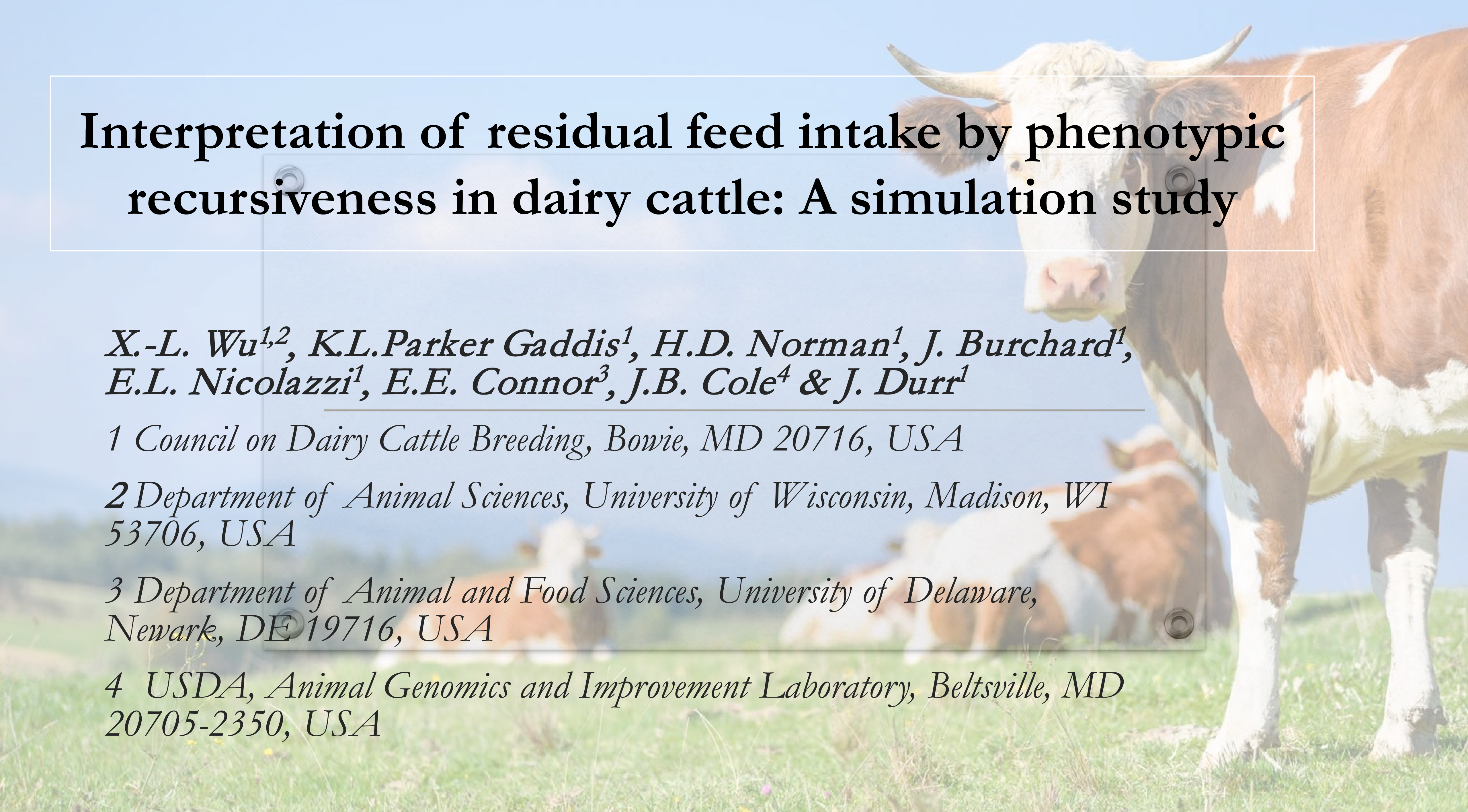


Interpretation of residual feed intake by phenotypic recursiveness in dairy cattle: A simulation study

X.-L. Wu[1,2], K.L.Parker Gaddis[1], H.D. Norman[1], J. Burchard[1], E.L. Nicolazzi[1], E.E. Connor[3], J.B. Cole[4] & J. Durr[1]

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

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

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

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

Residual feed intake (RFI)

- Feed efficiency is an important trait for dairy cattle breeding and management because feed costs comprise a large portion of various costs associated with dairy production.
- Feed efficiency is defined either as a ratio trait or as a regression (residual) trait (Berry and Crowley, 2013). In the past decades, residual feed intake (RFI) has become increasingly popular as a measure of net feed efficiency.
- RFI was initially proposed by Koch et al. (1963) as the residuals from the regression of feed intake on various energy sinks.
- In essence, RFI represents a resource allocation theory, from which animal deviants can be identified, these being animals that require more or less feed than predicted (Herd, 2009).

Statistical models for evaluating RFI

- Single-trait analysis
  - S1: Two-stage linear regression
  - S2: One-step linear regression
- Multiple-trait analysis
  - M1: Multiple-trait, mixed-effects models (MT)
  - M2: Recursive structural coefficient models (RSEM)

Single-trait, two-stage linear regression

- ❑ Stage #1
  - ○ A linear regression fits dry matter intake (DMI) as a linear function of energy sink traits, e.g., metabolic body weight (MBW), energy-corrected milk (ECM) or milk net energy (MILKNE), and changes in body weight. The residual is taken to be the RFI phenotype.

$$\boldsymbol{y}_1 = \sum_{j=2}^{k} b_{1k}\boldsymbol{y}_k + \boldsymbol{r}_1$$

- ❑ Stage #2
  - ○ The RFI phenotype is fitted by a mixed-effects model to estimate the genetic values and relevant genetic parameters.

$$\boldsymbol{r}_1 = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{Z}_1\boldsymbol{a}_1 + \boldsymbol{e}_1$$

where: $\boldsymbol{a}_1 \sim N(0, \boldsymbol{A}\sigma_a^2)$

Single-trait, one-step linear regression

- Combining these two modeling stages leads to a one-step approach without the need to estimate the residuals as the RFI phenotypes (e.g., Templeman et al., 2015).

$$\boldsymbol{y}_1 = \sum_{j=2}^{k} b_{1k}\boldsymbol{y}_k + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{Z}_1\boldsymbol{a}_1 + \boldsymbol{e}_1$$

- Fitting phenotypes as regressor variables in a linear regression was 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, mixed-effects model (Two-stage)

- ❑ Fit a multiple-traits, mixed-effects model, and estimate variance-covariance matrices (Note that accurately estimating (co)variance components often require a considerably large dataset)

$$\mathbf{G} = \begin{pmatrix} \sigma^2_{a_1} & \boldsymbol{G}_{1,2:k} \\ \boldsymbol{G}_{1,2:k} & \boldsymbol{G}_{2:k} \end{pmatrix};\ \mathbf{R} = \begin{pmatrix} \sigma^2_{e_1} & \boldsymbol{R}_{1,2:k} \\ \boldsymbol{R}_{1,2:k} & \boldsymbol{R}_{2:k} \end{pmatrix}; \boldsymbol{V}_p = \boldsymbol{V}_G + \boldsymbol{V}_R = \begin{pmatrix} \sigma^2_{p_1} & \boldsymbol{P}_{1,2:k} \\ \boldsymbol{P}_{1,2:k} & \boldsymbol{P}_{2:k} \end{pmatrix}$$

- ❑ Phenotypes or/and genetic values of RFI are obtained by a follow-up partial linear regression

Partial regression coefficients for RFI phenotypes: $\boldsymbol{b}_p = \boldsymbol{P}_{1,2:k}\, \boldsymbol{P}^{-1}_{2:k}$;

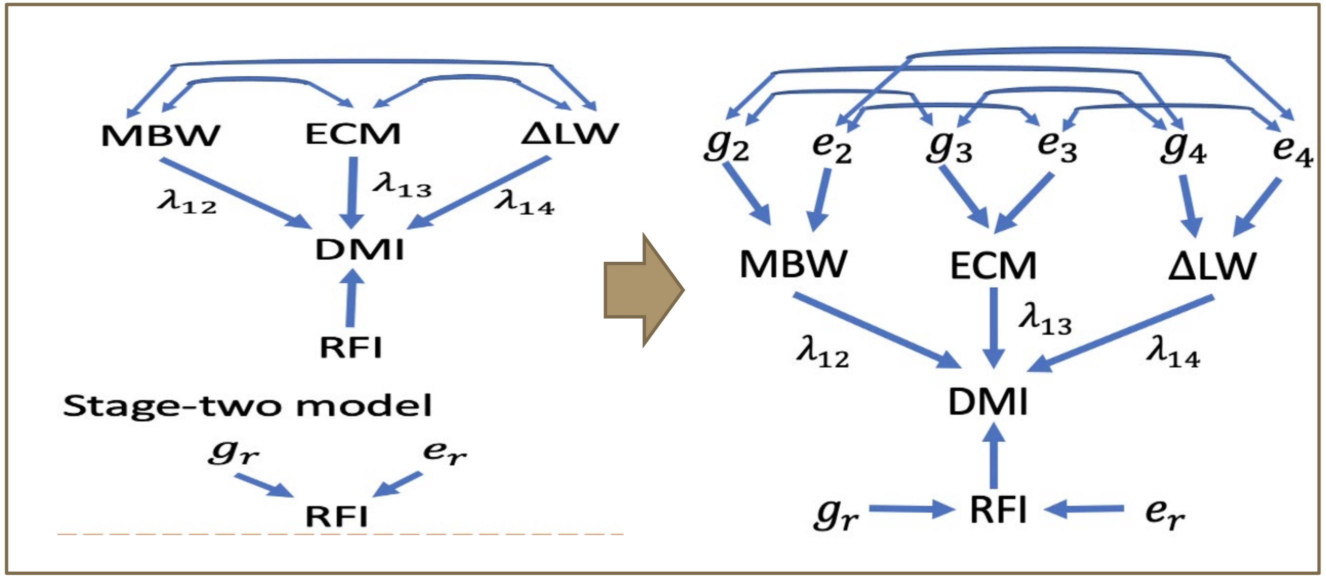
Partial regression coefficients for RFI genetic values: $\boldsymbol{b}_G = \boldsymbol{G}_{1,2:k}\, \boldsymbol{G}^{-1}_{2:k}$

Recursive structural equation model

- For the i-th individual

$$\boldsymbol{\Lambda}\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{a} + \mathbf{e}_i$$

$$\boldsymbol{\Lambda} = \begin{pmatrix} 1 & -\lambda_{12} & \cdots & -\lambda_{1k} \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$



RSEM: Variance-covariance matrices

RFI **DMI**

$$\boldsymbol{G}_0^{*}= \boldsymbol{\Lambda}^{-1}\boldsymbol{G}_0\boldsymbol{\Lambda}'^{-1}$$

$$\boldsymbol{G}_0 = \begin{pmatrix} \sigma_{a_r}^2 & 0 & \dots & 0 \\ 0 & \sigma_{a_2}^2 & \dots & \sigma_{a_{2k}} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \sigma_{a_{2k}} & \dots & \sigma_{a_4}^2 \end{pmatrix}; \quad \boldsymbol{G}_0^{*} = \begin{pmatrix} \sigma_{a_r}^2 + \Delta_a & \lambda_{12}\sigma_{a_2}^2 + \sum_{t\neq r,2}^{k}\lambda_{1t}\sigma_{a_{2t}} & \dots & \lambda_{1k}\sigma_{a_k}^2 + \sum_{t\neq r,k}^{k}\lambda_{1t}\sigma_{a_{kt}} \\ \lambda_{12}\sigma_{a_2}^2 + \sum_{t\neq r,2}^{k}\lambda_{1t}\sigma_{a_{2t}} & \sigma_{a_2}^2 & \dots & \sigma_{a_{2k}} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{1k}\sigma_{a_k}^2 + \sum_{t\neq r,k}^{k}\lambda_{1t}\sigma_{a_{kt}} & \sigma_{a_{2t}} & \dots & \sigma_{a_k}^2 \end{pmatrix}$$

$$R_0^{*}= \boldsymbol{\Lambda}^{-1}R_0\boldsymbol{\Lambda}'^{-1}$$

$$\boldsymbol{R}_0 = \begin{pmatrix} \sigma_{e_r}^2 & 0 & \dots & 0 \\ 0 & \sigma_{e_2}^2 & \dots & \sigma_{e_{2k}} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \sigma_{e_{2k}} & .. & \sigma_{e_k}^2 \end{pmatrix} \quad R_0^{*} = \begin{pmatrix} \sigma_{e_r}^2 + \Delta_e & \lambda_{12}\sigma_{e_2}^2 + \sum_{t\neq r,2}^{k}\lambda_{1t}\sigma_{e_{2t}} & \dots & \lambda_{1k}\sigma_{e_k}^2 + \sum_{t\neq r,k}^{k}\lambda_{1t}\sigma_{e_{kt}} \\ \lambda_{12}\sigma_{e_2}^2 + \sum_{t\neq r,2}^{k}\lambda_{1t}\sigma_{e_{2t}} & \sigma_{e_2}^2 & \dots & \sigma_{e_{2k}} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{1k}\sigma_{e_k}^2 + \sum_{t\neq r,k}^{k}\lambda_{1t}\sigma_{e_{kt}} & \sigma_{e_{2t}} & \dots & \sigma_{e_k}^2 \end{pmatrix}$$

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

Conditional posterior distribution: structural coefficients

$$\boldsymbol{\lambda}|else \sim MVN(\boldsymbol{\mu}_{\lambda}, \boldsymbol{V}_{\lambda})$$

where: $$\boldsymbol{V}_{\lambda} = \sigma_{e_1}^2 \times \begin{pmatrix} \sum_{i=1}^{n} y_{i2}^2 + \sigma_{e_1}^2 \tau^{-2} & \sum_{i=1}^{n} y_{i2} y_{i3} & \cdots & \sum_{i=1}^{n} y_{i2} y_{ik} \\ \sum_{i=1}^{n} y_{i3} y_{i2} & \sum_{i=1}^{n} y_{i3}^2 + \sigma_{e_1}^2 \tau^{-2} & \cdots & \sum_{i=1}^{n} y_{i3} y_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} y_{ik} y_{i2} & \sum_{i=1}^{n} y_{ik} y_{i3} & \cdots & \sum_{i=1}^{n} y_{ik}^2 + \sigma_{e_1}^2 \tau^{-2} \end{pmatrix}^{-1}$$

$$\boldsymbol{\mu}_{\lambda} = \sigma_{e_1}^{-2} \times \boldsymbol{V}_{\lambda} \times \begin{pmatrix} \sum_{i=1}^{n} y_{i2} w_{i1} + \sigma_{e_1}^2 \tau^{-2} \lambda_0 \\ \sum_{i=1}^{n} y_{i3} w_{i1} + \sigma_{e_1}^2 \tau^{-2} \lambda_0 \\ \ddots \\ \sum_{i=1}^{n} y_{ik} w_{i1} + \sigma_{e_1}^2 \tau^{-2} \lambda_0 \end{pmatrix}$$

The simplified algorithm

- A standard multiple-trait mixed-effects model analysis of energy sink traits, independent of the computed RFI phenotypes. This step can be implemented by Markov chain Monte Carlo, or simply by REML.
- Markov chain Monte Carlo sampling for the structural coefficients and model parameters for RFI.
- The variance-covariance components between DMI and energy sinks are based on the following relationships:

$$\mathbf{G}_0^* = \boldsymbol{\Lambda}^{-1}\boldsymbol{G}_0\boldsymbol{\Lambda}'^{-1};\ \mathbf{R}_0^* = \boldsymbol{\Lambda}^{-1}\mathbf{R}_0\boldsymbol{\Lambda}'^{-1}$$

where $\boldsymbol{G}_0 = \begin{pmatrix} \sigma_{a_r}^2 & \mathbf{0}' \\ \mathbf{0} & \mathbf{G}_{-r} \end{pmatrix}$, $\mathbf{R}_0 = \begin{pmatrix} \sigma_{e_r}^2 & \mathbf{0}' \\ \mathbf{0} & \mathbf{R}_{-r} \end{pmatrix}$.

Simulation parameters

$$\begin{array}{c} DMI \\ MBW \\ ECM \\ \Delta BW \end{array} \quad \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \boldsymbol{y}_3 \\ \boldsymbol{y}_4 \end{pmatrix} = \boldsymbol{\mu} + \begin{pmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \boldsymbol{a}_3 \\ \boldsymbol{a}_4 \end{pmatrix} + \begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \boldsymbol{e}_3 \\ \boldsymbol{e}_4 \end{pmatrix}$$

$$\boldsymbol{\mu} = \boldsymbol{0}; \quad \begin{pmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \boldsymbol{a}_3 \\ \boldsymbol{a}_4 \end{pmatrix} \sim \mathrm{MVN}(\boldsymbol{0}, \mathbf{G} \otimes \boldsymbol{A}); \begin{pmatrix} e_1 \\ \boldsymbol{e}_2 \\ \boldsymbol{e}_3 \\ \boldsymbol{e}_4 \end{pmatrix} \sim \mathrm{MVN}(\boldsymbol{0}, \mathbf{R} \otimes \mathbf{I})$$

$$\mathbf{G} = \begin{pmatrix} 0.399 & 0.205 & 0.191 & 0.012 \\ 0.205 & 0.524 & 0.033 & 0.037 \\ 0.191 & 0.033 & 0.287 & -0.005 \\ 0.012 & 0.037 & -0.005 & 0.048 \end{pmatrix}; \quad \mathbf{R} = \begin{pmatrix} 0.584 & 0.236 & 0.311 & 0.179 \\ 0.236 & 0.534 & 0.142 & 0.144 \\ 0.311 & 0.142 & 0.674 & -0.022 \\ 0.179 & 0.144 & -0.022 & 0.96 \end{pmatrix}$$

Structural coefficients/patrial regression

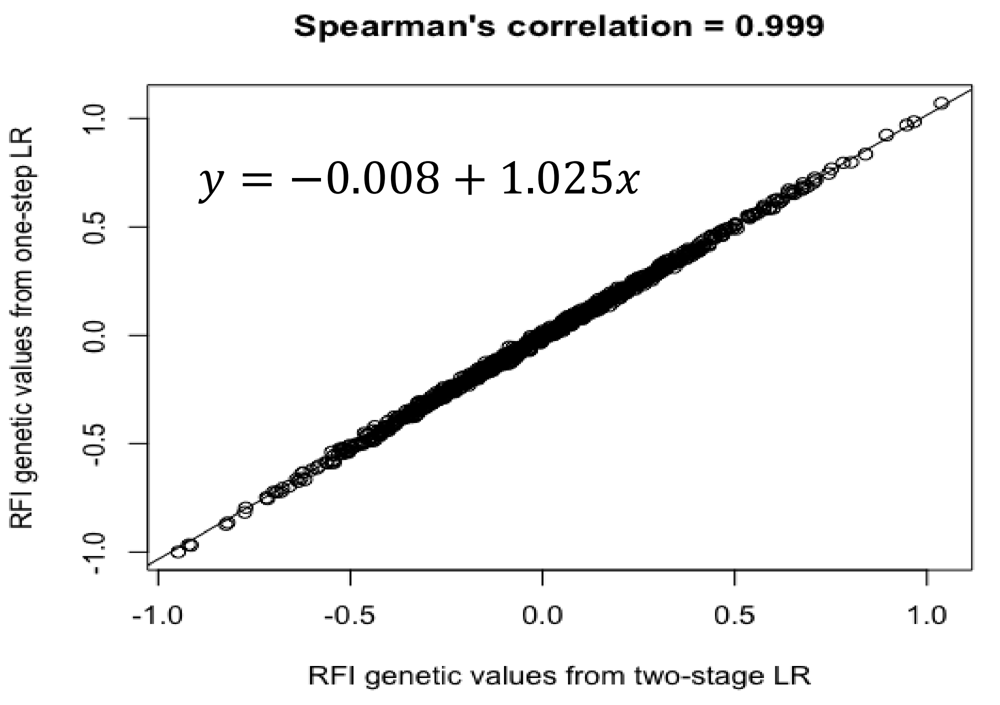
Energy sink	Two-stage LR		One-step LR		RSEM		MT		
	Mean	SD	Mean	SD	Posterior mean	Posterior SD	COVP	COVG	COVE
MBW	**0.327**	0.026	**0.313**	0.027	**0.312**	0.027	**0.327**	0.349	0.300
ECM	**0.498**	0.027	**0.470**	0.027	**0.469**	0.027	**0.498**	0.629	0.404
dLW	**0.127**	0.026	**0.137**	0.025	**0.137**	0.025	**0.127**	0.049	0.151

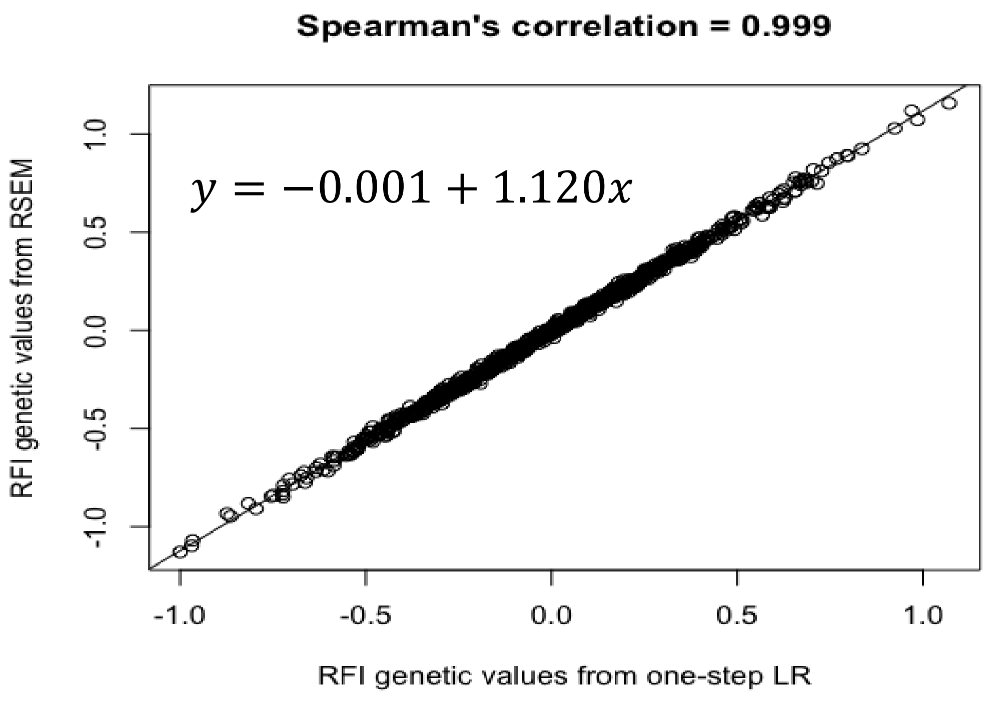
LR = single-trait linear regression; RSEM = Structural equation model; MT = multiple-traits, mixed effects model

COVP = partial regression coefficients from the MT model based on phenotypic variance-covariances;
COVG = partial regression coefficients from the MT model based on genetic variance-covariances.

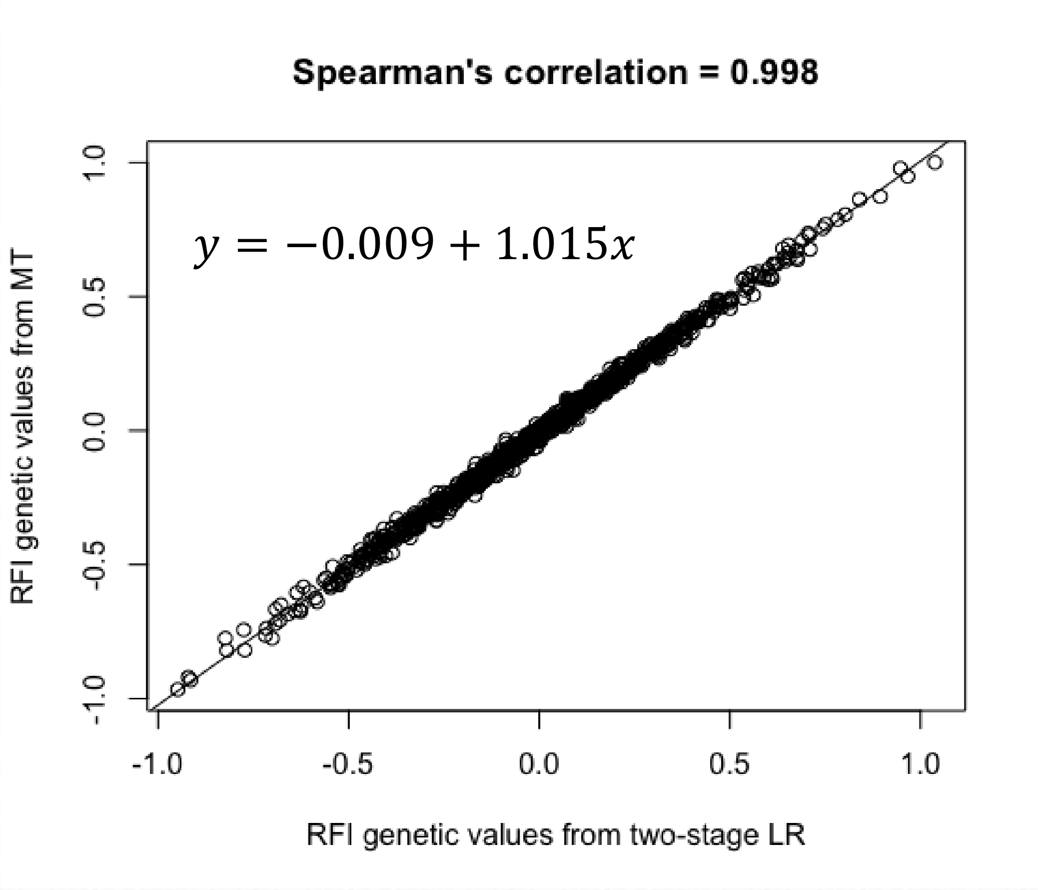
Two-stage LR, one-step LR, and RSEM

- The recursive model was equivalent to single-trait linear regression concerning the estimated RFI genetic values, but the recursive model has expanded the analytical capability to multiple traits with phenotypic relationships assumed.

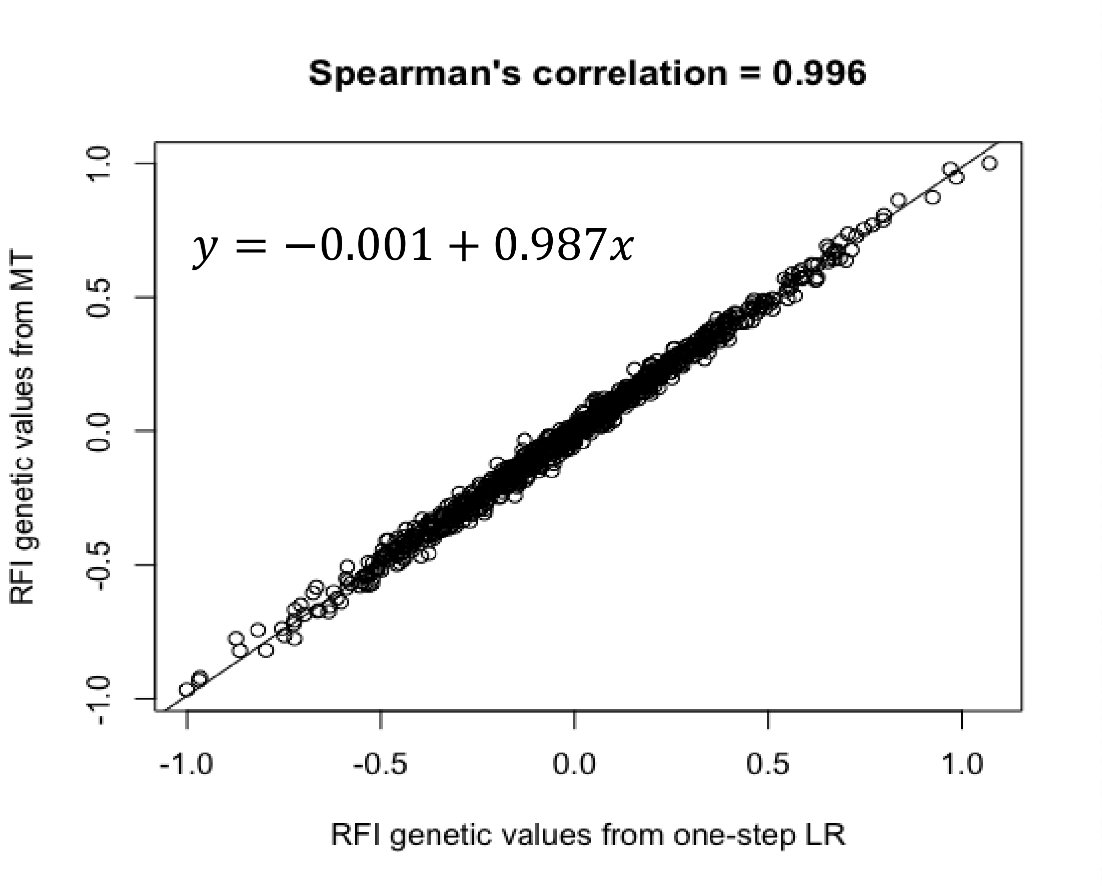




Two-stage LR, one-step LR, and MT



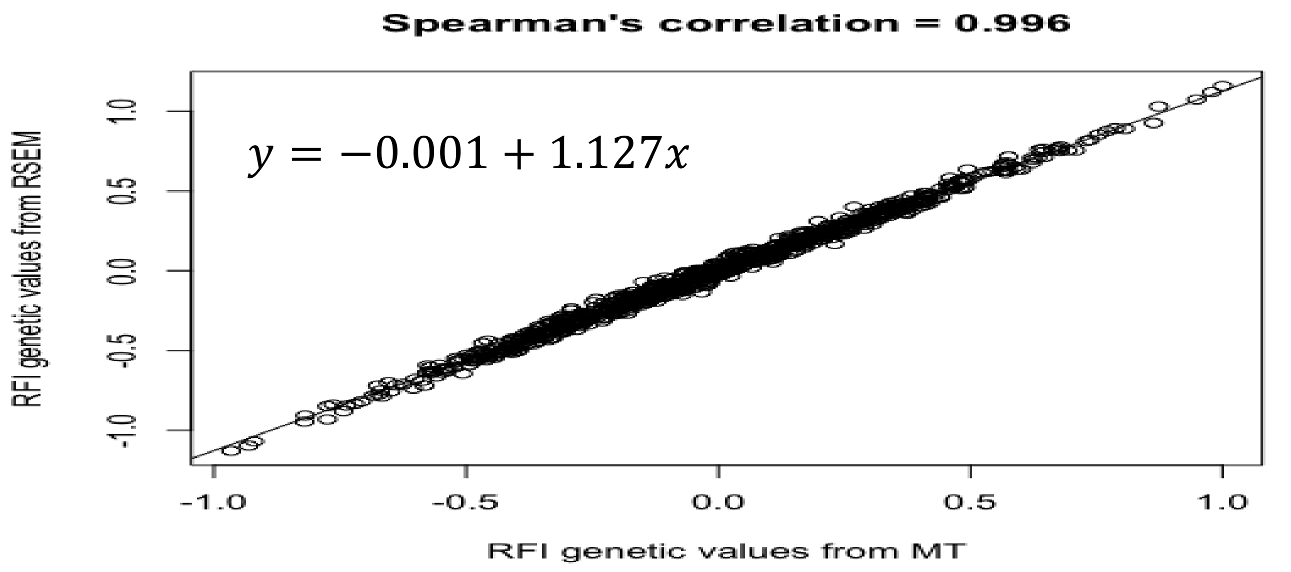
Spearman's correlation = 0.998



Spearman's correlation = 0.996

RSEM vs. MT model

- RSEM captures phenotypic recursive effects whereas the MT model derives the partial regression coefficients for RFI genetic values based on genetic variance-covariances only.
- These results were no indication of which model was more accurate because the true RFI genetic values were unknown, and because the simulation results were subject to the underlying assumptions, whether it favors one model or the other.



Conclusions

- We proposed a Bayesian structural equation model as a flexible, one-step, direct method for the genetic evaluation of RFI. A simplified algorithm is proposed, which facilitates dealing with large datasets in real applications.
- The recursive model was equivalent to single-trait linear regression concerning the estimated RFI genetic values, but the recursive model has expanded the analytical capability to multiple traits with phenotypic relationships assumed.
- The recursive model extends naturally to deal with heterogenous recursiveness that varied with subpopulations or varied genetic and residual relationships within the same population. Extending the recursive model to genomic prediction is straightforward too, which can be accomplished by replacing the additive genetic relationship matrix with a genomic relationship matrix.