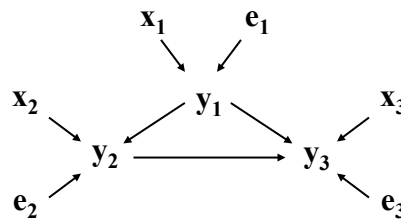


Structural Equation Models

Guilherme J. M. Rosa
University of Wisconsin-Madison

Structural Equation Models

⇒ Causal structure represented as a Directed Graph



Example of a causal structure, in which y 's represent measurements on three phenotypic traits, x 's and e 's represent known explanatory variables and residual factors affecting y 's, respectively.

Structural Equation Models

⇒ Graph represented by a set of structural equations

$$\begin{cases} y_1 = \beta_1 x_1 + e_1 \\ y_2 = \lambda_{21} y_1 + \beta_2 x_2 + e_2 \\ y_3 = \lambda_{31} y_1 + \lambda_{32} y_2 + \beta_3 x_3 + e_3 \end{cases}$$

• **Matrix notation:** $\mathbf{y} = \Lambda \mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$

observations → matrix of coefficients λ → residuals → fixed effects → exogenous covariates

$$\Lambda = \begin{bmatrix} 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ \lambda_{31} & \lambda_{32} & 0 \end{bmatrix}$$

Structural Equation Models

⇒ “Reduced Model”:

$$\mathbf{y} = \Lambda \mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad \mathbf{e} \sim N(\mathbf{0}, \mathbf{D})$$

$$\mathbf{y} - \Lambda \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

$$(\mathbf{I} - \Lambda) \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

$$\mathbf{y} = \Lambda^* \mathbf{X}\boldsymbol{\beta} + \Lambda^* \mathbf{e}, \quad \Lambda^* = (\mathbf{I} - \Lambda)^{-1}$$

$$\mathbf{y} = \boldsymbol{\mu}^* + \mathbf{e}^*, \quad \mathbf{e}^* = \Lambda^* \mathbf{e} \sim N(\mathbf{0}, \Lambda^* \mathbf{D} \Lambda^{*'})$$

Structural Equation Models

⇒ “Reduced Model”: Toy Example

$$\begin{cases} z = 2y \\ y = 3 \end{cases}$$


$$\begin{cases} z = 6 \\ y = 3 \end{cases}$$

Example of Application

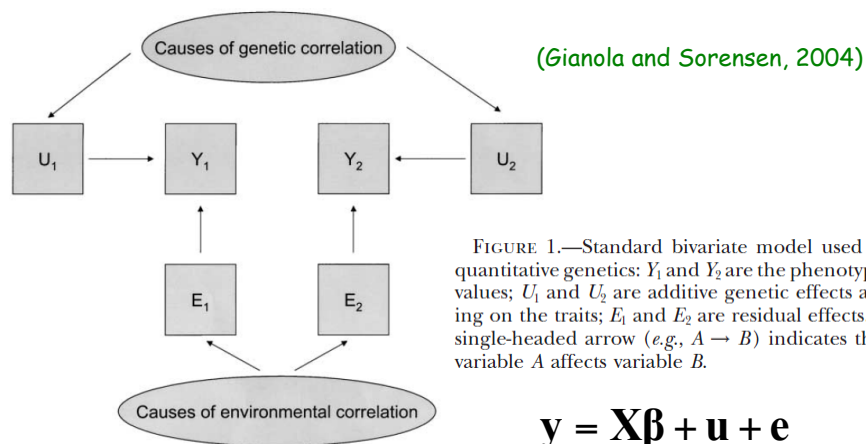
- ⇒ Completely randomized experiment to compare two treatments in terms of their effect on ribeye area in beef cattle
- ⇒ Data collected on ribeye area (RIB) and body weight (BW)
- ⇒ Results indicate a significant effect of treatment on RIB
- ⇒ Results indicate also a significant effect of treatment on BW
- ⇒ However, when BW is included as a covariate in the model for RIB, the treatment effect on RIB becomes non-significant



Structural Equation Models

- ⇒ Competing networks may be compared using some model selection criteria (e.g., LRT, AIC, BIC, or Bayesian approaches)
- ⇒ SEM intensively used in many fields, such as economics, psychometrics, social statistics, and biological sciences.
- ⇒ More recently, in quantitative genetics in the context of mixed model analysis (e.g., Gianola and Sorensen 2004)

SEM in Quantitative Genetics



$$r_{Y_1, Y_2} = h_1 h_2 r_{U_1, U_2} + e_1 e_2 r_{E_1, E_2}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} + \mathbf{e}$$

$$\mathbf{u} \sim N(\mathbf{0}, \mathbf{A}\sigma_a^2)$$

SEM in Quantitative Genetics

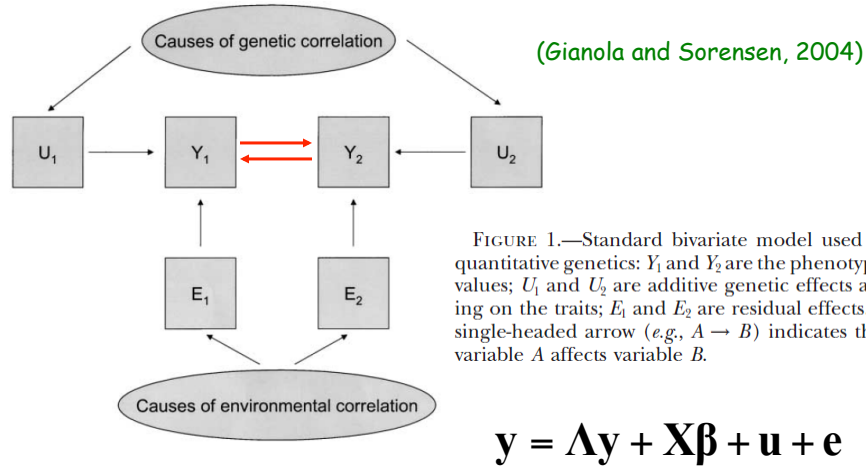


FIGURE 1.—Standard bivariate model used in quantitative genetics: Y_1 and Y_2 are the phenotypic values; U_1 and U_2 are additive genetic effects acting on the traits; E_1 and E_2 are residual effects. A single-headed arrow (e.g., $A \rightarrow B$) indicates that variable A affects variable B .

Example of Application

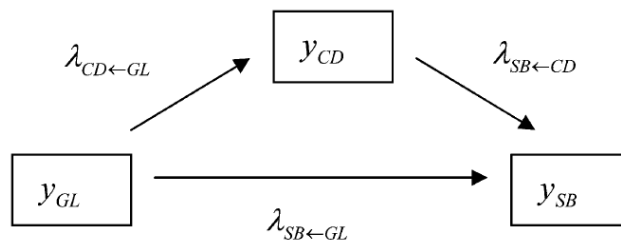


FIGURE 2.—Scheme of the recursive relationships between calving traits (y_{CD} , liability to calving difficulty; y_{SB} , liability to stillbirth; y_{GL} , phenotype for gestation length; the λ 's are structural coefficients).

Maturana EL, Wu X-L, Gianola D, Weigel KA and Rosa GJM. Exploring biological relationships between calving traits in primiparous cattle with a Bayesian recursive model. *Genetics* 181: 277-287, 2009.

Example of Application

$$\Lambda_k y_i = \mathbf{X}_i \mathbf{b} + \mathbf{Z}_{i(h)} \mathbf{h} + \mathbf{Z}_{i(s)} \mathbf{s} + \mathbf{Z}_{i(mgs)} \mathbf{mgs} + \mathbf{e}$$

$$\Lambda_k = \begin{bmatrix} 1 & 0 & 0 \\ -\lambda_{\text{CD} \leftarrow \text{GL}(k)} & 1 & 0 \\ -\lambda_{\text{SB} \leftarrow \text{GL}(k)} & -\lambda_{\text{SB} \leftarrow \text{CD}(k)} & 1 \end{bmatrix}$$

Estimates of recursive effects on the observable scale, by subgroup of gestation length

	261–267 days	268–273 days	274–279 days	280–291 days
$\lambda_{\text{CD} \leftarrow \text{GL}}$ (%/1 day GL)	0.06	0.24 ^a	0.37 ^a	0.47 ^a
$\lambda_{\text{SB} \leftarrow \text{GL}}$ (%/1 day GL)	-0.96 ^a	-0.57 ^a	-0.40 ^a	0.23 ^a
$\lambda_{\text{SB} \leftarrow \text{CD}}$ (%/1% CD)	0.61 ^a	0.60 ^a	0.60 ^a	0.60 ^a
$\lambda_{\text{SB} \leftarrow \text{CD} \leftarrow \text{GL}}$ ^b (%/1 day GL)	0.03	0.14 ^a	0.23 ^a	0.29 ^a
Overall effect of GL on SB (%/1 day GL)	-0.93 ^a	-0.43 ^a	-0.17	0.52 ^a

^a HPD_{99%} region does not include 0, where HPD is the highest posterior density.

^b $\lambda_{\text{SB} \leftarrow \text{CD} \leftarrow \text{GL}}$ corresponds to the effect of GL on SB mediated through CD.

Inferring Causal Phenotype Networks

⇒ *d-Separation* ('directed separation') concept: Verma and Pearl (1988), Pearl (1998), Geiger et al. (1990)

Two variables X and Y are said to be d-separated by Q if there is no active path between any X and Y conditionally on Q

• Example 1: y_2 is not a collider

• Example 2: y_2 is a collider

$\left\{ \begin{array}{l} y_3 \text{ is not independent of } y_1 \\ \text{Conditioned on } y_2, y_3 \text{ becomes} \\ \text{independent of } y_1 \end{array} \right.$

$\left\{ \begin{array}{l} y_1 \text{ and } y_3 \text{ unconditionally independent} \\ y_1 \text{ and } y_3 \text{ not independent,} \\ \text{conditioned on } y_2 \end{array} \right.$

$y_1 \longrightarrow y_2 \longrightarrow y_3$

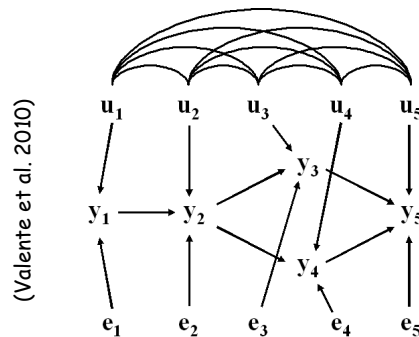
$y_1 \longrightarrow y_2 \longleftarrow y_3$

The IC Algorithm (Inductive Causation; Verma and Pearl 1991)

- Step 1:** Undirected graph (search for d-separations; connect adjacent variables)
- Step 2:** Partially oriented graph (search for colliders)
- Step 3:** Attempt to orient remaining undirected edges such that no new colliders or cycles are generated

Inferring Causal Phenotype Networks

- ⇒ Valente et al. (2010): Correlated genetic effects act as an additional source of phenotypic covariance, which may confound the search for causal structures



Example of network involving five phenotypic (observable) traits, and their corresponding additive genetic (u 's) and residual (e 's) effects; arcs connecting u 's represents genetic correlations

Inferring Causal Phenotype Networks

⇒ Valente et al. (2010): To restore the connection between causal structures and joint density, the joint distribution of phenotypes conditionally on additive genetic effects is used

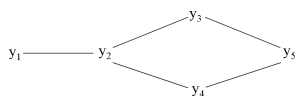
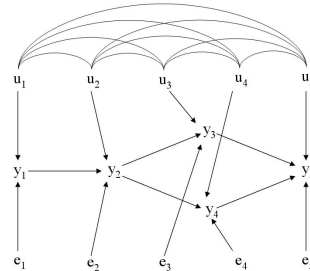
1. Fit a multiple trait model where additive genetic effects could be predicted based on pedigree information
2. Apply the IC algorithm to this matrix, returning a class of equivalent causal structures (i.e. causal structures that results in the same conditional independencies in the joint probability distribution)
3. Fit final structural equation model using the selected causal structure

Valente BD, Rosa GJM, de los Campos G, Gianola D and Silva MA. Searching for recursive causal structures in multivariate quantitative genetics mixed models. *Genetics* 185: 633-644, 2010.

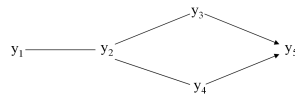
⇒ Valente et al. (2010): Simulation study

5 phenotypic traits
300 inbred lines
6 individuals per line

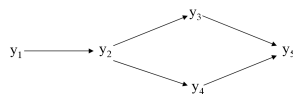
Model from which simulated data were drawn.



Undirected acyclic graph resulting from step 1 of the IC algorithm



Partially oriented graph retrieved by the IC algorithm



Structure chosen given prior beliefs on the direction of the edge between y_1 and y_2 (pointing towards y_2)

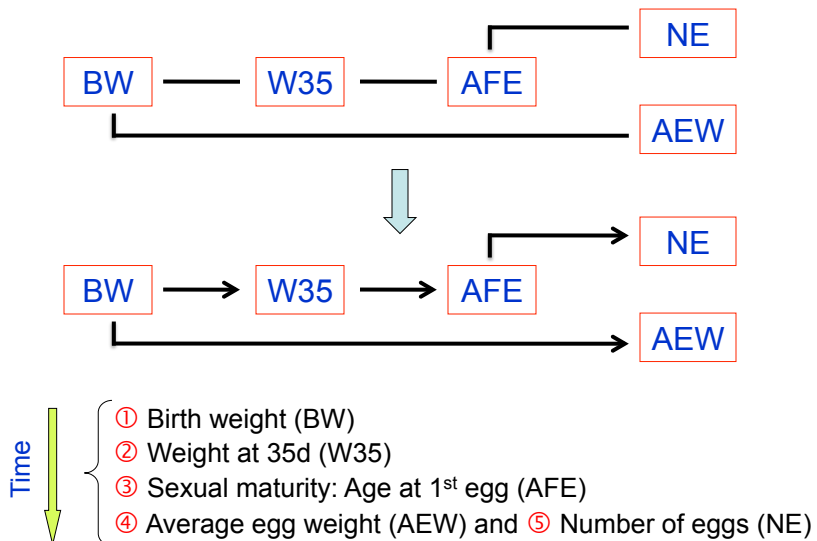
Example of Application

- Reproductive Traits in European Quail
- 892 females with observations
- Pedigree: 10,291 animals
- 5 phenotypic traits:
 - ① Birth weight
 - ② Weight at 35d
 - ③ Age at sexual maturity (1st egg)
 - ④ Average egg weight
 - ⑤ Rate of lay (# eggs)



Valente BD, Rosa GJM, Silva MA, Teixeira RB and Torres RA. Searching for phenotypic causal networks involving complex traits: an application to European quails. *Genet. Sel. Evol.* 43:37, 2011.

Example of Application





Example of Application 3: Milk Fatty Acid Composition in Dairy Cattle

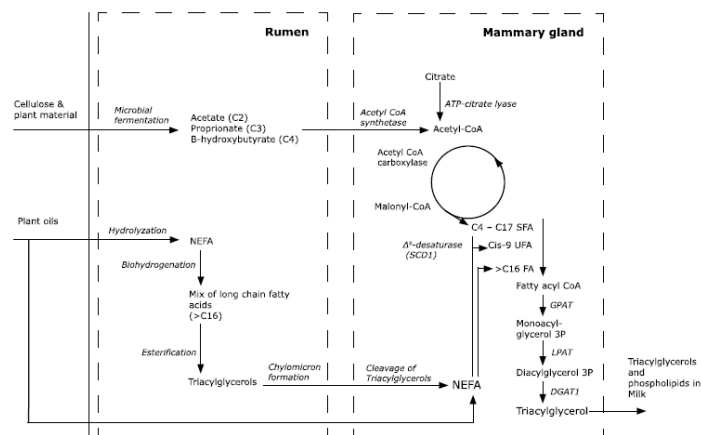


- 1,905 first lactation cows (63-282 DIM)
- 397 commercial farms
- Last four generations in pedigree: 4676 animals
- Fatty Acids (14 Total): C:4:0-C18:0, C10:1-C18:1, CLA (g/kg milk)

Bouwman AC, Valente BD, Janss LLG, Bovenhuis H and Rosa GJM. Exploring causal networks of bovine milk fatty acids in multivariate mixed model context. *Genetics Selection Evolution* 46:2, 2014.

Milk Fatty Acids Network

- Schematic overview of major milk fatty acids synthesis pathways



IC Algorithm, Results

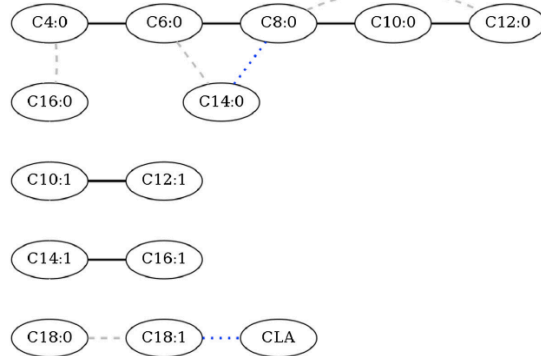


Figure 1 Network obtained from the inductive causation (IC) algorithm with different highest posterior density (HPD) intervals. The connections obtained with a HPD interval of 95% and 90% are given in black solid lines, with a HPD interval of 85% in grey dashed lines, and with a HPD interval of 80% in blue dotted lines.

- No unshielded colliders were recovered from the data and the resulting network was completely undirected

SEM, Results

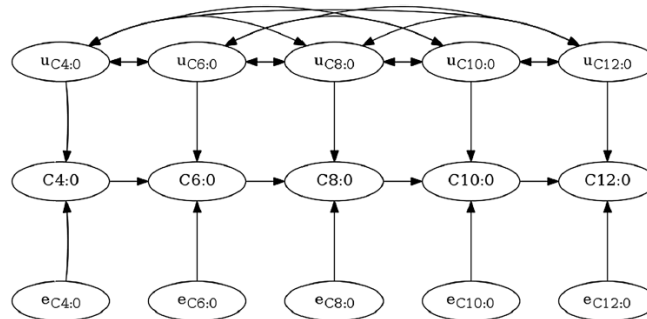
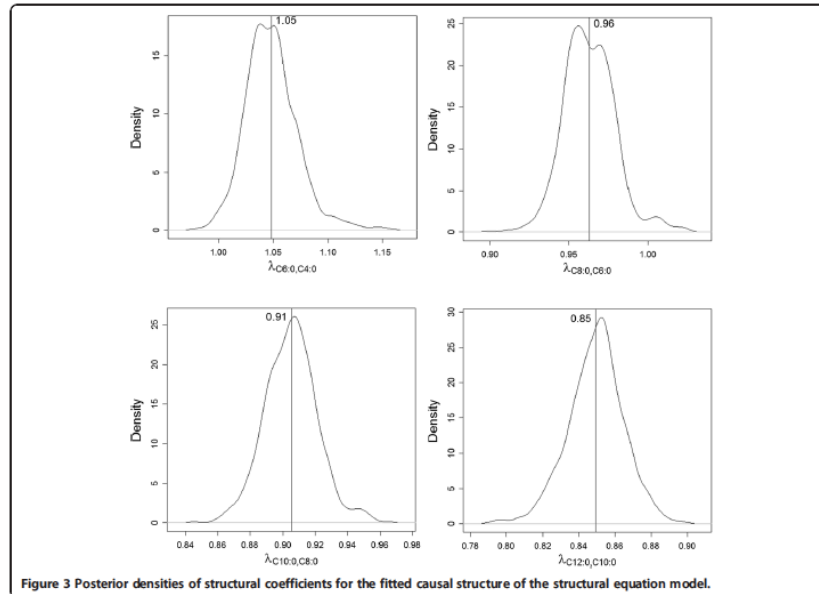


Figure 2 The fitted causal structure of the structural equation model. The edges in the fitted structure represent the causal relations for the observed variables (C4:0-C12:0), with independent residuals ($e_{C4:0}$ - $e_{C12:0}$) and correlated additive genetic effects ($u_{C4:0}$ - $u_{C12:0}$).

$$\Lambda = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \lambda_{C6:0,C4:0} & 0 & 0 & 0 & 0 \\ 0 & \lambda_{C8:0,C6:0} & 0 & 0 & 0 \\ 0 & 0 & \lambda_{C10:0,C8:0} & 0 & 0 \\ 0 & 0 & 0 & \lambda_{C12:0,C10:0} & 0 \end{bmatrix}$$

SEM, Results



Example of Application 4: Meat Quality Traits in Japanese Black Cattle

- 11,588 carcass records of Japanese Black cattle
- 57,523 animals in the pedigree
- Meat quality traits:

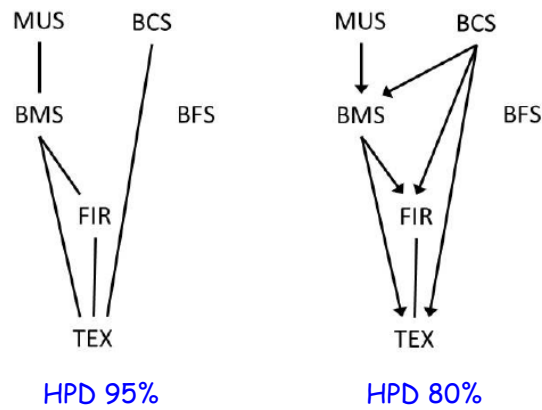
- ① Beef marbling score (BMS)
- ② Beef color score (BCS)
- ③ Firmness of beef (FIR)
- ④ Texture of beef (TEX)
- ⑤ Beef fat color score (BFS)
- ⑥ Mono unsaturated fatty acids / saturated fatty acids (MUS)





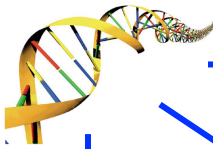
1 Dollar ~ 110 Yens

IC Algorithm, Results




Inoue K, Valente BD, Shoji N, Honda T, Oyama and Rosa GJM. Inferring phenotypic causal structures among meat quality traits and the application of a structural equation model in Japanese Black cattle. *Journal of Animal Science* 94:4133-4142, 2016.


SEM vs. MTM in Animal Breeding



Litter size, 0 days (y_1)



Litter size, 5 days (y_2)



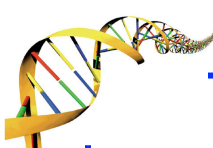
Litter weight at weaning (y_3)

$$\begin{cases} y_1 = u_1 + e_1 \\ y_2 = \lambda_{21}y_1 + u_2 + e_2 \\ y_3 = \lambda_{32}y_2 + u_3 + e_3 \end{cases}$$


Prediction

Valente BD, Rosa GJM, Gianola D, Wu XL and Weigel KA. Is structural equation modeling advantageous for the genetic improvement of multiple traits? *Genetics* 194: 561-572, 2013.


SEM vs. MTM in Animal Breeding



Litter size, 0 days (y_1)



Litter size, 5 days (y_2)



Litter weight at weaning (y_3)

$$\begin{cases} y_1 = u_1 + e_1 \\ y_2 = c \\ y_3 = \lambda_{32}c + u_3 + e_3 \end{cases}$$

Cross fostering

- SEM \approx MTM for predictions under stable conditions
- SEM superior if with modified scenarios (i.e., interventions)