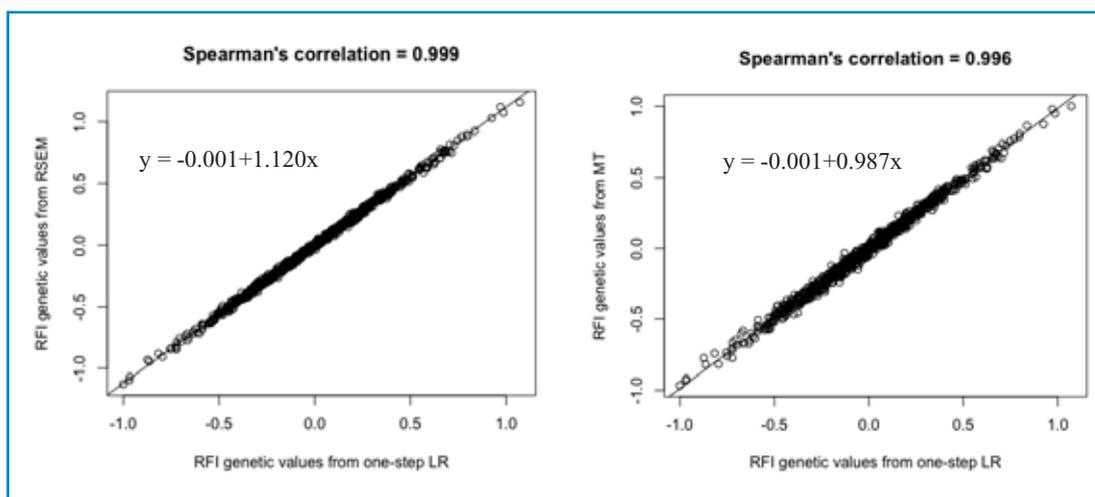


Figure 1. Comparing the estimated RFI genetic values from one-step linear regression (LR) and those from the recursive model (RSEM, left) or the multiple trait model (MT, right).

that RSEM is more accurate than the MT model, because the true RFI genetic values were unknown. It is possible to simulate RFI directly, but the simulated relationships depends on the assumptions, whether it favors one model or the other. Estimated heritability for RFI was 0.339 (two-stage LR), 0.348 (one-step LR), and 0.324 (RSEM).

In conclusion, we proposed RSEM as a direct genetic evaluation of RFI. Concerning the estimated RFI genetic values, RSEM was equivalent to the single-trait linear regression model, but RSEM expended the analytical capability to multiple traits with causative relationships assumed. RSEM can also provide the estimates of genetic parameters for RFI and all the involving traits, which are not discussed due to the page limitation. Finally, the recursive model extends very naturally to deal with heterogeneous structural coefficient matrices (Wu et al., 2007). Extending RSEM for RFI to genomic selection is just as straightforward, e.g., by replacing the additive relationship matrix with the genomic relationship matrix (VanRaden, 2008).

List of references

- Berry, D. P., and J. J. Crowley.** 2013. Cell Biology Symposium: genetics of feed efficiency in dairy and beef cattle. *J. Anim. Sci.* 91, 1594-613.
- Gianola, D., and D. Sorensen.** 2004. Quantitative genetic models for describing simultaneous and recursive relationships between phenotypes. *Genetics.* 167, 1407-1424.
- Koch, R. M., L. A. Swiger, D. Chambers, and K. E. Gregory.** 1963. Efficiency of feed use in beef cattle. *J. Anim. Sci.* 22, 486-494.