Quantitative Genetic Models for Describing Simultaneous and Recursive Relationships Between Phenotypes

Daniel Gianola*†,1 and Daniel Sorensen‡

*Departments of Animal Sciences, Dairy Science and Biostatistics and Medical Informatics, University of Wisconsin, Madison, Wisconsin 53706, †Department of Animal and Aquacultural Sciences, Agricultural University of Norway, N-1432 Ås, Norway and ‡Department of Animal Breeding and Genetics, Danish Institute of Agricultural Sciences, 8830 Tjele, Denmark

Manuscript received December 11, 2003
Accepted for publication March 5, 2004

ABSTRACT

Multivariate models are of great importance in theoretical and applied quantitative genetics. We extend quantitative genetic theory to accommodate situations in which there is linear feedback or recursiveness between the phenotypes involved in a multivariate system, assuming an infinitesimal, additive, model of inheritance. It is shown that structural parameters defining a simultaneous or recursive system have a bearing on the interpretation of quantitative genetic parameter estimates (e.g., heritability, offspring-parent regression, genetic correlation) when such features are ignored. Matrix representations are given for treating a plethora of feedback-recursive situations. The likelihood function is derived, assuming multivariate normality, and results from econometric theory for parameter identification are adapted to a quantitative genetic setting. A Bayesian treatment with a Markov chain Monte Carlo implementation is suggested for inference and developed. When the system is fully recursive, all conditional posterior distributions are in closed form, so Gibbs sampling is straightforward. If there is feedback, a Metropolis step may be embedded for sampling the structural parameters, since their conditional distributions are unknown. Extensions of the model to discrete random variables and to nonlinear relationships between phenotypes are discussed.

MULTIVARIATE models are of great importance in applied, evolutionary, and theoretical quantitative genetics. For example, in animal and plant breeding, the value of a candidate for selection as a prospective parent of the next generation often is a function of several traits, e.g., protein yield, milk somatic cell count, fertility, and life span in dairy cattle or yield and resistance to disease in wheat. In evolutionary genetics, the effects of natural selection on mean fitness depend on and retaken by the values of elements of the genetic variance-covariance matrix between quantitative characters (e.g., Cheverud 1984). Walsh (2003) and B. Walsh and M. Lynch (unpublished results) give a discussion of the dynamics of quantitative traits under multivariate selection.

A schematic of the standard multivariate model used in quantitative genetics is displayed in Figure 1, where a two-trait system is represented; for simplicity, all other explanatory variables are omitted. The diagram depicts a 2 × 1 vector of phenotypic values (Y1, Y2) expressed as a function (typically linear) of genetic effects (U1, U2), usually taken to be of an additive type, and of environmental or residual effects (E1, E2). The genetic and environmental effects are assumed to be independently distributed random vectors, following the bivariate normal distributions N(0, G0) and N(0, R0), respectively. Here,

\[
G_0 = \begin{bmatrix} \sigma_{a1}^2 & \sigma_{a12} \\ \sigma_{a12} & \sigma_{a2}^2 \end{bmatrix}
\]

(1)

and

\[
R_0 = \begin{bmatrix} \sigma_{e1}^2 & \sigma_{e12} \\ \sigma_{e12} & \sigma_{e2}^2 \end{bmatrix}
\]

(2)

are genetic and residual variance-covariance matrices, respectively. For example, \(\sigma_{a1}^2\) is the variance between additive genetic effects affecting trait 1, and \(\sigma_{a2}\) is the residual covariance between traits 1 and 2.

The standard model depicted in Figure 1 does not allow for feedback or recursive relationships between phenotypes, which may be present in many biological systems. A classical example of feedback (that is, when changes of a quantity indirectly influence the quantity itself) is given by Haldane and Priestley (1905) and retaken by Turner and Stevens (1959) and by Wright (1960). These authors modeled feedback relationships between the concentration of CO2 in the air (A) and in the alveoli of the lungs (C) and the depth of respiration (D). As shown in Figure 2, Turner and Stevens (1959) posit that A affects C; in turn, C and D have a...
feedback relationship. Wright (1960) introduces residuals V and W for the C and D variables, respectively, and makes a further extension of the model. The extension consists of including a variable X for the actual concentration of CO₂ in the lungs; in Figure 2, U represents “random” errors of technique; this is a “measurement error” model (Warren et al. 1974; Joreskog and Sorbom 2001). In the Turner-Stevens model, the effect of C on D is through a coefficient λ_CD, whereas λ_DC gives the rate of change of C with respect to D. Suppose that these two coefficients are not null, so that feedback takes place. Ignoring the actual biology of the problem, a model such as that of Turner and Stevens (1959) or of Wright (1960) implies the following: if A is increased and the relationship between C and A is such that C increases as well, then D will increase provided λ_CD is positive. Further, if λ_CD is positive, then C will increase further. If all the loops go in the same direction, there is positive feedback, which might lead to some equilibrium or to an eventual breakdown of the system (Turner and Stevens 1959). A second example of reciprocal interaction is the classical supply-demand problem of

---

**Figure 1.**—Standard bivariate model used in quantitative genetics: Y₁ and Y₂ are the phenotypic values; U₁ and U₂ are additive genetic effects acting on the traits; E₁ and E₂ are residual effects. A single-headed arrow (e.g., A → B) indicates that variable A affects variable B.

**Figure 2.**—Haldane and Priestley (1905) respiration relations. Models for describing feedback relationships between concentrations of CO₂ in the respired air (A), in the alveoli of the lungs (C), and depth of respiration; V, W, and U are residuals. Wright (1960) introduces the variable X, the actual concentration of CO₂ in the alveoli; i.e., C is an imperfect measure.
Quantitative Genetics of Simultaneity and Recursiveness

1409

econometrics (Wright 1925; Johnston 1972; Judge et al. 1985). Also, the existence of feedback inhibition is well known in genetic regulation. For instance, the product of a metabolic pathway may bind to a gene product (enzyme) catalyzing a previous step, to prevent the channeling of additional molecules through the pathway (Lewin 1985). A discussion of the implications of interactive enzyme systems in genetics is in Kacser and Burns (1981). They write:

In vivo enzymes do not act in isolation, but are kinetically linked to other enzymes via their substrates and products. These interactions modify the effect of enzyme variation on the phenotype, depending on the nature and quantity of the other enzymes present. An output of such a system, say a flux, is therefore a systemic property, and its response to variation at one locus must be measured in the whole system.

It has been long recognized in economics (e.g., Havelmo 1943) that the existence of lagged or of instantaneous feedback (often referred to as “simultaneity”) and of recursiveness between variables has implications on the understanding of multivariate systems, and that special statistical techniques are required for inference. Curiously, Sewall Wright’s work on feedback mechanisms has received scant attention in population/quantitative genetics, in spite of his influence on the aforementioned fields and the pervasiveness of such mechanisms in regulation, as noted. An explanation may reside in the fact that even though path analysis was “extremely powerful in the hands of Wright” (Kempthorne 1969), the lack of matrix representations in his writings hampered a general understanding of the method. This is especially true of Wright’s treatment of reciprocal interaction with lags (Wright 1960), which is difficult to follow. Also, Goldberger (1972, p. 988) noted: “when there are more estimating equations than unknown parameters, path analysis gives no systematic guide to efficient estimation,” a situation known as overidentification. At any rate, social scientists eventually embedded path analysis into the general framework of simultaneous systems and gave it a formal statistical structure (Goldberger 1972; Goldberger and Duncan 1973; Duncan 1975; Joreskög and Sorbom 2001).

Our concern here is with the consequences of the existence of simultaneous (“feedback”) and recursive relationships between phenotypic values on quantitative genetic parameters, as well as with statistical methods for appropriate inference. The outline of the article is as follows. GENETIC CONSEQUENCES OF SIMULTANEITY IN A TWO-TRAIT SYSTEM and GENETIC CONSEQUENCES OF RECURRENCE illustrate effects of simultaneity or recursiveness on simple two-trait systems. In particular, formulas are presented for heritability, for the offspring-parent regression, and for the genetic and environmental correlations between traits. Next, MATRIX REPRESENTATIONS shows that if four phenotypes are involved in a multivariate system, there can be as many as 128 mutually exclusive and exhaustive models for describing the relationships between phenotypic values. Formulas pertinent to multivariate selection (e.g., best prediction of genetic values) are given as well. Then, the likelihood function and identification of parameters sections are presented. BAYESIAN MODEL addresses statistical inference in a structural equations model from a Bayesian perspective. It is shown in the FULLY CONDITIONAL POSTERIOR DISTRIBUTIONS that, under normality assumptions, most conditional posterior distributions arising in the multivariate system are in recognizable form. The implication is that software for standard multiple-trait analysis of quantitative traits via Gibbs sampling (a Markov chain Monte Carlo method) can be adapted to handle simultaneity and recursiveness in a fairly direct manner. The article concludes with suggestions on how the approach can be extended to a wider class of models.

GENETIC CONSEQUENCES OF SIMULTANEITY IN A TWO-TRAIT SYSTEM

Let \( y_{i1} \) and \( y_{i2} \) be measurements on traits 1 and 2 observed in individual \( i \). For example, in an animal breeding setting, \( y_{i1} \) may represent the milk yield of dairy cow \( i \) and \( y_{i2} \) may be a proxy for the level of some disease (e.g., milk somatic cell count as an indicator of mastitis, a bacterial-related inflammation of the mammary gland). Suppose that biological knowledge admits that production affects disease and, in turn, that disease affects production. As noted earlier, we refer to this as a simultaneous or instantaneous feedback system, following econometric terminology (Zellner 1979; Judge et al. 1985). Assume that the relationship between production and disease can be represented by the two-equation linear system,

\[
y_{i1} = \lambda_{12} y_{i2} + \mathbf{x}_{i1}' \mathbf{b}_1 + u_{i1} + e_{i1},
\]

and

\[
y_{i2} = \lambda_{21} y_{i1} + \mathbf{x}_{i2}' \mathbf{b}_2 + u_{i2} + e_{i2}.
\]

Here, \( \lambda_{12} \) is the rate of change of level of production with respect to a disease index and \( \lambda_{21} \) is the gradient of disease with respect to production. A priori, one might expect \( \lambda_{12} \) to be negative and \( \lambda_{21} \) to be positive, since high output may be associated with “stress” in the dairy cow. It is assumed that \( \lambda_{12} \) and \( \lambda_{21} \) are homogeneous across individuals, but this can be relaxed. The vectors \( \mathbf{b}_1 \) and \( \mathbf{b}_2 \), often called fixed effects in the statistical literature (Searle 1971), are some location parameters such as age of the cow, parity, or breed affecting production (net of disease) and disease (net of production), respectively. Further, \( \mathbf{x}_{i1} \) and \( \mathbf{x}_{i2} \) are known incidence row vectors; whenever the same factors affect the two traits, the covariates are such that \( \mathbf{x}_{i1} = \mathbf{x}_{i2} = \mathbf{x}_i \), say. Finally, \( u_{i1} \) and \( u_{i2} \) are additive genetic effects (Fisher 1918; Bulmer 1980) intervening in the system and \( e_{i1} \) and \( e_{i2} \) are model residuals.
It is important to note that $u_{i1}$ ($u_{i2}$) above can be construed as an additive genetic effect affecting only production (disease) if and only if $\lambda_{i1}(\lambda_{i2}) = 0$; it is shown subsequently that $u_{i1}$ and $u_{i2}$ may affect each of the two traits. On the other hand, it is legitimate to view $u_{i1}$, say, as the additive genetic component of the linear combination $y_{i1} - \lambda_{i2}y_{i2}$. This is suggested by a rewriting of (3) and (4) as

\[ y_{i1} - \lambda_{i2}y_{i2} = x_{i1}'\beta_1 + u_{i1} + e_{i1} \tag{5} \]
\[ y_{i2} - \lambda_{i1}y_{i1} = x_{i2}'\beta_2 + u_{i2} + e_{i2} \tag{6} \]

More generally, $u_{i1}$ and $u_{i2}$ are genetic effects “controlling” system (5)–(6).

When a specification such as system (5)–(6) holds, generalized least-squares estimates (or maximum-likelihood estimates under standard normality assumptions) of the structural parameters $\lambda_{i1}$ (or $\lambda_{i2}$) are biased and inconsistent if obtained from a subset of equations (Johnston 1972). For instance, if a univariate analysis of trait 1 is conducted including $y_{i2}$ as a covariate (but ignoring the submodel for trait 2), $\lambda_{i1}$ is estimated with an upward bias (provided $\lambda_{i2} < 1$). The effects on model parameters are more cryptic as the dimensionality of the system increases, for example, when a system of five mutually interacting response variables is analyzed with a three-trait model.

The implications of system (5)–(6) on the interpretation of some parameters of interest in quantitative genetics analysis are considered next.

**Heritability**: Use (4) in (3) and solve for $y_{i1}$, to obtain the “reduced” model (a term from econometrics),

\[ y_{i1} = \mu_{i1} + u_{i1}^* + e_{i1}^* \tag{7} \]

where

\[ \mu_{i1}^* = \frac{x_{i1}'\beta_1 + \lambda_{i1}x_{i2}'\beta_2}{1 - \lambda_{i1}\lambda_{i2}} \]
\[ u_{i1}^* = \frac{u_{i1} + \lambda_{i1}u_{i2}}{1 - \lambda_{i1}\lambda_{i2}} \]

and

\[ e_{i1}^* = \frac{e_{i1} + \lambda_{i1}e_{i2}}{1 - \lambda_{i1}\lambda_{i2}} \]

with the random effects $u_{i1}^* \sim N(0, \sigma_{u1}^2)$ and $e_{i1}^* \sim N(0, \sigma_{e1}^2)$ being independently distributed. Note that both $\beta_1$ and $\beta_2$ intervene in $\mu_{i1}^*$, and that $u_{i1}^*$ is a linear combination of the system genetic effects $u_{i1}$ and $u_{i2}$. An implication of this is that estimates of location parameters and of predictions of random effects from standard univariate or multivariate analyses can be interpreted differently if simultaneity holds. Suppose data are missing at random (i.e., that selection is ignorable). In this case the fraction of the variance of trait 1 that can be attributed to additive genetic effects, or coefficient of heritability, is

\[ h_i^2 = \frac{\sigma_{u1}^2}{\sigma_{u1}^2 + \sigma_{e1}^2} \]

where

\[ \sigma_{u1}^2 = \frac{(1/(1 - \lambda_{i1}))^2(\sigma_{u1}^2 + 2\lambda_{i1}\sigma_{u2} + \lambda_{i1}^2\sigma_{e1}^2)}{(1/(1 - \lambda_{i2}))^2(\sigma_{u1}^2 + \sigma_{u2}^2 + 2\lambda_{i1}\sigma_{u12} + \lambda_{i1}\sigma_{u2} + \lambda_{i1}^2\sigma_{e1}^2)} + \lambda_{i1}(\sigma_{u2}^2 + \sigma_{e1}^2) \]

\[ = \frac{\sigma_{u1}^2 + 2\lambda_{i1}\sigma_{u2} + \lambda_{i1}^2\sigma_{e1}^2}{\sigma_{u1}^2 + \sigma_{u2}^2 + 2\lambda_{i1}\sigma_{u12} + \sigma_{u2} + \lambda_{i1}^2(\sigma_{u2}^2 + \sigma_{e1}^2)} \tag{8} \]

Observe that the apportionment of variance into genetic and environmental components depends nontrivially on the structural parameter $\lambda_{i2}$, but not on $\lambda_{i1}$ (the opposite being true for trait 2). If $\sigma_{u1}^2 = \sigma_{u2}^2 = 0$, the linear combinations or “composite” traits $y_{i1} - \lambda_{i2}y_{i2}$ and $y_{i2} - \lambda_{i1}y_{i1}$ are statistically independent (by virtue of normality); however, $h_i^2$ would still depend on $\lambda_{i2}$, as

\[ h_i^2 = \frac{\sigma_{u1}^2 + \lambda_{i1}^2\sigma_{e1}^2}{\sigma_{u1}^2 + \sigma_{u2}^2 + \lambda_{i1}^2(\sigma_{u12}^2 + \sigma_{e1}^2)} \]

The corresponding expression for the heritability of trait 2 is

\[ h_i^2 = \frac{\sigma_{u2}^2 + \lambda_{i2}^2\sigma_{e1}^2}{\sigma_{u2}^2 + \sigma_{u1}^2 + \lambda_{i2}^2(\sigma_{u12}^2 + \sigma_{e1}^2)} \]

**Regression of offspring on parent**: Let $y_{i1}$ be a record for trait 1 measured on the offspring of an individual with phenotype $y_{i1}$. The offspring-parent covariance is

\[ \text{Cov}(y_{i1}, y_{i1}) = \text{Cov}\left[ (x_{i1}'\beta_1 + \lambda_{i1}x_{i2}'\beta_2 + (u_{i1} + \lambda_{i1}u_{i2}) + (e_{i1} + \lambda_{i1}e_{i2}), \right. \]
\[ \left. 1 - \lambda_{i1}\lambda_{i2} \right] \]
\[ = \frac{1}{2} \left( (\sigma_{u1}^2 + \lambda_{i1}^2\sigma_{e1}^2) + \lambda_{i1}\sigma_{u12} \right) \]

Assuming zero covariances between environmental effects affecting records taken on different individuals, under additive inheritance one has that

\[ \text{Cov}(y_{i1}, y_{i1}) = \frac{1}{2} \left( \sigma_{u1}^2 + \lambda_{i1}^2\sigma_{e1}^2 + \lambda_{i1}\sigma_{u12} \right) \]

which is half of the numerator of the expression leading to (8). In the absence of feedback ($\lambda_{i2} = 0$), the offspring-parent covariance is always positive and equal to $\sigma_{u1}^2/2$. Under simultaneity, however, this covariance could be negative, provided that

\[ \sigma_{u12} > \frac{(\sigma_{u1}^2 + \lambda_{i1}^2\sigma_{e1}^2)}{2\lambda_{i2}} \]

The implication is that a variance component analysis based on the reduced model would, necessarily (because of the parameterization), return a positive fitted value of the offspring-parent covariance. On the other hand, this may not be so under a simultaneous equations model. If the observed covariance is negative, this should be construed as evidence against a specification failing to accommodate simultaneity, although there may be other reasons (e.g., maternal effects) for a negative offspring-parent covariance.

The regression of offspring on parent is

\[ b_{op} = \frac{1}{\sigma_{u1}^2 + \sigma_{u2}^2 + 2\lambda_{i1}\sigma_{u12} + \sigma_{u2} + \lambda_{i1}^2(\sigma_{u2}^2 + \sigma_{e1}^2)} \]

\[ \left( \frac{1}{\sigma_{u1}^2 + \lambda_{i1}(\sigma_{u2}^2 + \lambda_{i1}\sigma_{e1}^2) + \lambda_{i1}\sigma_{u12} + \lambda_{i1}^2(\sigma_{u2}^2 + \sigma_{e1}^2)} \right) \tag{9} \]
yielding the usual \( \sigma^2_{y_1}/[2(\sigma^2_{y_1} + \sigma^2_{y_2})] \) in the absence of feedback (\( \lambda_{12} = \lambda_{21} = 0 \)).

**Regression of one variable on another:** If \( x'_1 = x'_2 = x' \), the reduced models as in (7) become

\[
y_1 = x'_1\left(\frac{\beta_1 + \lambda_{12}\beta_2}{1 - \lambda_{12}^2} \right) + \left(\frac{u_1 + \lambda_{12}u_2}{1 - \lambda_{12}^2}\right) + \left(\frac{e_1 + \lambda_{12}e_2}{1 - \lambda_{12}^2}\right) = x'_1\beta'_1 + u'_1 + e'_1
\]

and

\[
y_2 = x'_2\left(\frac{\beta_2 + \lambda_{21}\beta_1}{1 - \lambda_{21}^2} \right) + \left(\frac{u_2 + \lambda_{21}u_1}{1 - \lambda_{21}^2}\right) + \left(\frac{e_2 + \lambda_{21}e_1}{1 - \lambda_{21}^2}\right) = x'_2\beta'_2 + u'_2 + e'_2
\]

(10)

Under normality, the regression function of trait 2 on trait 1 is

\[
E(y_2|y_1) = x'_2\beta'_2 + \frac{\sigma_{y_2}}{\sigma_{y_1}^2 + \sigma_{y_2}^2}(y_1 - x'_1\beta'_1).
\]

**Genetic and environmental correlations:** The reduced models (10) and (11) lead to

\[
\text{Cov}(y_1, y_2) = \text{Cov}(u'_1, u'_2) + \text{Cov}(e'_1, e'_2),
\]

where

\[
\text{Cov}(u'_1, u'_2) = \frac{(1 + \lambda_{12}\lambda_{21})\sigma_{u_1} + \lambda_{21}\sigma_{u_2}^2 + \lambda_{12}\sigma_{u_2}^2}{(1 - \lambda_{12}^2)^2}
\]

Similarly

\[
\text{Cov}(e'_1, e'_2) = \frac{(1 + \lambda_{12}\lambda_{21})\sigma_{e_1} + \lambda_{21}\sigma_{e_2}^2 + \lambda_{12}\sigma_{e_2}^2}{(1 - \lambda_{12}^2)^2}
\]

The genetic and environmental covariances depend on the \( \lambda \) coefficients and on the appropriate variances and covariances of each of the two composite traits.

The genetic correlation is

\[
\text{Corr}(u'_1, u'_2) = \frac{(1 + \lambda_{12}\lambda_{21})\sigma_{u_1} + \lambda_{21}\sigma_{u_2}^2 + \lambda_{12}\sigma_{u_2}^2}{\sqrt{(\sigma_{u_1}^2 + 2\lambda_{12}\sigma_{u_1} + \lambda_{12}\sigma_{u_2}^2)(\sigma_{u_2}^2 + 2\lambda_{21}\sigma_{u_2} + \lambda_{21}\sigma_{u_1}^2)}}
\]

(12)

and the expression for the residual correlation is similar. If \( \sigma_{y_2} = 0 \) (i.e., when the composite traits are genetically uncorrelated), (12) becomes

\[
\text{Corr}(u'_1, u'_2) = \frac{\lambda_{12}\sigma_{u_1}^2 + \lambda_{21}\sigma_{u_2}^2}{\sqrt{(\sigma_{u_1}^2 + \lambda_{12}^2\sigma_{u_2}^2)(\sigma_{u_2}^2 + \lambda_{21}^2\sigma_{u_1}^2)}}
\]

**GENETIC CONSEQUENCES OF RECURSIVENESS IN A TWO-TRAIT SYSTEM**

A recursive specification postulates, for instance, that \( y_{1t} \) affects \( y_{2t} \) but that the latter variable has no effect on \( y_{1t} \). An example is the maternal-effects model proposed by Falconer (1965) and examined by Koehn and Thompson (1977). This model postulates that the phenotype of an offspring is affected by the phenotype of its dam. In pigs, for instance, it is known that females born in larger litters tend to produce smaller litters (leading to negative values of the \( \lambda \) coefficients); conversely, females in smaller litters are expected to produce larger litters, etc. A recursive specification can be obtained from the “full model” by setting \( \lambda_{12} = 0 \) in (3) or (5), so that the system is now

\[
y_1 = x'_1\beta'_1 + u_1 + e_1,
\]

(13)

and

\[
y_2 = \lambda_{21}y_{1t} + x'_2\beta'_2 + u_2 + e_2.
\]

(14)

Assuming that the incidence vectors are such that \( x'_1 = x'_2 = x' \), use of (13) in (14) gives a reduced model for \( y_{2t} \):

\[
y_{2t} = x'_2(\beta'_2 + \lambda_{21}\beta'_1) + (u_2 + \lambda_{21}u_1) + (e_2 + \gamma_{21}e_1) = x'_2\beta'_2 + u'_2 + e'_2,
\]

where

\[
u'_2 \sim N(0, \sigma^2_{u_2} + 2\lambda_{21}\sigma_{u_2} + \lambda_{21}^2\sigma_{u_1}^2)
\]

and

\[
e'_2 \sim N(0, \sigma^2_{e_2} + 2\lambda_{21}\sigma_{e_2} + \lambda_{21}^2\sigma_{e_1}^2)
\]

so that

\[
\text{Var}(y_{2t}) = \sigma^2_{u_2} + \sigma^2_{e_2} + 2\lambda_{21}\sigma_{u_2} + \lambda_{21}^2\sigma_{u_1}^2 + \lambda_{21}^2\sigma_{e_1}^2 + 2\lambda_{21}\sigma_{e_2} + \lambda_{21}^2\sigma_{e_1}^2.
\]

**Heritability:** The heritability of trait 1 is the usual \( h^2 = \sigma^2_{y_1}/(\sigma^2_{y_1} + \sigma^2_{y_2}) \). Here, \( \sigma^2_{y_1} \) and \( \sigma^2_{y_2} \) are the additive genetic and residual variances of trait 1, respectively, contrary to the simultaneity situation, where these dispersion parameters pertain to the variation of genetic and residual effects affecting the distribution of \( y_1 - y_2 \). The coefficient of heritability of trait 2 has the form of (8), but with \( \lambda_{21} \) instead of \( \lambda_{12} \):

\[
h^2_2 = \frac{\sigma^2_{y_1} + 2\lambda_{21}\sigma_{y_1} + \lambda_{21}^2\sigma_{y_1}^2}{\sigma^2_{y_2} + \sigma^2_{y_2} + 2\lambda_{21}\sigma_{y_1} + \lambda_{21}^2\sigma_{y_1}^2 + 2\lambda_{21}\sigma_{y_2} + \lambda_{21}^2\sigma_{y_2}^2}.
\]

(15)

**Regression of offspring on parent:** The regression of offspring on parent depends on the trait or pairs of traits involved. Using the same notation as in the section for simultaneity, the offspring-parent covariance for trait 1 is

\[
\text{Cov}(y_{1t}, y_{1t}) = \text{Cov}(x'_1\beta'_1 + u_1 + e_1, x'_1\beta'_1 + u_1 + e_1) = \frac{1}{2} \sigma^2_{y_1}.
\]

Hence the regression of offspring on parent for trait 1 is simply \( h^2_1 \), the standard result from assuming additive inheritance.

Let \( y_{2t} \) be a record measured for trait 2 on the offspring of an individual with phenotype \( y_{2t} \). The offspring-parent regression, assuming that between-generation environmental effects are uncorrelated, has a similar form
Consider now the covariance between $y_{12}$, a record for trait 2 measured on the offspring of an individual with phenotype $y_{1}$. The offspring-parent covariance between such records is now

$$\text{Cov}(u_{2}, \lambda_{21}u_{1}, u_{i2}) = \frac{\sigma_{a_{21}} + \lambda_{21}\sigma^{2}_{a_{21}}}{2}$$

and the regression coefficient is

$$\lambda_{0,12} = \frac{1}{2} \left( \lambda_{21}h_{22}^{2} + \frac{\sigma_{a_{21}}}{\sigma^{2}_{a_{21}} + \sigma^{2}_{e_{12}}} \right). \quad (16)$$

Conversely, with $y_{1}$ being now a record for trait 1 measured on the offspring of an individual with phenotypic value $y_{2}$ the offspring-parent regression is

$$\lambda_{0,21} = \frac{\sigma_{a_{21}} + \lambda_{21}\sigma^{2}_{a_{21}}}{2[\sigma^{2}_{a_{21}} + \sigma^{2}_{e_{12}} + 2\lambda_{21}(\sigma_{a_{21}} + \sigma_{e_{12}}) + \lambda_{21}^{2}(\sigma^{2}_{a_{11}} + \sigma^{2}_{e_{11}})]}.$$  \quad (17)

**Regression of one variable on another:** Recall that $y_{1} = x'_{1}b_{1} + u_{1} + e_{1}$ and that $y_{2} = x'_{2}(\beta_{2} + \lambda_{21}b_{1}) + (u_{2} + \lambda_{21}u_{1}) + (e_{2} + \lambda_{21}e_{1})

$$= x'_{2}\beta_{2} + u_{2}^{*} + e_{2}^{*}.$$ Then

$$E(y_{2}|y_{1}) = x'_{2}(\beta_{2} + \lambda_{21}b_{1}) + \left(\lambda_{21} + \frac{\sigma_{a_{21}}}{\sigma^{2}_{a_{11}} + \sigma^{2}_{e_{11}}}\right)(y_{1} - x'_{2}b_{1}). \quad (18)$$

Conversely, the regression function of $y_{12}$ on $y_{1}$ is

$$E(y_{12}|y_{1}) = x'_{1}(b_{1} + \lambda_{21}b_{1}) + \left(\lambda_{21} + \frac{\sigma_{a_{21}}}{\sigma^{2}_{a_{11}} + \sigma^{2}_{e_{11}}}\right)(y_{1} - x'_{1}b_{1}). \quad (19)$$

The two preceding expressions reduce to the usual formulas under bivariate normality by letting $\lambda_{21} = 0$.

**Genetic and environmental correlations:** The genetic covariance between the two traits is

$$\text{Cov}(u_{1}, u_{2}^{*}) = \text{Cov}(u_{1}, u_{2} + \lambda_{21}u_{1}) = \sigma_{a_{12}} + \lambda_{21}\sigma^{2}_{a_{11}},$$

so that the genetic correlation is

$$\text{Corr}(u_{1}, u_{2}^{*}) = \frac{\sigma_{a_{12}} + \lambda_{21}\sigma^{2}_{a_{11}}}{\sqrt{\sigma^{2}_{a_{11}}(\sigma^{2}_{a_{12}} + 2\lambda_{21}\sigma^{2}_{a_{12}} + \lambda_{21}^{2}\sigma^{2}_{a_{11}})}}. \quad (20)$$

If $\sigma_{a_{12}} = 0$

$$\text{Corr}(u_{1}, u_{2}^{*}) = \frac{\lambda_{21}}{\sqrt{(\lambda_{21}^{2} + (\sigma^{2}_{a_{11}}/\sigma^{2}_{a_{12}}))}}.$$ in which case the sign of the genetic correlation depends on the sign of $\lambda_{21}$. Further, if the traits are scaled such that $\sigma^{2}_{a_{11}} = \sigma^{2}_{a_{12}} = 1$, the correlation is $\lambda_{21}/\sqrt{(1 + \lambda_{21}^{2})}$. It is interesting to observe that when genotypes are expressed in units of standard deviation, the genetic correlation is driven entirely by $\lambda_{21}$, which is a gradient operating at the phenotypic level.

Likewise, the environmental covariance and correlation are

$$\text{Cov}(e_{1}, e_{2}^{*}) = \text{Cov}(e_{1}, e_{2} + \lambda_{21}e_{1}) = \sigma_{e_{12}} + \lambda_{21}\sigma^{2}_{e_{11}},$$

and

$$\text{Corr}(e_{1}, e_{2}^{*}) = \frac{\sigma_{e_{12}} + \lambda_{21}\sigma^{2}_{e_{11}}}{\sqrt{\sigma_{e_{11}}^{2}(\sigma^{2}_{e_{12}} + 2\lambda_{21}\sigma^{2}_{e_{12}} + \lambda_{21}^{2}\sigma^{2}_{e_{11}})}}, \quad (21)$$

respectively.

**Matrix Representations**

**Many possible models:** A multivariate system may involve many response variables, as well as different levels of simultaneity and recursiveness. When the models comprise more than two traits, the issues and principles are as discussed above but the algebra is awkward. For example, consider the simultaneous-equation model for three traits given in Figure 3. Here, the three response variables $Y_{1}$, $Y_{2}$, and $Y_{3}$ have mutually reciprocal effects, so that there are six $\lambda$ coefficients or structural parameters in the model.

Several different models can be derived as special cases of the specification given in Figure 3. There are 64 models that can be viewed as “nested” within the diagram depicted. In general, for $K$ response variables there are $K(K - 1)$ structural coefficients ($\lambda$’s) in a fully simultaneous model. Since, in a given model, each coefficient can take the value $\lambda_{i}$ or be constrained to be 0 (when there is no “effect” of variable $j$ on variable $i$ in the latter case), there can be as many as $2^{K(K-1)}$ possible models for explaining relationships between the phenotypic variables; in practice, however, many of the models can be discarded on mechanistic grounds. For example, if all $\lambda$’s are set equal to 0 in Figure 3, this yields the standard trivariate model used for quantitative genetic analysis of three traits. Some other models that can arise are illustrated in Figures 4–6. A “cyclically recursive” model is depicted in Figure 4. Here, the causal relationship modeled is $Y_{1} \rightarrow Y_{2} \rightarrow Y_{3} \rightarrow Y_{1} \ldots$. For instance, consider a hypothetical situation where $Y_{1}$, $Y_{2}$, and $Y_{3}$ are phenotypic values in sibships of size 3, with the subscript indicating birth order. It may be that the chain of influences is such that the older sib (with phenotype $Y_{1}$) affects the second sib ($Y_{2}$) and so on, with the loop closing via an influence of the youngest on the oldest sib.

In Figure 5, it is hypothesized that $Y_{1}$ has an effect of the recursive type on both $Y_{2}$ and $Y_{3}$, but that there is simultaneity between $Y_{2}$ and $Y_{3}$; this is referred to as a recursive-simultaneous model. Here, $Y_{1}$ might be the concentration of a hormone regulating the production of two metabolites that are involved in a feedback relationship. In Figure 6, there is simultaneity between $Y_{2}$ and $Y_{3}$ with these two variables affecting $Y_{1}$; this is a simultaneous-recursive model: two biochemical products ($Y_{2}$ and $Y_{3}$) involved in the regulation of $Y_{1}$ may...
Figure 3.—Simultaneous-equations model for three variables: $Y_1$, $Y_2$, and $Y_3$ are the phenotypic values; $U_1$, $U_2$, and $U_3$ are additive genetic effects acting on the system; $E_1$, $E_2$, and $E_3$ are residual effects. A single-headed arrow (e.g., $A \rightarrow B$) indicates that variable $A$ affects variable $B$. Double-headed arrows denote correlations between pairs of variables. $\lambda_{ij}$ indicates the rate of change of variable $i$ with respect to variable $j$.

interact reciprocally toward the establishment of some equilibrium. Clearly, there is a constellation of modeling alternatives.

The preceding discussion illustrates that a general representation is needed for describing the full range of possibilities. For example, the two-variate simultaneous-

Figure 4.—Fully recursive model for three variables: $Y_1$, $Y_2$, and $Y_3$ are the phenotypic values; $U_1$, $U_2$, and $U_3$ are additive genetic effects acting on the system; $E_1$, $E_2$, and $E_3$ are residual effects. A single-headed arrow (e.g., $A \rightarrow B$) indicates that variable $A$ affects variable $B$. Double-headed arrows denote correlations between pairs of variables. $\lambda_i$ indicates the rate of change of variable $i$ with respect to variable $j$. 
equations system of Equations 3 and 4 can be put in matrix form as
\[
\begin{bmatrix}
1 & -\lambda_{21} \\
-\lambda_{21} & 1
\end{bmatrix}
\begin{bmatrix}
y_{i1} \\
y_{i2}
\end{bmatrix}
= \begin{bmatrix}
x'_{i1} & 0 \\
0 & x'_{i2}
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix}
+ \begin{bmatrix}
u_{i1} \\
u_{i2}
\end{bmatrix}
+ \begin{bmatrix}
e_{i1} \\
e_{i2}
\end{bmatrix}.
\]
(22)

This representation embeds four models \([K = 2, \text{ so } 2^{K(K-1)} = 4]\), including the simultaneous one. The other three models are the standard bivariate specification \((\lambda_{12} = \lambda_{31} = 0)\) and two recursive models \((\lambda_{12} = 0\) when \(y_1\) “affects” \(y_2\), but without a reciprocal effect; \(\lambda_{21} = 0\) when \(y_2\) affects \(y_1\)).

**Statistical structure:** Let there be \(K\) traits observed on individual or “cluster” (e.g., a family) \(i (i = 1, 2, \ldots, N)\), and write the system as
\[
Ay_i = X\beta + u_i + e_i,
\]
(23)
where \(y_i\) is a \(K \times 1\) vector of phenotypic measurements on the \(K\) traits of individual \(i\); \(A\) is a \(K \times K\) matrix containing at most \(K(K-1)\) unknown \(\lambda\) coefficients (all diagonal elements are equal to 1); and \(X\) is a \(K \times \Sigma_{j=1}^{K} p_j\) known incidence matrix with the form
\[
X =
\begin{bmatrix}
x'_{i1} & 0 & 0 & 0 \\
0 & x'_{i2} & 0 & 0 \\
& & \ddots & \ddots \\
& & & x'_{i(K-1)} & 0 \\
0 & 0 & \cdots & 0 & x'_{ik}
\end{bmatrix},
\]
where \(x'_{ij}\) is a row vector with \(p_j\) elements. Without loss of generality, it is assumed that \(X\) has full-column rank. Further, the location vector \(\beta\) is such that
\[
\beta =
\begin{bmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_{K-1} \\
\beta_K
\end{bmatrix},
\]
where \(\beta_j (j = 1, 2, \ldots, K)\) is \(p_j \times 1\). The vector \(u_i\) contains additive genetic effects of individual \(i\) for the \(K\) traits and, similarly, \(e_i\) is a vector of residual effects, distributed independently of \(u_i\).

If \(A\) has full rank, the reduced form of the model is
\[
y_i = A^{-1}X\beta + A^{-1}u_i + A^{-1}e_i
\]
\[
= \mu^* + u^* + e^*.
\]
(24)
where \(\mu^* = A^{-1}X\beta, u^* = A^{-1}u_i\) and \(e^* = A^{-1}e_i\). For example, in a two-trait simultaneous model, the elements of \(\mu^*, u^*,\) and \(e^*\) take the form given in (7). Assume now that
\[
u_i | G_0 \sim N(0, G_0); \quad i = 1, 2, \ldots, N,
\]
(25)
and
\[
e_i | R_0 \sim N(0, R_0)
\]
(26)
are mutually independent. Hence, the vectors of genetic and residual effects involved in the correlations shown

\[
\begin{array}{c}
\end{array}
\]

**Figure 5.**—Recursive simultaneous model for three variables: \(Y_1, Y_2,\) and \(Y_3\) are the phenotypic values; \(U_1, U_2,\) and \(U_3\) are additive genetic values acting on the system; \(E_1, E_2,\) and \(E_3\) are residual effects. A single-headed arrow (e.g., \(A \rightarrow B\)) indicates that variable \(A\) affects variable \(B\). Double-headed arrows denote correlations between pairs of variables. \(\lambda_i\) indicates the rate of change of variable \(i\) with respect to variable \(j\).
in Figures 3–6 follow multivariate normal distributions with covariance matrices $\mathbf{G}_0$ and $\mathbf{R}_0$, respectively, each having order $3 \times 3$. This implies that

$$u^*_i|\boldsymbol{A}, \mathbf{G}_0 \sim N(0, \mathbf{A}^{-1}\mathbf{G}_0\mathbf{A}^{-1}),$$

(27) and

$$e^*_i|\boldsymbol{A}, \mathbf{R}_0 \sim N(0, \mathbf{A}^{-1}\mathbf{R}_0\mathbf{A}^{-1})$$

are also independently distributed. Further, the marginal distribution of the phenotypic values for individual $i$ is

$$y_i|\boldsymbol{A}, \mathbf{R}_0, \mathbf{G}_0 \sim N(\mathbf{A}^{-1}\mathbf{X}\boldsymbol{p}, \mathbf{A}^{-1}(\mathbf{R}_0 + \mathbf{G}_0)\mathbf{A}^{-1})$$

(29)

**Genetic parameters and functions thereof:** The “multivariate heritability and coheritability” matrix can be defined as

$$\mathbf{H} = \mathbf{A}^{-1}\mathbf{G}_0\mathbf{A}^{-1}[\mathbf{A}^{-1}(\mathbf{R}_0 + \mathbf{G}_0)\mathbf{A}^{-1}]^{-1}$$

$$= \mathbf{A}^{-1}\mathbf{G}_0(\mathbf{R}_0 + \mathbf{G}_0)^{-1}\mathbf{A}.$$  

(30)

In the absence of simultaneity or recursiveness, $\mathbf{H} = \mathbf{G}_0(\mathbf{R}_0 + \mathbf{G}_0)^{-1}$, since $\mathbf{A}$ would be an identity matrix of order $K$ in this case. Note that the trace of (30),

$$\text{tr}(\mathbf{H}) = \text{tr}[\mathbf{G}_0(\mathbf{R}_0 + \mathbf{G}_0)^{-1}],$$

is free of the $\lambda$ coefficients. Now, using the measurements taken on individual $i$, the best predictor of $u^*_i$, in the sense of minimizing the mean square error of prediction among all possible functions of the data (Henderson 1973), is given by the conditional expectation function

$$E(u^*_i|y_i) = E(u^*_i) + \text{Cov}(u^*_i, y_i)\text{Var}^{-1}(y_i)(y_i - \Lambda^{-1}\mathbf{X}\boldsymbol{p})$$

$$= \mathbf{H}(y_i - \Lambda^{-1}\mathbf{X}\boldsymbol{p}).$$

(31)

In multiple-trait selection, animal and plant breeders are often interested in improving a linear combination of genetic values; e.g., $T_i = \mathbf{v}'u^*_i$, where $\mathbf{v}$ is a known $K \times 1$ vector of relative economic values (Smith 1936; Hazel 1943), and $u^*_i$ contains the “true” genetic values affecting the traits. Recall that the genetic values are $u_i$ only in the absence of feedbacks or recursiveness. Suppose the $N$ candidates are independently distributed, so that the density of the joint distribution of all genetic and phenotypic values is given by

$$p(u^*_i, y_i|\text{parameters}) = \prod_{i=1}^N p(u^*_i, y_i|\text{parameters}).$$

This is what Henderson (1963, 1973) termed an “equal information” situation. The best predictor of the “merit function” $T_i$ is

$$\hat{T}_i = E(v'u^*_i|y_i) = \mathbf{v}'\mathbf{H}(y_i - \Lambda^{-1}\mathbf{X}\boldsymbol{p})$$

$$= \mathbf{b}'(y_i - \Lambda^{-1}\mathbf{X}\boldsymbol{p}),$$

(32)

where

$$\mathbf{b} = \mathbf{H}'\mathbf{v} = \Lambda'(\mathbf{R}_0 + \mathbf{G}_0)^{-1}\mathbf{G}_0\Lambda^{-1}\mathbf{v}$$

(33)

is the classical “selection index” solution to the Smith-Hazel equations.

![Figure 6.—Simultaneous-recursive model for three variables: $Y_1$, $Y_2$, and $Y_3$ are the phenotypic values; $U_1$, $U_2$, and $U_3$ are additive genetic effects acting on the system; $E_1$, $E_2$, and $E_3$ are residual effects. A single-headed arrow (e.g., $A \rightarrow B$) indicates that variable $A$ affects variable $B$. Double-headed arrows denote correlations between pairs of variables. $\lambda_i$ indicates the rate of change of variable $i$ with respect to variable $j$.](image-url)
Var(y)_ib = Cov(y, iv'_u). Suppose that selection of a truncation type is based on \( \bar{T}_i \) in (32), such that a proportion \( \alpha \) of the candidates is kept as parents of the following generation. From the forms of (32) and (33), it follows that the mean of the distribution of \( \bar{T}_i \) in the unselected individuals is 0, since \( E(y)_i = \Lambda^{-1}X_\beta \). Under normality assumptions, standard theory (e.g., Bulmer 1980; Falconer and Mackay 1996) gives a mean of the selected individuals, \( E_\alpha(\bar{T}_i) = \frac{i\mathbf{Var}(\bar{T}_i)}{i\mathbf{Var}(\bar{T}_i) + \mathbf{Var}(\epsilon_i^o)}(y_i - E(y)_i) \].

For a two-trait simultaneous system, it was seen earlier that \( \frac{\mathbf{Var}(u_{11}^o)}{\mathbf{Var}(u_{11}^o) + \mathbf{Var}(\epsilon_{11}^o)} = \frac{\sigma_{z1}^2 + 2\lambda_{12}\sigma_{z2} + \lambda_{12}^2\sigma_{z1}^2}{\sigma_{z1}^2 + \sigma_{z2}^2 + 2\gamma_{12}(\sigma_{z12} + \sigma_{z2}) + \lambda_{12}^2(\sigma_{z12}^2 + \sigma_{z2}^2)} \) and \( E(y)_i = \mu_y + \lambda_{12}\mu_z \). Hence \( E_\alpha(\bar{T}_i) = \frac{i(\sigma_{z1}^2 + 2\gamma_{12}\sigma_{z2} + \lambda_{12}^2\sigma_{z1}^2)}{\sqrt{\sigma_{z1}^2 + \sigma_{z2}^2 + 2\gamma_{12}(\sigma_{z12} + \sigma_{z2}) + \lambda_{12}^2(\sigma_{z12}^2 + \sigma_{z2}^2)}} \). When \( h_{12} = 0 \), this reduces to the usual \( E_\alpha(\bar{T}_i) = h_i = \alpha h_i^o \), provided selection is based on \( \bar{T}_i = h_i[y_i - E(y)_i] \). The covariance matrix between additive genetic values of related individuals \( i \) and \( i' \) is

\[
\mathbf{Cov}(u_i^o, u_{i'}^o) = \mathbf{Cov}(\Lambda^{-1}u_i, u_{i'}^o\Lambda^{-1}) = a_{ii'}\Lambda^{-1}\mathbf{G}_0\Lambda^{-1},
\]

where \( a_{ii'} \) is twice the coefficient of coancestry between \( i \) and \( i' \).

**LIKELIHOOD FUNCTION**

Consider system (23) in conjunction with the normality assumptions (25) and (26), and regard the vector \( \mathbf{A} \) as “data.” The model for the entire data vector can be written as

\[
\begin{bmatrix}
\mathbf{A}_{y1} \\
\mathbf{A}_{y2} \\
\vdots \\
\mathbf{A}_{yN}
\end{bmatrix} = \begin{bmatrix}
\mathbf{X}_1 \\
\mathbf{X}_2 \\
\vdots \\
\mathbf{X}_N
\end{bmatrix} \beta + \mathbf{Z}_u + \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_N
\]

\[
= \mathbf{X}\beta + \mathbf{u} + \mathbf{e}, \quad (35)
\]

where \( \mathbf{u} \) comprises additive genetic effects for all individuals and all traits (\( \mathbf{u} \) may include additive genetic effects of individuals without records), and \( \mathbf{Z} \) is an incidence matrix of appropriate order. If all individuals have records for all traits, \( \mathbf{Z} \) is an identity matrix of order \( NK \times NK \) otherwise, columns of 0’s for effects of individuals without phenotypic measurements would be included in \( \mathbf{Z} \). In view of the normality assumptions (25) and (26), one can write

\[
\mathbf{u}|\mathbf{G}_0 \sim N(0, \mathbf{A} \otimes \mathbf{G}_0)
\]

and

\[
\mathbf{e}|\mathbf{R}_0 \sim N(0, \mathbf{I} \otimes \mathbf{R}_0),
\]

where \( \mathbf{A} \) is a matrix of additive genetic relationships (or of twice the coefficients of coancestry) between individuals in a genealogy, and \( \otimes \) indicates Kronecker product. Note that \( \mathbf{I} \otimes \mathbf{R}_0 \) reflects the assumption that all individuals with records possess phenotypic values for each of the \( K \) traits. This is not a requirement, but it simplifies somewhat the treatment that follows.

Given \( \mathbf{u} \), the vectors \( \mathbf{A} \) are mutually independent (since all \( \mathbf{e} \) vectors are independent of each other), so the joint density of all \( \mathbf{A} \) is

\[
p(\mathbf{A}_{y1}, \mathbf{A}_{y2}, \ldots, \mathbf{A}_{yN}|\mathbf{A}, \beta, \mathbf{u}, \mathbf{R}_0)
\]

\[
\times \frac{1}{|\mathbf{R}_0|^{1/2}} \exp \left[ -\frac{1}{2} \sum_{i=1}^N \mathbf{(A}_i - \mathbf{X}\beta - \mathbf{Z}_u)' \mathbf{R}_i^{-1}(\mathbf{A}_i - \mathbf{X}\beta - \mathbf{Z}_u) \right],
\]

where \( \mathbf{Z} \), is an incidence matrix that “picks up” the \( K \) breeding values of individual \( i \) and relates these to its phenotypic records \( y \). Making a change of variables from \( \mathbf{A}_i \) to \( \mathbf{y}_i \) (\( i = 1, 2, \ldots, N \)), the determinant of the Jacobian of the transformation is \( |\mathbf{A}| \). Hence, the density of \( \mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]' \) is

\[
p(\mathbf{y}|\mathbf{A}, \beta, \mathbf{u}, \mathbf{R}_0) \times \frac{1}{|\mathbf{A}^-1\mathbf{R}_0 \mathbf{A}^-1|^1} \times \exp \left[ -\frac{1}{2} \sum_{i=1}^N \mathbf{(y}_i - \mathbf{A}^-1(\mathbf{X}\beta + \mathbf{Z}_u))' \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{A}^-1(\mathbf{X}\beta + \mathbf{Z}_u)) \right]
\]

This is the density of the product of the \( N \) normal distributions

\[
\mathbf{y}|\mathbf{A}, \beta, \mathbf{u}, \mathbf{R}_0 \sim N(\mathbf{A}^{-1}(\mathbf{X}\beta + \mathbf{Z}_u), \mathbf{A}^{-1}\mathbf{R}_0 \mathbf{A}^{-1}),
\]

highlighting that the data generation process can be represented in terms of the reduced model (24), with the only novelty here being the presence of the inci-
dence matrix \( Z_\alpha \), with the latter being a \( K \times K \) identity matrix in (24). Hence, the entire data vector can be modeled as

\[
[y_1, y_2, \ldots, y_N] = \begin{bmatrix} \Lambda^{-1}X_\alpha \end{bmatrix} \beta + \begin{bmatrix} \Lambda^{-1}z_1 \cr \Lambda^{-1}z_2 \cr \vdots \cr \Lambda^{-1}z_N \end{bmatrix} u_i + \begin{bmatrix} e_1^T \cr e_2^T \cr \vdots \cr e_N^T \end{bmatrix}
\]

\[
= X_\alpha \beta + Z_\alpha u + e^* ,
\]

where \( X_\alpha \) is an \( NK \times \Sigma_{\beta=1}^p \beta \) matrix (again, assuming that each of the \( N \) individuals has measurements for the \( K \) traits), and \( Z_\alpha \) has order \( NK \times (N + P)K \), where \( P \) is the number of individuals in the genealogy lacking phenotypic records (the corresponding columns of \( Z_\alpha \) being null). Observe that (38) is in the form of a standard multiple-trait mixed-effects linear model, save for the fact that the incidence matrices depend on the unknown structural coefficients contained in \( \Lambda \). Hence

\[
p(y|\Lambda, \beta, u, R_\alpha) \\
\propto \frac{1}{|R_\alpha|^{1/2}} \exp \left[ -\frac{1}{2} (y - X_\alpha \beta - Z_\alpha u) R_\alpha^{-1} (y - X_\alpha \beta - Z_\alpha u)^T \right],
\]

where

\[
\text{Var}(e^*) = R_\alpha = I_N \otimes \Lambda^{-1} R_\alpha \Lambda^{-1}
\]

is a block-diagonal matrix consisting of \( N \) blocks of order \( K \times K \) and all such blocks are equal to \( \Lambda^{-1} R_\alpha \Lambda^{-1} \). It follows that \( y_i|\Lambda, \beta, u, R_\alpha \sim N(X_\alpha \beta + Z_\alpha u, R_\alpha) \). Hence, if simultaneity or recursiveness holds, the estimator of the residual variance-covariance matrix from a reduced model analysis is actually estimating \( \Lambda^{-1} R_\alpha \Lambda^{-1} \); this has a bearing on the interpretation of the parameter estimates.

Since it is assumed that \( u_i|G_\alpha \sim N(0, A \otimes G_\alpha) \), the likelihood function is given by

\[
L(\Lambda, \beta, R_\alpha, G_\alpha) \propto \left[ N(X_\alpha \beta + Z_\alpha u, R_\alpha) N(0, A \otimes G_\alpha) \right] du
\]

\[
\propto N(X_\alpha \beta + Z_\alpha (A \otimes G_\alpha) Z_\alpha^T).
\]

(40)

This likelihood has the same form as that for a standard multivariate mixed-effects model, except that, here, additional parameters (the nonnull elements of \( \Lambda \)) appear in both the location and dispersion structures of the reduced model (38). A pertinent issue, then, is whether or not all parameters in the model, that is, \( \Lambda, \beta, R_\alpha, \) and \( G_\alpha \), can be identified (i.e., estimated uniquely) from the likelihood. This is discussed in the following section.

IDENTIFICATION OF PARAMETERS

This is dealt with only briefly here as extensive treatments can be found in econometrics treatises such as Johnston (1972) and Judge et al. (1988); a readable account is in Goldberger (1998).

Consider the system of \( K \) response variables (23), and reorganize it as

\[
Ay_i + X\beta = e_i
\]

where \( \beta = -\beta \) and \( e_i = u_i + e_i \) is a residual. It is convenient to lump the sum of the two random effects into a single residual for the treatment that follows. Rewrite

\[
\begin{bmatrix} x_{1i} \cr 0 \cr \vdots \cr 0 \cr 0 \cr \tilde{\beta}_1 \cr \vdots \cr \tilde{\beta}_K \end{bmatrix} = Bx_i
\]

where \( x_i \), now a column vector of order \( \Sigma_{\beta=1}^p \beta \times 1 \), and \( B = K \times \Sigma_{\beta=1}^p \beta \). In practice, it suffices to keep the distinct explanatory variables in \( x_i \), e.g., if herd effects affect all traits in the system, only a single set of incidence variables needs to be considered. With this notation, (41) can be put as

\[
\begin{bmatrix} \lambda_1 y_i + b_1^T x_i \\ \lambda_2 y_i + b_2^T x_i \\ \vdots \\ \lambda_{K-1} y_{i-1} + b_{K-1}^T x_{i-1} \\ \lambda_K y_i + b_K^T x_i \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{K-1} \\ e_K \end{bmatrix},
\]

\[
\lambda_1 y_i + b_1^T x_i = e_i, \quad j = 1, 2, \ldots, K, \quad i = 1, 2, \ldots, N,
\]

constitutes the \( j \)th equation of the system. Compactly, the system is

\[
Ay_i + Bx_i = e_i.
\]

(43)

The reduced model is expressible as

\[
y_i = -\Lambda^{-1} Bx_i + \Lambda^{-1} e_i
\]

\[
= Bx_i + \tilde{e}_i,
\]

where \( B = -\Lambda^{-1} B \) is a \( K \times \Sigma_{\beta=1}^p \beta \) matrix of reduced model parameters, and \( \tilde{e}_i = \Lambda^{-1} e_i \). The system in (43) contains, at least potentially, the following number of parameters: \( K^2 \) (all elements of \( \Lambda \), including the 1’s in the diagonal), \( K \sum_{\beta=1}^p \beta \) (all elements of \( B \), including the null ones), plus \( K(K + 1) \) (the distinct elements of \( R_\alpha \).
and \( G_0 \). It is assumed that these two variance-covariance matrices can be separated in the estimation procedure, which depends on the genetic structure of the data set. Letting \( p = \Sigma_{p,p} / K \), the total number of parameters in the system is \( S = K^2(2 + p) + K \). In the reduced model, on the other hand, the number of potential parameters is \( K^2 p \) (the order of \( \Pi \)), plus \( K(K + 1) \) (the elements of \( \Sigma_{g,g} = \Lambda^{-1} G_0 A^{-1} \) and those of \( \Sigma_{b,b} = \Lambda^{-1} R_0 A^{-1} \)), yielding \( R = K^2(1 + p) + K \) as the total number of parameters. To obtain unique estimates of the parameters in \( \Lambda, B, G_0 \), and \( R_0 \), \( S - R = K^2 \) restrictions are needed for uniqueness. These can be of four types (Judge et al., 1988), as follows.

1. "Normalization" restrictions: set the diagonal elements of \( \Lambda \) to 1, so that the parameters in equation \( j \) are expressed relative to this constant of proportionality. This yields \( K \) restrictions, so an additional \( K^2 - K = K(K - 1) \) restrictions are still needed.

2. Exclusion restrictions: some of the \( \lambda \) coefficients may be 0, as in a recursive model, or the elements of \( B \) may not appear in each of the equations.

3. Restrictions in the form of a linear combination of parameters in the same equation or across equations.

4. Restrictions on the variance-covariance matrices \( G_0 \) and \( R_0 \) (typically, such restrictions are not employed in quantitative genetic analysis).

Formal procedures for evaluation of identification of equations are described by Johnston (1972) and Judge et al. (1988). Suppose that \( \Pi \) is given and that one wishes to estimate uniquely (identify) the parameters in \( \Lambda \) and in \( B \). Briefly, note that the parameters of the reduced model, \( \Pi = -\Lambda^{-1} B \), satisfy \( \Pi \Lambda + B = 0 \) or, equivalently,

\[
[\Lambda \ B] \Pi = 0. \quad (45)
\]

Consider now row \( j \) of (45) and write it as

\[
\lambda_j' \Pi + b_j' = 0.
\]

Transposing, this yields

\[
[\Pi' \ I_{j,j} \Sigma_{g-g} \Sigma_{b-b}] \begin{bmatrix} \lambda_j \\ b_j \end{bmatrix} = 0. \quad (46)
\]

This defines a system of equations on \( K + \Sigma_{g-g} \) unknowns in which the rank of the known coefficient matrix is \( \Sigma_{g-g} \). Hence, restrictions are needed to identify the unknown parameters \( \lambda \) and \( b \) of equation \( j \) of the system. The restrictions can be denoted (Judge et al., 1988) as

\[
R \begin{bmatrix} \lambda_j \\ b_j \end{bmatrix} = 0, \quad (47)
\]

where \( R \) is a \( J \times (K + \Sigma_{g-g}) \) matrix of rank \( J < K + \Sigma_{g-g} \). For example, an exclusion restriction can be indicated by filling the appropriate row of \( R \) with 0's, save for a 1 in the position corresponding to the element of \( [\lambda_j', b_j']' \) to be excluded from equation \( j \). Since the normalization restrictions set all diagonal elements of \( \Lambda \) as equal to 1, and then \( \lambda_j = 1 \) (\( \lambda_j \) is the \( j \)th element of \( \lambda \)), this implies that (47) must provide \( K - 1 \) linearly independent relationships, so that one can arrive at the \( K \) restrictions needed. Now, combine (46) and (47), to arrive at the system

\[
[\Pi' \ I_{j,j} \Sigma_{g-g} \Sigma_{b-b}] \begin{bmatrix} \lambda_j \\ b_j \end{bmatrix} = 0, \quad (48)
\]

where \( R = [R^A \ R^b] \) is given in partitioned form, and the coefficient matrix must have rank \( K - 1 + \Sigma_{g-g} \) to obtain unique estimates of \( \lambda \) and \( b \). Johnston (1972) and Judge et al. (1988) state that the rank of the coefficient matrix is \( K - 1 + \Sigma_{g-g} \) if and only if

\[
[R^A \ R^b] \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \ldots \\ \lambda_K \\ b_1 \\ b_2 \\ \ldots \\ b_K \end{bmatrix} = 0 \quad (49)
\]

is \( K - 1 \). Note that the preceding matrix has order \( J \times K \) and that column \( j \) is null by virtue of (47). Hence, for (49) to possess rank \( K - 1 \), it must be that \( J \geq K - 1 \); i.e., a condition for identification of equation \( j \) is that the number of restrictions \( J \) must be at least \( K - 1 \) (recall that \( K \) is the number of traits in the system). However, this is not sufficient: as stated, the rank of (49) must be \( K - 1 \).

In short, if the rank of (49) is \( K - 1 \), equation \( j \) is just identified, meaning that the relationship between the reduced model parameters and the \( \lambda \)'s and \( b \)'s in the equation is unique. If the rank is larger than \( K - 1 \), the equation is overidentified, meaning that there are many ways in which the structural model parameters can be expressed as a function of the elements of \( \Pi \). In these two cases, the \( \lambda \)'s and \( b \)'s may be inferred efficiently, using methods that employ all information available in the data, e.g., maximum-likelihood or Bayesian procedures. Finally, if the rank of (49) is smaller than \( K - 1 \), equation \( j \) is underidentified, and the structural parameters cannot be solved as a function of the reduced model parameters (Dreze and Richard 1983).

The preceding developments are illustrated with a two-trait simultaneous model. Suppose that \( y_{i1} \) and \( y_{i2} \) are measurements of systolic and diastolic blood pressure, respectively, taken on individual \( i \); assume that physiological knowledge postulates a feedback between the two variables. Let the models be

\[
y_{i1} = \lambda_{i1} y_{i2} + \beta_{i1} Age_i + \beta_{i2} Smokes_i + u_{i1} + \epsilon_{i1} \quad \text{and} \quad (49)
y_{i2} = \lambda_{i2} y_{i1} + \beta_{i2} Age_i + \beta_{i3} Drinking_i + \beta_{i4} Exer_i + u_{i2} + \epsilon_{i2},
\]

where \( Age \) is the age of \( i \) in years; \( Smokes \) is a binary variable (0 represents no smoking during the year prior to measurement and 1 represents smoking); \( Drinking \) is an estimate of the amount of alcohol \( i \) consumed in
the year previous to the blood pressure test, ignoring a possible error of measurement, and Exer measures the extent to which \( i \) exercises. The \( u \) and \( e \) variables are additive genetic and residual effects, as before, and the \( \lambda \)'s and \( \beta \)'s are the structural model parameters. Here, \( K = 2 \) and the number of \( x \) variables is 5, since the two intercepts \( \beta_{11} \) and \( \beta_{21} \) are related to the measurements via the same incidence variate, which takes the value 1 for all \( i \). The first equation has three “beta coefficients” (\( \beta_{11}, \beta_{12}, \text{and } \beta_{13} \)) and the second has four (\( \beta_{21}, \beta_{22}, \beta_{24}, \text{and } \beta_{25} \)). Before normalization

\[
\mathbf{A}_{2 \times 2} = \begin{bmatrix}
\lambda_{11} & -\lambda_{12} \\
-\lambda_{21} & \lambda_{22}
\end{bmatrix},
\]

and

\[
\mathbf{B}_{2 \times 3} = \begin{bmatrix}
-\beta_{11} & -\beta_{12} & -\beta_{13} \\
-\beta_{21} & -\beta_{22} & -\beta_{24} & -\beta_{25}
\end{bmatrix} \begin{bmatrix}
1 \\
\text{Age}_{i} \\
\text{Smoking}_{i} \\
\text{Drinking}_{i} \\
\text{Exer}_{i}
\end{bmatrix}.
\]

Equation 1 of the system uses the two exclusions \( \beta_{14} = \beta_{15} = 0 \). Hence, (49) is

\[
\mathbf{R} \begin{bmatrix}
\mathbf{A}' \\
\mathbf{B}'
\end{bmatrix} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} = \begin{bmatrix}
0 \\
0
\end{bmatrix}.
\]

The rank of this matrix is 1 (which is \( K - 1 \)), so that the equation is identified. Equation 2 of the system employs the exclusion \( \beta_{23} = 0 \) so that (49) is

\[
\begin{bmatrix}
\lambda_{11} & -\lambda_{12} \\
-\lambda_{12} & \lambda_{22} \\
-\beta_{11} & -\beta_{21} \\
-\beta_{21} & -\beta_{22} \\
\beta_{13} & 0 \\
0 & -\beta_{24} \\
0 & -\beta_{25}
\end{bmatrix} = \begin{bmatrix}
-\beta_{13} & 0
\end{bmatrix}.
\]

Since the rank of this matrix is 1, the second equation is identified as well. Hence, \( \lambda_{12}, \lambda_{22}, \text{and the elements of } \mathbf{B}_1 = [\beta_{11}, \beta_{12}, \beta_{13}]' \text{ and of } \mathbf{B}_2 = [\beta_{21}, \beta_{22}, \beta_{24}, \beta_{25}] \text{ can be estimated uniquely.}

### BAYESIAN MODEL

**General:** The form of the likelihood function given in (40) suggests that obtaining maximum-likelihood estimates of the structural model parameters \( \mathbf{A}, \mathbf{B}, \mathbf{R}_n, \) and \( \mathbf{G}_n \) is not an easy matter, with a main difficulty being the fact that \( \mathbf{A} \) is unknown. On the other hand, if the elements of this matrix were given, the setting would be as in a multivariate mixed-effects linear model, so standard procedures, such as the expectation-maximization (EM) algorithm, could be employed for computing the likelihood-based estimates. Another complication is that, typically, highly nonlinear functions of the parameters must be inferred. For example, see the forms of the mean \( \mu_k^* \) in model (7) and of the coefficient of heritability in (8). Intuitively, asymptotic approximations to the sampling distribution of the maximum-likelihood estimates may be relatively less accurate at a given sample size when the parametric function of interest is nonlinear than when it is linear. Note, however, that \( \mu_k^* \) and (8) may be inferred from the reduced model, via the standard multivariate parameterization. In special circumstances, one can form estimators of the structural parameters from statistics derived from the reduced model. These are called “indirect” procedures in econometrics (Johnston 1972).

Also, inferring random effects is of great importance in applied quantitative genetics (e.g., animal, plant, or tree breeding), and their best predictor would take a form such as in (32). In practice, however, calculations require replacing the unknown structural parameters by their maximum-likelihood estimates, that is, computing

\[
\hat{E}(u^*|y) = \hat{\Lambda}^{-1}\hat{G}_n(\hat{\mathbf{R}}_n + \hat{\mathbf{G}}_n)^{-1}\hat{\Lambda}(y - \hat{\Lambda}^{-1}\mathbf{X}\hat{\beta}).
\]

If interest focuses on the “system” genetic effects, the statistic would be

\[
\hat{E}(u^*|y) = \hat{\mathbf{G}}_n(\hat{\mathbf{R}}_n + \hat{\mathbf{G}}_n)^{-1}\hat{\Lambda}(y - \hat{\Lambda}^{-1}\mathbf{X}\hat{\beta}).
\]

The finite sample properties of the resulting empirical predictors are unknown. A common Bayesian criticism (Box and Tiao 1973; Gianola and Fernando 1986; Sorensen and Gianola 2002) is that (50) does not take the uncertainty (error of estimation) of the estimates of the parameters into account.

An alternative is to adopt a Bayesian approach, where inferences about structural parameters, random effects, or functions thereof are made from their marginal posterior distributions (Zellner 1971, 1979; Box and Tiao 1973; Gelman et al. 1995; Carlin and Louis 2000; Sorensen and Gianola 2002). A review of some of the issues in simultaneous models from an econometric perspective is in Zellner (1979), Dreze and Richard (1983), Judge et al. (1985), and Koop (2003). A salient feature of the Bayesian analysis is its ability to produce exact finite sample inference, as well as to override potential underidentification of parameters. If proper priors are adopted for all parameters in a model, all posterior distributions are proper as well (Bernardo and Smith 1994; O’Hagan 1994). However, unless the parameters are identifiable in the likelihood, the influence of the prior does not dissipate asymptotically. An example of this is in Carlin and Louis (2000) and in Gianola’s Quantitative Genetics of Simultaneity and Recursiveness.
nola and Sorensen (2002). These authors discuss a situation where two random variables have the distributions $X_i \sim N(\mu, \sigma^2)$ and $Y_i \sim N(\eta, \sigma^2)$, say, but $Z_i = X_i + Y_i$ ($i = 1, 2, \ldots, N$) is observed. Here, maximum likelihood can estimate $\mu + \eta$, but not $\mu$ or $\eta$ separately. The individual means cannot be inferred separately because an infinite number of combinations of the elementary parameters confer the same likelihood to $\mu + \eta$. On the other hand, a Bayesian analysis with proper priors assigned to both $\mu$ and $\eta$ gives distinct, proper, marginal posterior distributions of $\mu$ and $\eta$. However, the usual asymptotic domination of the prior by the likelihood does not occur. Even when sample sizes are of infinite size, the prior matters; see Dreze and Richard (1983) and O'Hagan (1994). For this reason, it is always important to investigate parameter identification, as discussed in the preceding section. A Bayesian analysis of the structural model (35) is presented subsequently.

**Prior and posterior distributions:** Represent all unknown parameters of the model by $\theta = [\lambda, \beta, u, R_0, G_0]$, where $\lambda$ is a vector containing the unknown elements of $\Lambda$, after normalization; for example, in a fully simultaneous model with $K = 3$, $\lambda$ would contain six $\lambda$ coefficients. No distinction is made here between fixed and random elements in the parameter vector since all unknowns are treated as unobservable random variables in the Bayesian setting. Bayes’ theorem gives as density of the posterior distribution

$$p(\theta | y, H) \propto p(y | \theta) p(\theta | H),$$

where $p(\theta | H)$ is the prior density of $\theta$, $H$ represents the collection of all known hyperparameters, and $p(y | \theta)$ is the density of the data. Note that $p(y | \theta) = p(y | \Lambda, \beta, u, R_0)$, since, given $u$, the data-generation model does not depend on the genetic covariance matrix $G_0$. Take as joint prior density of all parameters

$$p(\theta | H) = p(\lambda | H_0) p(\beta | H_0) p(u | H_0) p(R_0 | H_0) p(G_0 | H_0),$$

(51)

where, for example, $p(\lambda | H_0)$ is the density of the prior distribution of $\lambda$ and $H_0$ denotes a set of known hyperparameters. Take as prior distribution of $\lambda$ the Gaussian process

$$\lambda | \lambda_0, \tau^2 \sim N(1\lambda_0, \tau^2),$$

(52)

where $1$ is a vector of ones, $\lambda_0$ is a known prior mean, common to all $\lambda$’s, and $\tau^2$ is a tuning parameter, adjusting the degree of sharpness of the prior. Further, we assign the priors

$$\beta | \beta_0, \beta^2 \sim N(1\beta_0, I\beta^2),$$

(53)

$$u | G_0 \sim N(0, G_0),$$

(54)

$$R_0 | \nu_R, V_R \sim IW(\nu_R, V_R),$$

(55)

and

$$G_0 | \nu_G, V_G \sim IW(\nu_G, V_G),$$

(56)

where $IW(\nu_R, V_R)$ and $IW(\nu_G, V_G)$ denote $K$-dimensional Wishart distributions with “degrees of freedom” parameters $\nu_R$ and $\nu_G$, respectively, and scale matrices $V_R$ and $V_G$. Collecting (52–56), the joint prior density of all parameters in (51) is

$$p(\lambda, \beta, u, R_0, G_0 | \lambda_0, \tau^2, \beta_0, \beta^2, \nu_R, V_R, \nu_G, V_G) = N(\lambda | \lambda_0, \tau^2) N(\beta | \beta_0, \tau^2) N(u | 0, G_0) IW(R_0 | \nu_R, V_R) IW(G_0 | \nu_G, V_G).$$

(57)

Combination of (57) with the data density (37) gives as joint posterior density of all estimands

$$p(\theta | y, H) = \prod_{i=1}^N N(y_i | X_i \beta + Z_{ii}, \Lambda^{-1} R_0 \Lambda^{-1}) \times N(\lambda | \lambda_0, \tau^2) N(\beta | \beta_0, \tau^2) N(u | 0, G_0) IW(R_0 | \nu_R, V_R) IW(G_0 | \nu_G, V_G).$$

(58)

**FULLY CONDITIONAL POSTERIOR DISTRIBUTIONS**

Often, the fully conditional posterior distributions of the parameters of a model can be ascertained from the joint posterior density, (58) in this case, by retaining the parts varying with the parameter (or group of parameters) of interest and treating the remaining parts as known (e.g., Sorensen and Gianola 2002). This procedure is employed in what follows.

**Distributions of $\beta$ and $u$:** Using the preceding concept in (58), one has that

$$p(\beta, u | R_0, G_0, \lambda, y, H) = \prod_{i=1}^N N(y_i | X_i \beta + Z_{ii}, \Lambda^{-1} R_0 \Lambda^{-1}) N(\beta | 1\beta_0, I\beta^2) N(u | 0, G_0).$$

Now, since the value of $\lambda$ is given in this conditional distribution, one can treat this vector as known and form the pseudo-data vectors $A_{iy}$ ($i = 1, 2, \ldots, N$). Thus

$$p(\beta, u | R_0, G_0, \lambda, y, H) \propto p(A_{y1}, A_{y2}, \ldots, A_{yN} | \beta, u, R_0) N(\beta | 1\beta_0, I\beta^2) N(u | 0, G_0),$$

where $p(A_{y1}, A_{y2}, \ldots, A_{yN} | \beta, u, R_0)$ is as in (36). Retaining only the terms that depend on $\beta$ and $u$, the explicit form of the conditional posterior density above is

$$p(\beta, u | R_0, G_0, \lambda, y, H) \propto \exp \left\{ \sum_{i=1}^N (\lambda - \lambda_0) R_0^{-1} (\beta - \beta_0) \right\} \times \frac{\exp \left\{ -\frac{1}{2\beta^2} (\beta - 1\beta_0)^T (\beta - 1\beta_0) \right\} \exp \left\{ \frac{1}{2\beta^2} (\beta - 1\beta_0)^T (\beta - 1\beta_0) \right\}}{2^{\nu_G}}.$$

(59)

Note in (59) that the influence of the prior distribution of $\beta$ on conditional (or marginal) inferences can be tempered by taking a large value of the spread parameter $\beta^2$. The preceding expression can be recognized as the posterior density of the location parameters in a Gaussian linear model with proper priors and known variance components (Lindley and Smith 1972; Gianola and Fernando 1986; Sorensen and Gianola 2002). It is well known that that the corresponding distribution is
where
\[
\begin{pmatrix}
\hat{\alpha} \\
\hat{\beta}
\end{pmatrix} - \begin{pmatrix}
C_{\hat{\alpha}\hat{\beta}} \\
C_{\hat{\beta}\hat{\beta}}
\end{pmatrix}^{-1} \begin{pmatrix}
X'(I \otimes R_i')y + b'^{-1}\hat{\beta} \\
Z'(I \otimes R_i')y
\end{pmatrix},
\]
and
\[
\begin{pmatrix}
X'(I \otimes R_i')X + b'^{-1}I \\
Z'(I \otimes R_i')Z + A' \otimes G_i'^{-1}
\end{pmatrix}^{-1},
\]

Samples from (60) can be obtained directly, or by the method of Garcia-Cortés and Sorensen (1996), or via Gibbs sampling, element-wise or block-wise (e.g., Wang et al. 1993, 1994; Sorensen and Gianola 2002).

Distributions of \(G_0\) and \(R_0\): Retaining in (58) the parts that vary with \(G_0\) gives

\[
p(G_0|\beta, u, R_0, \lambda, y, H) \propto N(u|0, G_0)IW(G_0|\nu_G, V_G),
\]

taking the explicit form

\[
p(G_0|\beta, u, R_0, \lambda, y, H) \propto |G^{1/2}|^{\nu_G + \nu_G/2} \exp\left(-\frac{1}{2} u'(A^{-1} \otimes G_i')u\right)
\]

\[
\times |G_i'|^{\nu_G/2}\exp\left(-\frac{1}{2} tr (G_i'G_i')\right)
\]

\[
\times |G_i'|^{\nu_G/2}\exp\left(-\frac{1}{2} tr (G_i'\sum_{i=1}^3 \alpha^i u_i + V_c)\right),
\]

where \(\alpha^i\) is the element in row \(i\) and column \(j\) of \(A^{-1}\). This reveals that the conditional posterior distribution of \(G_0\) is the inverse Wishart process

\[
G_0|\beta, u, R_0, \lambda, y, H \sim IW\left(G_0|\nu_G + \sum_{i=1}^3 \alpha^i u_i + V_c\right),
\]

where \(\nu\) is the order of \(A\). Similarly, the density of the conditional posterior distribution of \(R_0\) is

\[
p(R_0|\beta, u, G_0, \lambda, y, H)
\]

\[
\propto \prod_{i=1}^N |Y_i|A^{-1}(X_i\beta + Z_iu)\Lambda^{-1}R_0A^{-1},
\]

Since the elements of \(A\) are given in this distribution, one can also write

\[
p(R_0|\beta, u, G_0, \lambda, y, H) \propto N(Y|A^{-1}X_i\beta + Z_iu, R_0|\nu_R, V_R)
\]

\[
= |R_0|^{-\nu_R/2}\exp\left(-\frac{1}{2} (Y_i - X_i\beta - Z_iu)'R_0^{-1}(Y_i - X_i\beta - Z_iu)\right)
\]

\[
\times |R_0|^{\nu_R/2}\exp\left(-\frac{1}{2} tr (R_0)V_R\right)
\]

\[
= |R_0|^{-\nu_R/2}\exp\left(-\frac{1}{2} tr (R_0\sum_{i=1}^3 \alpha^i u_i + V_c)\right),
\]

where \(r_i = Y_i - X_i\beta - Z_iu\). This is the density of the inverse Wishart distribution.

Distribution of the elements of \(\lambda\): All conditional posterior distributions considered so far are recognizable. However, this is not so for the elements of \(\lambda\). Retaining in (58) the parts that vary with the structural coefficients \(\Lambda\) yields

\[
p(\lambda|\beta, u, R_0, G_0, y, H)
\]

\[
\propto \prod_{i=1}^N |N(y_i\Lambda^{-1}(X_i\beta + Z_iu), \Lambda^{-1}R_0A^{-1})N(\lambda|I\lambda_0, I\tau^2)\].
\]

This has the form

\[
p(\lambda|\beta, u, R_0, G_0, y, H) \propto |\lambda|^{\nu_{\lambda}/2}\exp\left(-\frac{1}{2} \sum_{i=1}^N (\lambda - X_i\beta - Z_iu)'R_0^{-1}(\lambda - X_i\beta - Z_iu)\right)
\]

\[
\times \exp\left(-\frac{1}{2\tau^2}(\lambda - 1\lambda_0)'(\lambda - 1\lambda_0)\right).
\]

Hence, contrary to (60), (64), and (66), the conditional posterior distribution of \(\lambda\) is not standard, except in a fully recursive system (Zellner 1971), as noted below. Subsequently, a Metropolis scheme (Tanner 1993; Gelman et al. 1995) is developed for drawing samples from the distribution having (67) as density.

Consider \(\lambda_i\), and rewrite it (put \(K = 3\), to illustrate) as

\[
\lambda_i = \begin{bmatrix}
\lambda_{12} \\
\lambda_{13} \\
\lambda_{21} \\
\lambda_{23} \\
\lambda_{31} \\
\lambda_{32}
\end{bmatrix},
\]

\[
= \begin{bmatrix}
\lambda_{12} - \lambda_{13} \lambda_{23} \\
\lambda_{13} - \lambda_{23} \lambda_{32} \\
\lambda_{21} - \lambda_{23} \lambda_{31} \\
\lambda_{23} - \lambda_{21} \lambda_{32} \\
\lambda_{31} - \lambda_{12} \lambda_{23} \\
\lambda_{32} - \lambda_{12} \lambda_{31}
\end{bmatrix}
\]

\[
= Y_i - Y_i'\Lambda',
\]

where

\[
Y_i = \begin{bmatrix}
y_2 \\
y_3 \\
y_2 \\
y_3 \\
y_3 \\
y_3
\end{bmatrix},
\]

\[
\Lambda_i = \begin{bmatrix}
\lambda_{12} \\
\lambda_{13} \\
\lambda_{21} \\
\lambda_{23} \\
\lambda_{31} \\
\lambda_{32}
\end{bmatrix},
\]

Using this representation in (67) above gives

\[
p(\lambda|\beta, u, R_0, G_0, y, H) \propto |\lambda|^{\nu_{\lambda}/2}\exp\left(-\frac{1}{2} \sum_{i=1}^N (w_i - Y_i\lambda)'R_0^{-1}(w_i - Y_i\lambda)\right)
\]

\[
\times \exp\left(-\frac{1}{2\tau^2}(\lambda - 1\lambda_0)'(\lambda - 1\lambda_0)\right),
\]

where \(w_i = Y_i - X_i\beta - Z_iu\). The two quadratic forms on \(\lambda\) appearing in the exponents can be combined using standard results (e.g., Box and Tiao 1973; Sorensen and Gianola 2002), yielding

\[
p(\lambda|\beta, u, R_0, G_0, y, H) \propto |\lambda|^{\nu_{\lambda}/2}\exp\left(-\frac{1}{2} (\lambda - \hat{\lambda})'(\lambda - \hat{\lambda})\right).
\]

where
\[ \hat{\lambda} = \left( \sum_{i=1}^{N} Y_i R \cdot Y_i + \tau^{-2} I \right)^{-1} \left( \sum_{i=1}^{N} Y_i R \cdot \mathbf{w}_i + \tau^{-2} I \phi \right), \]

and

\[ V_\lambda = \left( \sum_{i=1}^{N} Y_i R \cdot Y_i + \tau^{-2} I \right)^{-1}. \]

An important simplification occurs in a fully recursive system, e.g., when \( y_i \) is a linear function of \( y_j \) and \( y_{j+1} = f(y_j, y_{j+1}); \) here, \( \Lambda \) is a triangular matrix, so that \( |\Lambda| = 1. \)

In this case, the conditional posterior distribution of \( \lambda \) is exactly the normal process \( N(\hat{\lambda}, V_\lambda) \), so sampling is straightforward.

For the general case, our Metropolis algorithm for drawing samples from (68) uses the multivariate normal distribution \( N(\hat{\lambda}, V_\lambda) \) for generating candidates. The algorithm proceeds as outlined below:

1. Draw a candidate \( \lambda^* \) from \( N(\hat{\lambda}, V_\lambda) \), with \( \hat{\lambda} \) and \( V_\lambda \) computed from the given \( \beta, u, R, G, y, \) and \( H \) values, which we refer to as state of the system at time \( t \).
2. Calculate the acceptance probability

\[ \alpha = \frac{|\lambda^*|^{N/2} \exp \left( -\frac{1}{2} \lambda^* - \lambda^* (V_\lambda)^{-1} \lambda^* - \lambda (V_\lambda)^{-1} \lambda \right)}{|\lambda^{N/2} \exp \left( -\frac{1}{2} \lambda^{N/2} \lambda (V_\lambda)^{-1} \lambda (V_\lambda)^{-1} \lambda \right)}. \]

Set

\[ \lambda = \begin{cases} \lambda^* & \text{with probability } \alpha, \\ \lambda^{(t-1)} & \text{otherwise.} \end{cases} \]

**MARKOV CHAIN MONTE CARLO AND INFERENCE FROM THE SAMPLES**

Drawing marginal inferences about unknown aspects of the joint posterior distribution (58) is impossible by analytical means, since this requires high-dimensional integration of unwieldy expressions. Instead, we propose to infer features of interest by iterative sampling via Markov chain Monte Carlo methods. The requisite theory is presented in Gilks et al. (1996), Robert and Casella (1999), and Sorensen and Gianola (2002).

Briefly, the idea is to create a Markov chain with (58) as equilibrium distribution. In our context, the sampling starts from some initial point \( \theta^{(0)} \) inside the parameter space, with updates obtained by successive looping through the fully conditional posterior distributions (60), (64), (66), and (68). This defines what has been termed a “Metropolis within Gibbs” algorithm: the draws from the known fully conditionals constitute the Gibbs proposals (accepted with probability equal to 1), with the Metropolis step for \( \lambda \) completing a loop of the algorithm (if the system is recursive, this becomes a Gibbs step as well). The early samples are discarded as burn-in, i.e., prior to declaration of convergence, and successive samples (\( m \), say) are collected subsequently until a sufficiently small Monte Carlo error of estimation of features of the posterior has been attained. Typically, the estimators of the features are ergodic averages. A discussion of convergence analysis is in Cowles and Carlin (1996).

To illustrate, consider inferring the offspring-parent regression in a simultaneous two-trait model, as discussed in Genetic Consequences of Simultaneity in a Two-Trait System. The Monte Carlo estimator of the regression is

\[ \hat{b}_{xy} = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{b_i} \frac{1}{b_i} \left( \sigma_{al}^2 + \lambda_{1l}^2 \right) \left( \sigma_{al}^2 + \lambda_{2l}^2 \right) + \lambda_{al}^2 \left( \sigma_{al}^2 + \alpha_{al}^2 \right), \]

where, for example, \( \sigma_{al}^2 \) and \( \lambda_{al}^2 \) are samples from the marginal posterior distributions of \( \sigma_{al}^2 \) and \( \lambda_{al}^2 \), respectively. A second example is that of inferring the additive genetic value of an individual for trait 1 in connection with model (7). The mean of the posterior distribution is estimated as

\[ \hat{E}(u_{i|}^a | y, H) = \frac{1}{m} \sum_{i=1}^{m} u_{i|}^a + \lambda_{al}^2 u_{i|}^a. \]

Further, in the context of multivariate selection, the estimate of the posterior mean of the merit of candidate \( i \) would be, using (32) and (33),

\[ \hat{T}_i = \frac{1}{m} \sum_{i=1}^{m} \Lambda_i (R_{i|} + G_{i|})^{-1} (\Lambda_i | y_i - X \beta_{i}). \]

**CONCLUSION**

This article proposes an extension of the standard linear models for analysis of quantitative traits in genetics to situations in which there is feedback or recursiveness between phenotypic values. In addition, techniques for parameter inference are described, with a focus on Bayesian analysis via Markov chain Monte Carlo methods. The representations develop a merger of standard quantitative genetics theory and modern tools from Bayesian inference with existing knowledge from econometrics and sociology, i.e., simultaneous systems and structural equations models (Joreskog and Sorbom 2001).

Wright (1925) pioneered on this type of treatment of multivariate systems, beginning with his “corn and hog correlations,” as noted by Goldberger (1972), and reaching a climax in Wright (1960), where a difficult path analysis of “reciprocal interaction with or without lag” is presented. Curiously, although Wright’s ideas were foundational in animal breeding and quantitative genetics, his work on feedback and on joint determination of systems of variables did not receive much attention (if at all) in these fields, even though biological systems typically display reciprocity between variables, with instantaneous or delayed feedbacks. Thus, it is perplexing that this type of work has been ignored in quantitative genetics. Path analysis (Wright 1921) was not favored by influential statistical geneticists such as
Charles Henderson (C. Henderson, personal communication) and Oscar Kempthorne (KEMPTHORNE 1969), who preferred variance component models and matrix treatments instead. For instance, calculation of additive relationships and of inbreeding coefficients is straightforward with tabular (matrix) methods (HENDERSON 1976; QUAAAS 1976), and the inverse of large relationship matrices is crucial in genetic evaluation via the mixed-model algorithm (see GIANOLA 2001, for a vignette). Similar calculations are awkward, at best, if done with the method of path coefficients, and it is not obvious how the needed inverse can be generated directly from the paths. Also, animal breeders often need to account for nuisance fixed effects (the explanatory variables are often discrete) and for interaction variance components in the models. This also contributed to the fading away of path analysis in animal breeding, since it was not clear how these problems could be treated in the standard framework of path analysis. Arguably, a heavy emphasis on large-scale computations may have distracted animal and plant breeders away from the modeling process, where a path diagram has an unparalleled eloquence.

On the other hand, path analytic methods have had an important impact in sociometrics. For example, see DUNCAN (1975) for an account of structural equation models. Further, the LISREL software (now in version 8; JORESKOG and SORBOM 2001) for estimation of structural coefficients has been used in sociology since the mid 1970s. A discussion of the use of structural equation models in biology is in SHIPLEY (2002), with emphasis on causal inference; an abundant literature on estimation and testing in structural models can be found in this text. In this context, our article can be viewed as an attempt to reclaim, for quantitative genetics, the heritage of Wright modeling ideas, although in the light of modern machinery developed mainly by econometricians.

As pointed out in matrix representations, there can be many competing models for explaining relationships in a multivariate system. Thus, model selection issues are paramount. The standard Bayesian tool kit for model selection and criticism, including the Bayes factor (BERNARDO and SMITH 1994; O’HAGAN 1994) and posterior predictive checks (GELMAN et al. 1995; SORENSEN and WAAGEPETERSEN 2003), can be employed in connection with simultaneity and recursiveness without extra difficulties, other than those related to the computations.

The methods presented here can be extended in several directions. For example, some of the response variables involved may be discrete. Here, one could introduce latent variables, such as the typical liability variate of quantitative genetics (THOMPSON 1979; GIANOLA 1982), and then model feedback or recursiveness at that level. Also, nonlinearity (in the parameters) may be dictated by mechanistic considerations. For example, a gene product may accumulate following the trajectory of some nonlinear growth curve, and one may wish to capture such behavior in the model. Lack of linearity makes identification and computation more difficult, but the Bayesian Markov chain Monte Carlo machinery works satisfactorily with nonlinear systems (GILKS et al. 1996). Another extension consists of a hierarchical modeling of the $\lambda$ coefficients controlling feedback or recursion. Here, we have assumed that the model depends on a single set of $\lambda$ parameters. However, these may be clustered, e.g., family aggregation of parameters, or be affected by a locus with major effects. This sort of multier specification can be handled via the usual Bayesian modeling of hierarchies, in the sense of LINDLEY and SMITH (1972).

We have focused on feedback or recursiveness at the phenotypic level. However, these effects might also take place at the level of the genes, in which case the model could be written as

$$\mathbf{A}_y = \mathbf{X}\mathbf{\beta} + \mathbf{A}_u u + \mathbf{e},$$

with a “purely genetic” (phenotypic) feedback obtained by setting $\mathbf{A}_y (\mathbf{A}_u)$ equal to an identity matrix. It would be of interest to study parameter identification issues in this more general model.

There are situations specific to animal breeding where introducing simultaneity in the model may alleviate some inadequacies of standard treatments. Consider, for example, dairy cows with first and second lactation milk yield records. It is known that if a cow has a good first record of performance, then there is a chance that she will receive preferential nutrition and management, relative to her herd mates. One can think of an effect of the first record on the second even in the absence of a mechanistic basis for this when the breeding value and permanent environmental effect of the cow are included in the statistical models for genetic evaluation. Similarly, one may consider including an effect of second on first records on similar grounds, as if the future had an influence on the present.

In short, there is a plethora of potential scenarios in biology where recursiveness or simultaneity enters into the realm of the possible. In spite of the towering influence of Sewall Wright, these phenomena have been essentially ignored in quantitative genetics. It is hoped that the techniques described in this article will contribute toward a deeper understanding of complex traits.

The authors thank James Crow, Arthur Goldberger, Guillerme Rosa, and Arnold Zellner for useful comments. Research was supported by the Wisconsin Agriculture Experiment Station and the Danish Institute of Agricultural Sciences and by grants from the National Research Initiatives Competitive Grants Program/U.S. Department of Agriculture (99-35205-8162 and 2003-35205-12833) and the National Science Foundation (DEB-0089742).

LITERATURE CITED

