An alternative Interpretation of residual feed intake by phenotypic recursive relationships in dairy cattle

Xiao-Lin Wu,[1,2]* Kristen L. Parker Gaddis,[1] Javier Burchard,[1] H. Duane Norman,[1] Ezequiel Nicolazzi,[1] Erin E. Connor,[3] John B. Cole,[4] and Joao Durr[1]

[1] Council on Dairy Cattle Breeding, Bowie, MD 20716
[2] Department of Animal and Dairy Sciences, University of Wisconsin, Madison, WI 53706
[3] Department of Animal and Food Sciences, University of Delaware, Newark, DE 19716
[4] USDA, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD 20705-2350

* Corresponding author: nick.wu@uscdcb.com

JDS Communications 2021; 2:1-6.
https://doi.org/10.3168/jdsc.2021-0080

Appendix 1 A simplified algorithm to implement the recursive structural equation model for RFI via Markov chain Monte Carlo simulation

The Bayesian implementation of the recursive model for RFI follows Gianola and Sorensen (2004) and Wu et al. (2007, 2008, 2010). A simplified algorithm is described below. The RFI phenotypes (i.e., $y_{i1} - \sum_{t=2}^{k} \lambda_{1t}\, y_{it}$) are uncorrelated with the phenotypes of energy sinks (Kennedy et al., 1993). According to the path theory, a zero phenotypic correlation (r_p) between RFI and an energy sink (indexed by j, for $j = 2, \ldots, k$) implies that either 1) $(1 - h_{RFI})(1 - h_j) r_{e_{RFI} e_j} = -h_{RFI} h_j r_{a_{RFI} a_j}$ or 2) $r_{e_{RFI} e_j} = 0$ and $r_{a_{RFI} a_j} = 0$, where h stands for the square root of heritability, and $r_{a_{RFI} a_j}$ and $r_{e_{RFI} e_j}$ are the genetic and residual correlation, respectively, between RFI and energy sink j, assuming a total determination by these two components. We took the latter approach by forcing the genetic and residual covariance between RFI and energy sinks to be zeros, because we intended to have RFI as a measure of net feed efficiency, independent of energy sinks. That is,

$$\boldsymbol{G}_0 = \begin{pmatrix} \sigma^2_{a_1} & 0 & \ldots & 0 \\ 0 & \sigma^2_{a_2} & \ldots & \sigma_{a_2 a_k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \sigma_{a_k a_2} & \ldots & \sigma^2_{a_k} \end{pmatrix} \tag{1}$$

$$\boldsymbol{R}_0 = \begin{pmatrix} \sigma_{e_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{e_2}^2 & \dots & \sigma_{e_2 e_k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \sigma_{e_k e_2} & \cdot\cdot & \sigma_{e_k}^2 \end{pmatrix} \tag{2}$$

The conditional posterior distribution of structural coefficients does not depend on any unknown parameters of energy sinks, assuming zero genetic and residual correlations between RFI and energy sinks. This feature drastically simplifies the posterior inference of structural coefficient matrix and unknown parameters for RFI. Denote $\boldsymbol{\lambda} = (\lambda_{12}, \lambda_{13}, \cdots, \lambda_{1k})'$. We assumed a multivariate normal prior distribution (MVN) for $\boldsymbol{\lambda}$. That is, $\boldsymbol{\lambda}|\lambda_0, \tau^2 \sim MVN(\mathbf{1}\lambda_0, \mathbf{I}\tau^2)$, where $\mathbf{1}$ is a $(k-1) \times 1$ vector of ones, $\mathbf{I}$ is a $(k-1) \times (k-1)$ identity matrix, and λ_0 and τ^2 are hyper-parameters. Then, the conditional posterior distribution of $\boldsymbol{\lambda}$ is also a multivariate normal distribution (Gianola and Sorensen, 2004; Wu et al., 2007), independent of the equations for energy sinks. The conditional posterior means of $\boldsymbol{\lambda}$ are:

$$E(\boldsymbol{\lambda}|\boldsymbol{else}) = \begin{pmatrix} \sum_{i=1}^n y_{i2}^2 + \sigma_{e_1}^2 \tau^{-2} & \sum_{i=1}^n y_{i2} y_{i3} & \cdots & \sum_{i=1}^n y_{i2} y_{ik} \\ \sum_{i=1}^n y_{i3} y_{i2} & \sum_{i=1}^n y_{i3}^2 + \sigma_{e_1}^2 \tau^{-2} & \cdots & \sum_{i=1}^n y_{i3} y_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n y_{ik} y_{i2} & \sum_{i=1}^n y_{ik} y_{i3} & \cdots & \sum_{i=1}^n y_{ik}^2 + \sigma_{e_1}^2 \tau^{-2} \end{pmatrix}^{-1} \times \begin{pmatrix} \sum_{i=1}^n y_{i2} w_{i1} + \sigma_{e_1}^2 \tau^{-2} \lambda_0 \\ \sum_{i=1}^n y_{i3} w_{i1} + \sigma_{e_1}^2 \tau^{-2} \lambda_0 \\ \ddots \\ \sum_{i=1}^n y_{ik} w_{i1} + \sigma_{e_1}^2 \tau^{-2} \lambda_0 \end{pmatrix} \tag{3}$$

where **else** represents the data and all other unknown model parameters, and $w_{i1} = y_{i1} - (\mu_1 + \mathbf{x}'_{i1}\boldsymbol{\beta}_1 + \boldsymbol{z}'_{i1}\boldsymbol{a}_1)$, for $i = 1, \dots, n$. Similarly, the conditional posterior distribution of location parameters (i.e., fixed and random effects) and scaling parameters (variance components), respectively, for RFI does not involve any unknown parameters for energy sinks either. The conditional posterior means of the RFI location parameters are the following:

$$E\left(\begin{matrix} \mu_1 \\ \boldsymbol{\beta}_1 \\ \boldsymbol{\alpha}_1 \end{matrix} \middle| \boldsymbol{else}\right) = \begin{pmatrix} \mathbf{1}'\boldsymbol{y}_1^* \\ \boldsymbol{x}_1'\boldsymbol{y}_1^* + \sigma_{e_1}^2 \omega^{-2} \boldsymbol{\beta}_0 \\ \boldsymbol{z}_1'\boldsymbol{y}_1^* \end{pmatrix} \begin{pmatrix} \mathbf{1}'\mathbf{1} & \boldsymbol{x}_1 & \boldsymbol{z}_1 \\ \boldsymbol{x}_1' & \boldsymbol{x}_1'\boldsymbol{x}_1 + \sigma_{e_1}^2 \omega^{-2} \mathbf{I} & \boldsymbol{x}_1'\boldsymbol{z}_1 \\ \boldsymbol{z}_1' & \boldsymbol{z}_1'\boldsymbol{x}_1 & \boldsymbol{z}_1'\boldsymbol{z}_1 + \sigma_{e_1}^2 \sigma_{a_1}^{-2} \mathbf{A}^{-1} \end{pmatrix}^{-1} \tag{4}$$

where $\boldsymbol{y}_1^* = \begin{pmatrix} y_{11} - \sum_{j=2}^k \lambda_{1k} y_{1k} \\ \vdots \\ y_{n1} - \sum_{j=2}^k \lambda_{1k} y_{nk} \end{pmatrix}$, $\boldsymbol{1}$ is a $n \times 1$ vector of ones, and $\mathbf{I}$ is an identity matrix of appropriate dimensions. In the above, a flat prior is assumed for the overall mean. Multivariate normal prior distributions are assumed for fixed-effect and random-effects: $Pr(\boldsymbol{\beta}_1) =$

$MVN(\boldsymbol{\beta}_0, \omega^2\mathbf{I})$ and $Pr(\boldsymbol{a}_1) = MVN\left(\mathbf{0}, \mathbf{A}\sigma_{a_1}^2\right)$, where $\mathbf{A}$ is a numeric additive genetic relationship matrix, $\sigma_{a_1}^2$ is the RFI genetic variance, and $\boldsymbol{\beta}_0$ and ω^2 are hyper-parameters. Given inverse Chi-squared prior distributions to the genetic and residual variances, $Pr(\sigma_{u_1}^2) = Inv - Chi(v_a, v_a S_a^2)$ and Pr $(\sigma_{e_1}^2) = Inv - Chi(v_e, v_e S_e^2)$, their conditional distributions are also inverse Chi-squared distributions:

$$\sigma_{u_1}^2 | else \sim Inv - Chi(q + v_a, \boldsymbol{a}_1' \boldsymbol{A}^{-1} \boldsymbol{a}_1 + v_a S_a^2) \tag{5}$$

$$\sigma_{e_1}^2 | else \sim Inv - Chi(n + v_e, \boldsymbol{e}_1' \boldsymbol{e}_1 + v_e S_e^2) \tag{6}$$

Computing sub-models for 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. In the latter case, 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.

Finally, the variance-covariance components for DMI are computed from observing the following relationships: $\mathbf{G}_0^* = \boldsymbol{\Lambda}^{-1}\mathbf{G}_0\boldsymbol{\Lambda}'^{-1}$, $\mathbf{P}_0^* = \boldsymbol{\Lambda}^{-1}\mathbf{P}_0\boldsymbol{\Lambda}'^{-1}$, and $\mathbf{R}_0^* = \boldsymbol{\Lambda}^{-1}\mathbf{R}_0\boldsymbol{\Lambda}'^{-1}$, where $\boldsymbol{G}_0 = \begin{pmatrix} \sigma_{a_r}^2 & \mathbf{0}' \\ \mathbf{0} & \mathbf{G}_{-r} \end{pmatrix}$, $\mathbf{P}_0 = \begin{pmatrix} \sigma_{p_r}^2 & \mathbf{0}' \\ \mathbf{0} & \mathbf{P}_{-r} \end{pmatrix}$, $\mathbf{R}_0 = \begin{pmatrix} \sigma_{e_r}^2 & \mathbf{0}' \\ \mathbf{0} & \mathbf{R}_{-r} \end{pmatrix}$. Note that the variances and covariances between energy sinks remain unchanged after transformations. That is,

$$\boldsymbol{G}_0^* = \boldsymbol{\Lambda}^{-1}\boldsymbol{G}_0\boldsymbol{\Lambda}'^{-1} = \begin{pmatrix} \sigma_{a_1}^2 + \Delta_a & \lambda_{12}\sigma_{a_2}^2 + \sum_{t\neq 1,2}^{k}\lambda_{1t}\sigma_{a_2a_t} & \cdots & \lambda_{1k}\sigma_{a_k}^2 + \sum_{t\neq 1,k}^{k}\lambda_{1t}\sigma_{a_ka_t} \\ \lambda_{12}\sigma_{a_2}^2 + \sum_{t\neq 1,2}^{k}\lambda_{1t}\sigma_{a_2a_t} & \sigma_{a_2}^2 & \cdots & \sigma_{a_2a_k} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{1k}\sigma_{a_k}^2 + \sum_{t\neq 1,k}^{k}\lambda_{1t}\sigma_{a_ka_t} & \sigma_{a_ka_2} & \cdots & \sigma_{a_k}^2 \end{pmatrix} \tag{7}$$

$$\boldsymbol{R}_0^* = \boldsymbol{\Lambda}^{-1}\boldsymbol{R}_0\boldsymbol{\Lambda}'^{-1} = \begin{pmatrix} \sigma_{e_1}^2 + \Delta_e & \lambda_{12}\sigma_{e_2}^2 + \sum_{t\neq 1,2}^{k}\lambda_{1t}\sigma_{e_2e_t} & \cdots & \lambda_{1k}\sigma_{e_k}^2 + \sum_{t\neq 1,k}^{k}\lambda_{1t}\sigma_{e_ke_t} \\ \lambda_{12}\sigma_{e_2}^2 + \sum_{t\neq 1,2}^{k}\lambda_{1t}\sigma_{e_2e_t} & \sigma_{e_2}^2 & \cdots & \sigma_{e_2e_k} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{1k}\sigma_{e_k}^2 + \sum_{t\neq 1,k}^{k}\lambda_{1t}\sigma_{e_ke_t} & \sigma_{e_ke_2} & \cdots & \sigma_{e_k}^2 \end{pmatrix} \tag{8}$$

where $\Delta_x = \sum_{t'=2}^{k}\lambda_{1t'}^2\sigma_{x_{t'}}^2 + \sum_{t'=2}^{k}\left(\lambda_{1t'}\sum_{t\neq 1,t'}^{k}\lambda_{1t}\sigma_{x_tx_{t'}}\right)$, for $x = a$ and e, respectively.

Appendix 2 Heritability and genetic correlation

The heritability for RFI and DMI, respectively, are defined as:

$$h^2_{RFI} = \frac{\sigma^2_{a_1}}{\sigma^2_{a_1}+\sigma^2_{e_1}} \tag{9}$$

$$h^2_{DMI} = \frac{\sigma^2_{a_1}+\Delta_a}{(\sigma^2_{a_1}+\Delta_a)+(\sigma^2_{e_1}+\Delta_{e_r})} \tag{10}$$

The heritability for an energy sink trait takes a similar formula as (9). The genetic correlation between RFI and an energy sink is fixed at zero. The genetic correlation between DMI and an energy sink trait is not zero, which is computed as follows:

$$r_{a_{1t'}} = \frac{\lambda_{1t'}\sigma^2_{a_{t'}}+\sum^k_{t\neq 1,t'}\lambda_{1t}\sigma_{a_t a_{t'}}}{\sqrt{(\sigma^2_{a_1}+\Delta_a)\times\sigma^2_{a_{t'}}}}, \text{ for } t' = 2, \dots, k \tag{11}$$

Appendix 3 Partial regression coefficients, given fixed and random effects

Consider equation (3) in Wu et al. (2021) and replace the structural coefficients, λ_{1j}, by partial regression coefficients, b_j, for $j = 2, \dots, k$. If we move all the fixed and random effects to the left-hand side of the equation and keep the energy sinks and the residual on the right-hand side, it becomes:

$$y_{i1} - \mu_1 - \mathbf{x}'_{i1}\boldsymbol{\beta}_1 - \boldsymbol{z}'_{i1}\boldsymbol{a}_1 = (y_{i2} \quad y_{i3} \quad \cdots \quad y_{ik})\begin{pmatrix} b_2 \\ b_3 \\ \vdots \\ b_k \end{pmatrix} + \boldsymbol{e}_1 \tag{12}$$

Then, the least-square (LS) solutions of the partial regression coefficients are the following:

$$\left(\begin{matrix} \hat{b}_2 \\ \hat{b}_3 \\ \vdots \\ \hat{b}_k \end{matrix}\middle| \mu_1, \boldsymbol{\beta}_1, \boldsymbol{a}_1\right) = \begin{pmatrix} \sum^n_{i=1} y^2_{i2} & \sum^n_{i=1} y_{i2}y_{i3} & \cdots & \sum^n_{i=1} y_{i2}y_{ik} \\ \sum^n_{i=1} y_{i3}y_{i2} & \sum^n_{i=1} y^2_{i3} & \cdots & \sum^n_{i=1} y_{i3}y_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum^n_{i=1} y_{ik}y_{i2} & \sum^n_{i=1} y_{ik}y_{i3} & \cdots & \sum^n_{i=1} y^2_{ik} \end{pmatrix}^{-1} \begin{pmatrix} \sum^n_{i=1} y_{i2}\, w_{i1} \\ \sum^n_{i=1} y_{i3}\, w_{i1} \\ \ddots \\ \sum^n_{i=1} y_{ik} w_{i1} \end{pmatrix} \tag{13}$$

where $w_{i1} = y_{i1} - \mu_1 - \mathbf{x}'_{i1}\boldsymbol{\beta}_1 - \boldsymbol{z}'_{i1}\boldsymbol{a}_1$. Note that (3) coincides precisely with (13) if we let $\tau^2 \to \infty$ in (3), which is equivalent to assigning flat priors to structural coefficients in (13). In other

words, the conditional posterior means of structural coefficients agree with (or asymptotically equivalent to) the partial regression coefficients in one-step LR, given μ_1, $\boldsymbol{\beta}_1$ and $\boldsymbol{\alpha}_1$, if we ignore the prior values (or the impact of priors diminishes when the data dominates the posteriors).

Appendix 4 A fully-recursive model based on the modified Cholesky Decomposition (Lu et al., 2015)

Consider the phenotypic relationships, for example, in the present example. The $L\Sigma L'$ decomposition implies fully recursive relationships for the traits (ordered by y_{i4}, y_{i3}, y_{i2}, and y_{i1}). Here, L is the unit lower triangular matrix, which corresponds to the structural coefficient matrix in RSEM, as follows:

$$\Lambda' = L' = \begin{pmatrix} 1 & -b_{12} & -b_{13} & -b_{14} \\ 0 & 1 & -b_{23} & -b_{24} \\ 0 & 0 & 1 & -b_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{14}$$

where $b_{jj'}$ is the effect (i.e., partial regression coefficient) from trait j' to trait j, and $y_{ij} = \sum_{j'=1}^{j-1} b_{jj'} y_{ij'}$, for $j = 2, \ldots, 4$. The covariance matrix between the reparameterized variables (y_{i4}, $y_{i3} - b_{34}y_{i4}$, $y_{i2} - \sum_{j=3}^{4} b_{1j}y_{ij}$, and $y_{i1} - \sum_{j=2}^{4} b_{1j}y_{ij}$) is diagonal, meaning that they are mutually independent. Following the same Bayesian modeling settings to implementing the reparameterized MT model as a Bayesian recursive model, we derive the conditional posterior means of the structural coefficients as follows:

$$E\left(\begin{pmatrix} b_{34} \\ b_{23} \\ b_{24} \\ b_{12} \\ b_{13} \\ b_{14} \end{pmatrix} \middle| else\right)$$

$$= \begin{pmatrix} \frac{\sigma_{e_1}^2}{\sigma_{e_3}^2}\sum y_{i4}^2 + \frac{\sigma_{e_1}^2}{\tau^2} & 0 & 0 & & & \\ 0 & \frac{\sigma_{e_1}^2}{\sigma_{e_2}^2}\sum y_{i3}^2 + \frac{\sigma_{e_1}^2}{\tau^2} & \frac{\sigma_{e_1}^2}{\sigma_{e_2}^2}\sum y_{i3}y_{i4} & & 0 & \\ 0 & \frac{\sigma_{e_1}^2}{\sigma_{e_2}^2}\sum y_{i3}y_{i4} & \frac{\sigma_{e_1}^2}{\sigma_{e_2}^2}\sum y_{i4}^2 + \frac{\sigma_{e_1}^2}{\tau^2} & & & \\ & & & \sum y_{i2}^2 + \frac{\sigma_{e_1}^2}{\tau^2} & \sum y_{i2}y_{i3} & \sum y_{i2}y_{i4} \\ & 0 & & \sum y_{i2}y_{i3} & \sum y_{i3}^2 + \frac{\sigma_{e_1}^2}{\tau^2} & \sum y_{i3}y_{i4} \\ & & & \sum y_{i2}y_{i4} & \sum y_{i3}y_{i4} & \sum y_{i4}^2 + \frac{\sigma_{e_1}^2}{\tau^2} \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \frac{\sigma_{e_1}^2}{\sigma_{e_3}^2} \sum y_{i4} w_{i3} + \frac{\sigma_{e_1}^2 \lambda_0}{\tau^{-2}} \\ \frac{\sigma_{e_1}^2}{\sigma_{e_2}^2} \sum y_{i3} w_{i2} + \frac{\sigma_{e_1}^2 \lambda_0}{\tau^{-2}} \\ \frac{\sigma_{e_1}^2}{\sigma_{e_2}^2} \sum y_{i4} w_{i2} + \frac{\sigma_{e_1}^2 \lambda_0}{\tau^{-2}} \\ \sum y_{i2} w_{i1} + \frac{\sigma_{e_1}^2 \lambda_0}{\tau^{-2}} \\ \sum y_{i3} w_{i1} + \frac{\sigma_{e_1}^2 \lambda_0}{\tau^{-2}} \\ \sum y_{i4} w_{i1} + \frac{\sigma_{e_1}^2 \lambda_0}{\tau^{-2}} \end{pmatrix} \tag{15}$$

In the above, the equations for b_{12}, b_{13}, b_{14} in (15) are identical to those for λ_{12}, λ_{13}, λ_{14} in (3), suggesting that both models are equivalent for evaluating RFI, assuming phenotypic recursive effects. However, they differed in the assumed relationships between energy sinks. For example, the model by Lu et al. (2015) assumed recursive effects from milk energy (MILKE) on MBW, but the feedback effect from MBW to MILKE did not exist. In contrast, the relationships between energy sinks are correlational in a recursive model.

Appendix 5 Partial regression coefficients based on a multiple-trait model for RFI and energy sinks

Based on the standardized phenotypes, the phenotypic partial regressions are derived as follows, which is also in a comparable form with (3):

$$\boldsymbol{b} = \boldsymbol{c}_{12} \boldsymbol{V}_{22}^{-1}$$

$$= \begin{pmatrix} \sum_{i=1}^{n} y_{i1} y_{i2} & \sum_{i=1}^{n} y_{i1} y_{i3} & \cdots & \sum_{i=1}^{n} y_{i1} y_{ik} \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{n} y_{i2}^2 & \sum_{i=1}^{n} y_{i2} y_{i3} & \cdots & \sum_{i=1}^{n} y_{i2} y_{ik} \\ \sum_{i=1}^{n} y_{i3} y_{i2} & \sum_{i=1}^{n} y_{i3}^2 & \cdots & \sum_{i=1}^{n} y_{i3} y_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} y_{ik} y_{i2} & \sum_{i=1}^{n} y_{ik} y_{i3} & \cdots & \sum_{i=1}^{n} y_{ik}^2 \end{pmatrix}^{-1} \tag{16}$$

where $\boldsymbol{c}_{12}$ is a vector of the phenotypic covariances between DMI and the energy sink traits, and $\boldsymbol{V}_{22}$ is the phenotypic variance-covariance matrix for the energy sink traits.

References

Gianola, D., and D. Sorensen. 2004. Quantitative genetic models for describing simultaneous and recursive relationships between phenotypes. Genetics. 167:1407-1424.

Wu, X-L., B. Heringstad, Y. M. Chang, G. de Los Campos, and D. Gianola. 2007. Inferring relationships between somatic cell score and milk yield using simultaneous and recursive models. J. Dairy Sci. 90:3508–3521.

Wu, X-L., B. Heringstad, and D. Gianola. 2008. Exploration of lagged relationships between mastitis and milk yield in dairy cows using a Bayesian structural equation Gaussian-threshold model. Genet. Sel. Evol. 40: 333–357.

Wu, X-L., B. Heringstad, and D. Gianola. 2010. Bayesian structural equation models for inferring relationships between phenotypes: a review of methodology, identifiability, and applications. J. Anim. Breed. Genet. 127:3–15.