STRUCTURAL EQUATION MODELS IN HUMAN BEHAVIOR GENETICS

Arthur S. Goldberger

October 2002

This is the revised version of a paper presented at the NSF Symposium on Identification and Inference for Econometric Models, held at Berkeley in August 2001. It is scheduled to appear in a festschrift honoring Thomas Rothenberg. The research reported here derives from joint work with Leon J. Kamin. I am grateful also to Donald Hester and Molly Martin for instructive advice.

STRUCTURAL EQUATION MODELS IN HUMAN BEHAVIOR GENETICS

Arthur S. Goldberger
University of Wisconsin

Introduction

That IQ is a highly heritable trait has been widely reported. Rather less well-known are recent reports in major scientific journals such as those announcing that the heritability of controllable life events is 53% among women and 14% among men (Saudino et al., 1997), while the heritabilities of inhibition of aggression, openness to experience, and right-wing authoritarianism are respectively 12%, 40%, and 50% (Pedersen et al., 1989; Bergeman et al., 1993; McCourt et al., 1999). It seems that milk and soda intake are in part heritable, but not the intake of fruit juice or diet soda (de Castro, 1993).

These reported heritabilities are parameter estimates obtained in structural modeling of measures taken on pairs of siblings -- prototypically, identical (monozygotic) twins and fraternal (dizygotic) twins, some reared together and others reared apart. The models are of the linear random effects type, in which an observed trait -- a phenotype -- is expressed in terms of latent factors -- genetic and environmental -- whose prespecified cross-twin correlations differ by zygosity and rearing status. Estimation is by maximum likelihood applied to the phenotypic variances and covariances. Heritability, the key parameter of interest, refers to the proportion of the variance of the phenotype that is attributable to the variance of the genetic factors.

Regarding these studies, various issues arise. Those that I will touch on here include: identification, nonnegativity constraints, alternative estimators, pretest estimation, conditioning of the design

matrix, multivariate analyses, and the objectives of structural modeling. Some of these issues were featured in Thomas Rothenberg's dissertation (1972), a remarkable book that led me to appreciate the generality of the minimum chi-square principle in estimation, and the contrast between equality and inequality constraints in efficient estimation.

In the present paper, I will focus on the SATSA project -- the Swedish Adoption/Twin Study of Aging -- which, from the early 1980s on, has assembled a sample of adult twin pairs: approximately 200 MZT (identical twins reared together), 200 DZT (fraternal twins reared together), 100 MZA (identical twins reared apart), and 150 DZA (fraternal twins reared apart). The fraternal twins are all same-sex pairs. The twins have been assessed in person and via mail questionnaires on several occasions, on a wide range of traits, some cognitive and others relating to personality, temperament, and recollections of childhood upbringing. Concerns about the representativeness of the samples and the reliability and validity of the measures were raised in Goldberger & Kamin (1998) and Kamin & Goldberger (2002). I suppress those concerns here in order to focus on the modeling.

Primary Model

The specification of the main SATSA model is captured as follows. Consider a typical individual, whose phenotype (observable trait value) Y is determined by unobservable factors as

(1) $$Y = \alpha_1 G + \alpha_2 D + \alpha_3 S + \alpha_0 U.$$

Here G is the additive genetic factor, D is the nonadditive genetic factor, S is the shared environment factor, and U is the nonshared environment factor. (The distinction between the two genetic factors will be exposited

later). Assume that the factors are uncorrelated, and standardize all variables to have zero means and unit variances, so that the phenotypic variance is

$$(2) \qquad V(Y) = \alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_0^2 = 1 \ .$$

The individual is paired with his or her sibling, whose phenotype is determined as

$$(3) \qquad Y' = \alpha_1 \, G' + \alpha_2 \, D' + \alpha_3 \, S' + \alpha_0 \, U' .$$

Across the sibling pair, all factor correlations are assumed to be zero except perhaps for those that link one sibling's additive genetic, nonadditive genetic, and shared environment factors with the corresponding factors of the other sibling. So the phenotypic sibling covariance is

$$(4) \qquad C(Y,Y') = C(G,G') \ \alpha_1^2 + C(D,D') \ \alpha_2^2 + C(S,S') \ \alpha_3^2 \ .$$

Referring to identical and fraternal twins (MZs and DZs), reared together and apart (Ts and As), those factor covariances are assumed to be

$$\begin{aligned} &C(G,G') = 1 \text{ for MZs, } 1/2 \text{ for DZs} \\ (5) \qquad &C(D,D') = 1 \text{ for MZs, } 1/4 \text{ for DZs} \\ &C(S,S') = 1 \text{ for Ts, } 0 \text{ for As} \ . \end{aligned}$$

With all variables standardized, covariances are also correlations. The consequence is that in the population, the phenotypic correlations for the four twin types are

$$\begin{aligned} &\text{MZT} \qquad \rho_1 = \alpha_1^2 + \alpha_2^2 + \alpha_3^2 \\ (6) \quad &\text{DZT} \qquad \rho_2 = \alpha_1^2/2 + \alpha_1^2/4 + \alpha_3^2 \\ &\text{MZA} \qquad \rho_3 = \alpha_1^2 + \alpha_2^2 \\ &\text{DZA} \qquad \rho_4 = \alpha_1^2/2 + \alpha_2^2/4 \ . \end{aligned}$$

Let $\beta_1 = \alpha_1^2$, $\beta_2 = \alpha_2^2$, and $\beta_3 = \alpha_3^2$, and define the vectors

$$(7)\quad \underset{\sim}{\rho} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \end{bmatrix}, \quad \underset{\sim}{x}_1 = \begin{bmatrix} 1 \\ 0.5 \\ 1 \\ 0.5 \end{bmatrix}, \quad \underset{\sim}{x}_2 = \begin{bmatrix} 1 \\ 0.25 \\ 1 \\ 0.25 \end{bmatrix}, \quad \underset{\sim}{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

Then SATSA's Primary Model has this linear specification for the population phenotypic correlations:

$$(8)\quad \underset{\sim}{\rho} = \underset{\sim}{x}_1 \beta_1 + \underset{\sim}{x}_2 \beta_2 + \underset{\sim}{x}_3 \beta_3.$$

These βs are components of phenotypic variance, as is the implied nonshared environment component, $\beta_0 = \alpha_0^2 = 1 - (\beta_1 + \beta_2 + \beta_3)$. The parameter β_1 is called narrow heritability, while the sum $\beta_1 + \beta_2$ is called broad heritability.

With 4 correlations expressed in terms of 3 parameters, there is 1 equality restriction, namely

$$\rho_1 - \rho_3 = \rho_2 - \rho_4,$$

which says that the difference between MZ and DZ correlations is the same whether the twins are reared together or apart. Further, with all 4 βs assumed to be nonnegative, there is an inequality restriction, namely

$$\rho_3/4 \leq \rho_4 \leq \rho_3/2,$$

which says that the DZA correlation should lie between one-fourth and one-half of the MZA correlation.

Given random samples from each of the 4 twin groups, one might take observed phenotypic correlations r_1, r_2, r_3, r_4, interpret them as estimates of the population correlations, and estimate $\underset{\sim}{\beta} = (\beta_1\ \beta_2\ \beta_3)'$ by running the least-squares linear regression of the 4 x 1 vector $\underset{\sim}{r} = (r_1\ r_2\ r_3\ r_4)'$ on the 4 x 3 matrix $\underset{\sim}{X} = (\underset{\sim}{x}_1\ \underset{\sim}{x}_2\ \underset{\sim}{x}_3)$, thus minimizing $\sum_{i=1}^{4} (r_i - \rho_i)^2$.

A more appropriate procedure would take account of the fact that the

variance of a sample correlation coefficient depends on the population correlation coefficient as well as the sample size, and choose values for the β–estimates to minimize

$$\sum_{i=1}^{4} w_i (r_i - \rho_i)^2,$$

where $w_i = n_i/(1 - r_i^2)^2$, with n_i being the number of observations in the i–th twin group.

However, it is most convenient to work with Fisher's z–transforms of correlation coefficients, namely

$$z = (1/2) \log[(1 + r)/(1 - r)],$$

$$\zeta = (1/2) \log[(1 + \rho)/(1 - \rho)],$$

relying on the presumption that in random sampling, sample size n, the variable z is distributed approximately $N(\zeta, 1/n)$; see Wilks (1962, p. 276). So a particular application of the minimum chi-square principle, which I label FZLS, chooses values for the β–estimates to minimize

$$\sum_{i=1}^{4} n_i (z_i - \zeta_i)^2,$$

which amounts to a straightforward, albeit nonlinear, regression problem. With 4 observations and 3 parameters, the minimized criterion provides an asymptotic $\chi^2(1)$ statistic of model fit which can serve to test the equality restriction $\rho_1 - \rho_3 = \rho_2 - \rho_4$.

I have oversimplified the procedure of the SATSA group in several respects. They do not standardize the observed variables, but rather work with variances and covariances, taking β_0 as a free parameter. (They then rescale parameter estimates ex post to obtain the proportional components of variance). They do not use FZLS, but rather Gaussian ML, following Neale & Cardon (1992, Chapters 6–7). Often, they take as data 8 phenotypic variances: for each twin group, the between–family and within–family

components. That gives them 3 additional degrees of freedom for model fit, which are implicitly allocated to the hypothesis that the four phenotypic variances are the same. This may be an interesting hypothesis, but has little to do with behavior–genetic theory. Sometimes they work with 12 observed phenotypic variances and covariances: for each twin group, a variance for twin A, a variance for twin B, and a covariance. This gives 4 additional degrees of freedom for model fit, which are implicitly allocated to equating the phenotypic variances for twins A and B in each twin group. The labelling of the twins was arbitrary, so those 4 additional degrees of freedom are in effect allocated to the hypothesis that SATSA's own assignment of the labels was in fact random. This is hardly an interesting hypothesis, and has nothing to do with behavior–genetic theory. (The economists Ashenfelter & Krueger (1994), working with twins, albeit not with behavior genetics, also treat an arbitrary labelling of Twin A and Twin B as meaningful).

Typically, the SATSA group residualize the observed traits on age and gender before beginning the modeling exercise, but occasionally they introduce age into the model itself as a covariate. This adds 2 parameters (a population age variance and a population trait–on–age slope), and adds 12 observed moments: for each twin group, the covariance of twin A's trait with age, the covariance of twin B's trait with age, and the variance of age. (As usual, twins in Sweden have the same age). In this manner, Lichtenstein et al. (1992) were able to report a total of 18 degrees of freedom for model fit, while the core of the model in correlation terms had just 1.

Genetic Theory

The genetic basis for this line of research is minimal. The biological content of the model, after all, consists of the ratios 1/2 and 1/4 for DZ twins relative to MZ twins. (It is true that the theory does extend to cover kinships other than twins). The formal distinction between the two genetic factors should be familiar to econometricians: it is the distinction between the conditional expectation function and the best linear predictor. Consider a gene with two variants (alleles), - and +. At this locus, an individual may be --, -+, +-, or ++. Score these as Z = 0, 1, 1, 2, and consider the distribution of phenotypes Y for persons of each score Z. If $E(Y|Z)$ is linear, that is if the expected observable trait for heterozygotes (Z = 1) is halfway between those for homozygotes (Z = 0 and Z = 2), then only an additive genetic factor is present. If $E(Y|Z)$ is nonlinear, for example if the expected observed trait for Z = 2 is the same as for Z = 0, then a nonadditive genetic factor is present. In that case, the $BLP(Y|Z)$ gives the additive factor, and the deviations $E(Y|Z) - BLP(Y|Z)$ give the nonadditive factor. So the two genetic factors are uncorrelated by construction. The Appendix sketches why, under certain assumptions, MZs and DZs correlate 1/2 and 1/4 on those two factors. Remarkably, the argument for a single locus extends directly to multiple loci. On all this, see Falconer & Mackay (1996, Chapters 7–9).

In SATSA's Primary Model, identification is obtained by ruling out many possibilities a priori. Covariance between an individual's genetic factors and shared environment factor is not allowed, conventional wisdom on the role of parents to the contrary notwithstanding. Nor is there any allowance for the possibility that the separated twins were placed into

similar environments. Nor is there any allowance for MZTs to have more similar environments than DZTs, that is, for C(S,S') to differ by zygosity; any excess phenotypic similarity of MZTs over DZTs is attributed to their excess genetic similarity. Joseph (1998) provides a critical assessment of the evidence in favor of this "equal environment assumption". Even the specified ratios 1/2 and 1/4 are not sacred; those values are valid under random mating, but would be different if there is assortative mating for the trait.

It is quite ironic that the assumptions of the behavior–genetic model refer so directly to social behavior, rather than biological processes.

Secondary Model

On occasion, the SATSA group adopts an alternative model that makes allowance for some environmental similarity for twins reared apart, thus addressing the objection that the separated twins may not have been reared in randomly different environments. This is accomplished by replacing the nonadditive genetic factor with a "selective placement" or "correlated environment" factor which correlates perfectly across twins of all types. In terms of the display in (7), replace $\underset{\sim}{x}_2$ with a new variable $\underset{\sim}{x}_4$ whose value is 1 for all 4 twin groups. Then $\underset{\sim}{x}_4 = (1, 1, 1, 1)'$ and the Secondary Model has

$$\underset{\sim}{\rho} = \underset{\sim}{x}_1 \gamma_1 + \underset{\sim}{x}_4 \gamma_2 + \underset{\sim}{x}_3 \gamma_3. \qquad (9)$$

As the SATSA group recognizes, it is not feasible to include both genetic factors along with the new environmental factor because exact collinearity would result: $\underset{\sim}{x}_4 = 3\,\underset{\sim}{x}_1 - 2\,\underset{\sim}{x}_2$.

The design columns in (9) span the same space as those in (8), so the Secondary Model implies the same equality constraint, namely $\rho_1 - \rho_3 = \rho_2 - \rho_4$. However, with all its γs assumed nonnegative, the implied inequality constraint is now

$$\rho_3/2 \leq \rho_4 \leq \rho_3 ,$$

which says that the DZA correlation should lie between 1/2 and 1 times the MZA correlation. So SATSA researchers are either attracted to this model immediately when the observed DZ correlations run high relative to the MZ correlations, or else choose it retroactively after observing that the nonnegativity constraint binds when fitting the Primary Model.

Agnostic Model

The need to choose between the Primary and Secondary Models may be avoided by freeing up the relation between ρ_4 and ρ_3. This can be accomplished by allowing two distinct genetic factors, one for MZs and one for DZs. If we let $\underset{\sim}{x}_5 = (1\ 0\ 1\ 0)'$ and $\underset{\sim}{x}_6 = (0\ 1\ 0\ 1)'$, we can write the Agnostic Model as

(10) $$\underset{\sim}{\rho} = \underset{\sim}{x}_5 \delta_1 + \underset{\sim}{x}_6 \delta_2 + \underset{\sim}{x}_3 \delta_3 .$$

Observe that $\underset{\sim}{x}_5 = -\underset{\sim}{x}_1 + 2\,\underset{\sim}{x}_2$ and $\underset{\sim}{x}_6 = 4\,\underset{\sim}{x}_1 - 4\,\underset{\sim}{x}_2$, so the columns of this design span the same space as the previous ones did, and this model implies the same equality constraint, $\rho_1 - \rho_3 = \rho_2 - \rho_4$. One may suppose that all 3 δs are nonnegative, and also $\delta_1 > \delta_2$ (genetic similarity greater for MZs than for DZs). But even with all δs assumed nonnegative, it allows ρ_4 to range from 0 up to ρ_3. (A similar idea was employed by Lykken et al. (1988)). Adopting the Agnostic Model would reduce the need to follow SATSA's model selection strategy, and would also dispel some of the

mystique of a rigorous biological foundation for the SATSA analyses.

Skeptical Model

A crucial feature of SATSA's models is that they make no allowance for environmental resemblance to differ between MZTs and DZTs, as a result say of more similar treatment by parents and peers. A simple alternative 3—parameter specification would include the additive genetic factor $\underset{\sim}{x}_1$, an MZT shared environment factor $\underset{\sim}{x}_7 = (1\ 0\ 0\ 0)'$, and a DZT shared environment factor $\underset{\sim}{x}_8 = (0\ 1\ 0\ 0)'$. This Skeptical Model can be written as

(11) $$\underset{\sim}{\rho} = \underset{\sim}{x}_1 \theta_1 + \underset{\sim}{x}_7 \theta_2 + \underset{\sim}{x}_8 \theta_3 \quad .$$

These design columns span a different space, and the single equality restriction is now

$$\rho_4 = \rho_3/2.$$

One may suppose that all 3 θs are nonnegative, and also $\theta_2 > \theta_3$ (environmental similarity greater for MZTs than DZTs). This model has, as far as I know, not been used by the SATSA researchers, and only rarely by other behavior geneticists, e.g. Loehlin (1987, pp. 122—126).

Identification and Constraints

In practice it is rare for SATSA to publish estimates of a full 3—factor version of either their Primary or Secondary Model. Almost invariably, one or another of the three factors will be dropped and a reduced model fitted and reported. That happens either when one of the estimated parameters is "nonsignificant", or when their algorithm (which apparently precludes negative estimates) finds the nonnegativity constraint to be binding and sets the offending parameter at zero. As a consequence of

this general–to–specific strategy, almost always the only model published is a reduced 2–factor, or 1–factor, one.

In particular, throughout the SATSA publications, one rarely -- perhaps 5% of the time -- finds traits for which both additive and nonadditive genetic variance components are estimated to be nonzero. It is not hard to see why. Consider the Primary Model. If only MZT and DZT data were available, it would be impossible to distinguish between the additive and nonadditive genetic components. The availability of separated twins formally identifies β_1, β_2, and β_3, but the identification is tenuous. Treating the 4 x 3 design matrix $\underset{\sim}{X} = (\underset{\sim}{x}_1\ \underset{\sim}{x}_2\ \underset{\sim}{x}_3)$ as if it simply had 4 observations, the "correlation" (about zero) between x_1 and x_2 is 0.97. That high degree of collinearity carries over to FZLS and ML estimation, producing unreliable and negatively correlated estimates of β_1 and β_2. Pedersen, Plomin, Nesselroade, & McClearn (1992) cast their lot with the additive side; Plomin, Pedersen, Lichtenstein, & McClearn (1994), analyzing the same cognitive traits, cast their lot with the nonadditive side. Similar considerations apply to the Secondary Model.

Had the Agnostic Model been used, in many cases the SATSA group could have maintained a full 3–factor model for the correlation coefficients. For example, the condition $\delta_1 > \delta_2 > 0$ is equivalent to $\beta_1 + \beta_2 > 0.5\ \beta_1 + 0.25\ \beta_2 > 0$, implying $0.5\ \beta_1 + 0.75\ \beta_2 > 0$, and even when unconstrained estimates of one of those βs was negative, the corresponding estimates of the δs might be admissible.

Empirical Implementation

To illustrate the applications, for three traits selected from SATSA

publications I give estimates of the various models. In Pedersen et al. (1988), Extraversion and Neuroticism were each measured as the sum of yes–no responses (coded 1–0) to nine items drawn from a short form of the Eysenck Personality Inventory, residualized on age and gender. In Lichtenstein et al. (1992), Occupation was measured by four nonfarm occupational categories (coded 1, 2, 3, 4, then logged), gender–specific and residualized on age; I use only the results for men.

Table 1 refers to those three traits. First, the observed correlations are given along with sample sizes. Then come model–fitting results, with Roman letters denoting estimates of the corresponding Greek–letter parameters, and standard errors where available in parentheses. SATSA's ML estimates for the particular reduced Primary Model that they published are given, followed by my FZLS estimates for that model. (Reassuringly, our numbers are generally close; the exception, b_3 for Neuroticism, I take to be a misprint. For Neuroticism, Pedersen et al. (1988) also report and prefer the full Secondary Model, with parameter estimates $b_1 = .13$, $b_4 = .16$, $b_3 = .07$).

Then follow results of my fitting the full Primary, Secondary, Agnostic, and Skeptical Models by FZLS. Readers may, for each model, readily calculate the fitted correlations from the parameter estimates. And they may also estimate reduced versions of these models: the FZLS method requires only the r_is and n_is, which are often what is available in SATSA publications.

Throughout the table, the chi–square statistic is the minimized value of the FZLS criterion. Degrees of freedom for model fit are the number of correlations, 4, minus the number of parameters estimated, 3 for full

models and 2 for the reduced models. For their reduced models, the χ^2s for model fit approach or exceed significance by conventional standards. As is to be expected, chi-square values coincide when design matrices span the same space.

A curiosity of the SATSA analyses, one that is not inherent in the behavior–genetic approach, is that they typically formulate the model in terms of path coefficients (such as our αs) rather than the variance components (such as our βs, the squared αs). As a result, they report ML standard errors for estimated path coefficients, which do not translate into standard errors for the parameters of theoretical interest, namely the contributions to variance. For our FZLS estimates, standard errors are routinely calculated.

In the table, we observe that the total genetic component, $\beta_1 + \beta_2$, is estimated virtually the same whether the full or reduced Primary Model is used, and is the same (apart from rounding) as the estimate of δ_1 in the Agnostic Model. So it might be argued that broad heritability of each trait is clearly discernible in the data. On the other hand, we also observe that the Agnostic and Skeptical Models appear as plausible competitors for SATSA's preferred models, while providing alternative interpretations of the data. For example, the Skeptical Model attributes only 22% of the variance in Extraversion to genetic factors, rather than 40% or so.

The FZLS method does not constrain the parameter estimates to be nonnegative, and indeed for many of the SATSA data sets, FZLS produces negative estimates where SATSA would reduce the model and effectively report zeros. One could test the nonnegativity constraints, a task never undertaken by the SATSA researchers. For example, in the Primary Model for

Extraversion, forcing $b_1 = 0$ increases χ^2 by 3.45 (= 5.84 − 2.29), approaching significance by conventional standards for a single constraint. To be sure, the appropriate test procedure is that for inequality constraints, which is more tolerant of departures: see Kodde & Palm (1986), Wolak (1987). Recently, some behavior geneticists have reported confidence intervals using the profile likelihood. The source article is Neale & Miller (1997), which recommends discarding any negative portion of the interval, that is, left–truncating the interval at zero.

If the SATSA group insist on the requirement that all βs be nonnegative, it is because of their insistence on interpreting them as components of variance. Perhaps the frequent occurrence of binding constraints should serve as an indication that their general behavior–genetic approach is not valid. On the other hand, there is nothing in principle that precludes factors that contribute to dissimilarity rather than similarity of twins. Perhaps negative parameter estimates should not serve to reject a particular full model out of hand.

Pretesting Issues

SATSA's empirical implementation of the behavior–genetic approach is not a routine exercise, but involves a sequence of choices and stopping rules. Nothing about the track that leads to their final variant is accounted for when they engage in statistical inference. So the standard errors and confidence intervals that they do report are merely nominal. The pretesting issues associated with such model selection are not mentioned in the behavior–genetic reports nor in the standard textbook, Neale & Cardon (1992). My impression from the econometric and statistical literature is

that under pretesting, nominal standard errors are misleadingly low, so that actual precision is overstated.

To investigate this, Monte Carlo runs may be useful. I report on one here. Adopt the Primary Model with parameter values $\beta_1 = 0.4$, $\beta_2 = 0.1$, $\beta_3 = 0.3$, implying $\rho_1 = 0.800$, $\rho_2 = 0.525$, $\rho_3 = 0.500$, $\rho_4 = 0.225$. Take the sample sizes to be $n_1 = 100$, $n_2 = 100$, $n_3 = 50$, $n_4 = 100$. Generate sample correlations r_i ($i = 1,\ldots,4$), or rather the z-transforms thereof z_i, by random sampling from $z_i \sim N(\zeta_i, 1/n_i)$. Estimate parameters by FZLS, reducing the model and re-estimating when a parameter estimate is negative.

Table 2 summarizes results of a 1000-replication run. Column (1) gives, for estimation of the full model, the average parameter estimates, their average standard errors, and their actual standard deviations. The next three columns give that information conditionally for the three branches of the pretest estimator: column (2) refers to the 71 samples in which the additive genetic factor was dropped because its unrestricted coefficient estimate was negative, column (3) refers to the 349 samples in which the nonadditive factor was dropped because its unrestricted coefficient estimate was negative, and column (4) refers to the 580 samples in which all factors were retained because none of the unrestricted estimates were negative. In the rightmost column, the information is given unconditionally for the pretest estimator, blanks in columns (2) and (3) being treated as zeroes. We observe some bias in the pretest estimators of β_1 and β_2, and more variability in them than would be indicated by the standard errors for the reduced models. On the other hand, we observe that the sum $\beta_1 + \beta_2$ is virtually unbiasedly estimated by $b_1^* + b_2^*$.

Multivariate Models

Having analyzed dozens of observed traits individually in the same manner, the SATSA group has moved on to multivariate analyses, in which several phenotypes are modeled jointly in terms of latent factors. So now the concern is with accounting for covariances, as well as variances, of observed traits. For example, Lichtenstein & Pedersen (1995) analyze five phenotypes jointly: life events, loneliness, perceived support, quantity of relationships, and health.

Their structure may be captured as follows. For an individual,

$$\underset{\sim}{y} = \underset{\sim}{A}_1 \underset{\sim}{g} + \underset{\sim}{A}_3 \underset{\sim}{s} + \underset{\sim}{A}_0 \underset{\sim}{u}, \tag{12}$$

where the observed vector $\underset{\sim}{y}$ is 5 x 1, and the uncorrelated latent factors $\underset{\sim}{g}$, $\underset{\sim}{s}$, $\underset{\sim}{u}$ are 5 x 1 with identity variance matrices, while the parameter matrices $\underset{\sim}{A}_1$, $\underset{\sim}{A}_3$, and $\underset{\sim}{A}_0$ are at most lower triangular. (Nonadditive genetic factors are dropped a priori, so $\underset{\sim}{A}_2$ is absent). An individual is paired with his or her twin, for whom

$$\underset{\sim}{y}' = \underset{\sim}{A}_1 \underset{\sim}{g}' + \underset{\sim}{A}_3 \underset{\sim}{s}' + \underset{\sim}{A}_0 \underset{\sim}{u}'. \tag{13}$$

The now familiar assumptions are made about cross-twin correlations among the latent factors. Gaussian maximum-likelihood estimation of the parameter matrices yields a decomposition of the 5 x 5 variance matrix of $\underset{\sim}{y}$ into its genetic and environmental constituents. This leads Lichtenstein & Pedersen to conclude, for example, that of the 0.17 correlation between perceived support and health among women, 0.15 is due to genetic factors, and 0.02 to nonshared environment.

Following Neale & Cardon (1992, Chapter 12), they refer to their specification as a Cholesky model. Indeed, the recursive structure will be

familiar to macroeconomists, but here the ordering of the elements of y is to some extent arbitrary. Behavior geneticists credit Martin & Eaves (1977) for introducing the idea of multivariate twin modeling. In the same year the economists Behrman, Taubman, & Wales (1977) empirically implemented such a twin model, one with a natural recursive ordering running from education to initial occupation to current occupation to earnings.

Objectives

The stream of human behavior-genetic research tapped here represents structural modeling in several senses: the equations depict causal links rather than mere empirical associations, the regressions among observable variables are derived in terms of more fundamental parameters, the parameters of interest are not those of the conditional expectation of one observed variable given others. However, the requirement that one of the structural parameters may change while others remain unchanged has not been invoked by the behavior geneticists.

It is fair to ask what the objectives of the behavior-genetic exercises are. Should one be reassured by a finding that broad heritability $\beta_1 + \beta_2$ is estimated robustly? What indeed does one learn from a report that genetic factors account for, say, 50% of the variance of a certain trait? It might be argued that to the extent that a trait is heritable, it is not malleable, that is, not subject to change by policy intervention. That argument is incorrect. The geneticist Newton Morton (1974) wrote "one would be quite unjustified in claiming that heritability is relevant to educational strategy. The teacher confronted with a neighborhood in which a substantial fraction of the children appear uneducable by either academic

or vocational criteria seems to me like a physical therapist treating a case of poliomelitis: neither need be concerned with the extent to which susceptibility to the observed disorder is genetic." The geneticist Richard Lewontin (1974) wrote, "The fallacy is that a knowledge of the heritability of some trait in a population provides an index of the efficacy of environmental or clinical intervention in altering the trait either in individuals or in the population as a whole." In a review article that does recognize some contributions of the behavior–genetic approach, the developmental psychologist Maccoby (2000) wrote, "... high heritability of a trait does not imply that it is not also subject to the influence of environmental factors, or that it cannot be changed by alterations in environmental conditions."

But economists need not go that far afield. After all, the behavior–genetic parameters are effectively R^2s: they measure the proportion of the variation in an observed trait that is accounted for by variation in this or that latent factor. As Cain & Watts (1970) explained years ago, such measures of "importance" are simply not indicators of policy effectiveness. Their argument was applied to the heritability context by Goldberger (1979).

Appendix

Consider a single locus at which there are two possible alleles - and +, so that individuals are either --, -+, +-, or ++. Let Z = "the score" denote the number of +s an individual has at that locus, so Z = 0, 1, 2. For simplicity suppose that the two alleles are equally prevalent, and that in equilibrium, Prob(Z=0) = 1/4, Prob(Z=1) = 1/2, Prob(Z=2) = 1/4. Assuming that all phenotypic variance is genetic, for each Z there is a phenotype Y = Y(Z), which we can code as

$$Y(0) = -a \ , \quad Y(1) = b \ , \quad Y(2) = a \ .$$

Then $E(Y) = b/2$ and $V(Y) = a^2/2 + b^2/4$. The two terms in V(Y) are the additive and nonadditive genetic variances respectively. If b = 0, Y is linear in Z, the heterozygote's phenotype is halfway between those of the homozygotes: all genetic variance is additive. If a = 0, there is no linear component in Y(Z), the two homozygotes' phenotypes are the same: all genetic variance is nonadditive.

Denote the scores of husband, wife, and child by H, W, S respectively. It is easy to verify the tabulations of $Pr(S|H,W)$ below, and then $E(Y|H,W)$ for the two extreme cases. The final column gives the probabilities for each H,W combination under the assumption of random-mating equilibrium.

		Conditional probabilities			Expected phenotypes		
H	W	S = 0	S = 1	S = 2	If b = 0	If a = 0	Pr(H,W)
0	0	1	0	0	−a	0	1/16
0	1	1/2	1/2	0	−a/2	b/2	2/16
0	2	0	1	0	0	b	1/16

1 0	1/2	1/2	0	−a/2	b/2	2/16
1 1	1/4	1/2	1/4	0	b/2	4/16
1 2	0	1/2	1/2	a/2	b/2	2/16
2 0	0	1	0	0	b	1/16
2 1	0	1/2	1/2	a/2	b/2	2/16
2 2	0	0	1	a	0	1/16

Conditional on H,W, any two (non−MZ) siblings are drawn independently, so across all families, $C(Y,Y')$, the covariance of their phenotypes, is the same as the variance of the sibship means.

For the $b = 0$ case, where $E(Y) = 0$ and $V(Y) = a^2/2$, we calculate

$$V[E(Y|H,W)] = (a^2/16)\ (1 + 4/2 + 1) = a^2/4\ ,$$

which is one−half of the additive variance. For the $a = 0$ case, where $E(Y) = b/2$ and $V(Y) = b^2/4$, we calculate

$$E[E^2(Y|H,W)] = (b^2/16)(1 + 4/2 + 1 + 1) = b^2(5/16),$$

so

$$V[E(Y|H,W)] = b^2(5/16) - (b/2)^2 = b^2/16,$$

which is one−fourth of the nonadditive variance. (A similar calculation will show that parent and child share one-half of the additive variance, and none of the nonadditive variance).

The same conclusions follow when $Y(Z)$ has both additive and nonadditive components, when allele probabilities are unequal, when there is random variation in Y for given Z, and when multiple loci are introduced: Falconer & Mackay (1996, Chapter 9). When $Y(Z)$ is not deterministic, then one extreme case has $E(Y|Z)$ linear so $BLP(Y|Z) = E(Y|Z)$, and the other has $BLP(Y|Z)$ horizontal with $E(Y|Z)$ not constant.

References

O. Ashenfelter & A. Krueger (1994), "Estimates of the economic return to schooling from a new sample of twins," American Economic Review 84: 1157–1173.

J. Behrman, P. Taubman, & T. Wales (1977), "Controlling for and measuring the effects of genetics and family environment in equations for schooling and labor market success," in P. Taubman, ed. Kinometrics: The Determinants of Socio–economic Success Within and Between Families, Amsterdam: North–Holland.

C. S. Bergeman, H. M. Chipuer, R. Plomin, N. L. Pedersen, G. E. McClearn, J. R. Nesselroade, P. T. Costa, & R. R. McCrae (1993), "Genetic and environmental effects on openness to experience, agreeableness, and conscientiousness: an adoption/twin study," Journal of Personality 61: 159–179.

G. G. Cain & H. W. Watts (1970), "Problems in making inferences from the Coleman Report," American Sociological Review 35: 228–242.

J. M. de Castro (1993), "A twin study of genetic and environmental influences on medical and psychological traits," Physiology and Behavior 54: 677–687.

D. S. Falconer & T. F. C. Mackay (1996), Introduction to Quantitative Genetics, Fourth Edition, Harlow, Essex: Longmont.

A. S. Goldberger (1979), "Heritability", Economica 46: 327–347.

A. S. Goldberger & L. J. Kamin (1998), "Behavior–genetic modeling of twins: a deconstruction," University of Wisconsin: Social Systems Research

Institute Working Paper 9824, 77 pp.

J. Joseph (1998), "The equal environment assumption of the classical twin method: a critical analysis," Journal of Mind and Behavior 19: 325–358.

L. J. Kamin & A. S. Goldberger (2002), "Twin studies in behavioral research: a skeptical view," Theoretical Population Biology 61: 83–95.

D. A. Kodde & F. C. Palm (1986), "Wald criteria for jointly testing equality and inequality restrictions," Econometrica 54: 1243–1248.

R. C. Lewontin (1974), "The analysis of variance and the analysis of causes," American Journal of Human Genetics 26: 400–411.

P. Lichtenstein, N. L. Pedersen, & G. E. McClearn (1992), "The origins of individual differences in occupational status and educational level," Acta Sociologica 35: 13–31.

P. Lichtenstein & N. L. Pedersen (1995), "Social relationships, stressful life events, and self–reported physical health: genetic and environmental influences," Psychology and Health 10: 295–319.

J. C. Loehlin (1987), Latent Variable Models, Hilldale, NJ: Erlbaum.

D. T. Lykken, W. G. Iacono, K. Haroian, M. McGue, & T. J. Bouchard (1988), "Habituation of the skin conductance response to strong stimuli: a twin study," Psychophysiology 25: 4–15.

E. E. Maccoby (2000), "Parenting and its effects on children: on reading and misreading behavior genetics," Annual Review of Psychology 51: 1–27.

N. G. Martin & L. J. Eaves (1977), "The genetical analysis of covariance structure," Heredity 38: 79–95.

N. E. Morton (1974), "Analysis of family resemblance I.

Introduction," American Journal of Human Genetics 26: 318–330.

K. McCourt, T. J. Bouchard, D. T. Lykken, A. Tellegen, & M. Keyes (1999), "Authoritarianism revisited: genetic and environmental influences examined in twins reared apart and together," Personality and Individual Differences 27: 985–1014.

M. C. Neale & L. R. Cardon (1992), Methodology for Genetic Studies of Twins and Families, Dordrecht: Kluwer Academic.

M. C. Neale & M. B. Miller (1997), "The use of likelihood–based confidence intervals in genetic models," Behavior Genetics 27: 113–120.

N. L. Pedersen, P. Lichtenstein, R. Plomin, U. DeFaire, G. E. McClearn, & K. A. Matthews (1989). "Genetic and environmental influences for Type A–like measures and related traits: a study of twins reared apart and twins reared together," Psychosomatic Medicine 51: 428–440.

N. L. Pedersen, R. Plomin, G. E. McClearn, & L. Friberg (1988), "Neuroticism, extraversion, and related traits in adult twins reared apaprt and reared together," Journal of Personality and Social Psychology 55: 950–957.

N. L. Pedersen, R. Plomin, J. T. Nesselroade, & G. E. McClearn (1992), "A quantitative genetic analysis of cognitive abilities during the second half of the life span," Psychological Science 3: 346–353.

R. Plomin, N. L. Pedersen, P. Lichtenstein, & G. E. McClearn (1994), "Variability and stability in cognitive abilities are largely genetic later in life," Behavior Genetics 24: 207–215.

T. J. Rothenberg (1972), Efficient Estimation with A Priori Information, New Haven: Yale University Press.

K. J. Saudino, N. L. Pedersen, P. Lichtenstein, G. E. McClearn, & R.

Plomin (1997), "Can personality explain genetic influences on life events?," Journal of Personality and Social Psychology 72: 196–206.

S. S. Wilks (1962), Mathematical Statistics , New York: John Wiley & Sons.

F. A. Wolak (1987), "An exact test for multiple inequality and equality constraints in the linear regression model," Journal of the American Statistical Association 82: 782–793.

TABLE 1. ALTERNATIVE MODELS FOR THREE TRAITS

	Extraversion				Neuroticism				Occupation			
OBSERVED												
Correlations	r_1	r_2	r_3	r_4	r_1	r_2	r_3	r_4	r_1	r_2	r_3	r_4
	.54	.06	.30	.04	.41	.24	.25	.28	.82	.36	.44	.44
Sample sizes	n_1	n_2	n_3	n_4	n_1	n_2	n_3	n_4	n_1	n_2	n_3	n_4
	150	204	95	220	151	204	202	201	38	42	24	36
MODEL-FITTING												
Reduced Primary Model												
	b_1	b_2	b_3	χ^2	b_1	b_2	b_3	χ^2	b_1	b_2	b_3	χ^2
SATSA ML	0	.41	.07		.31	0	.10		.60	0	.09	
My FZLS	0	.40	.06	5.84	.35	0	.06	3.83	.59	0	.20	3.46
	..	(.07)	(.06)		(.07)	..	(.06)		(.11)	..	(.11)	
Full Models												
Primary	b_1	b_2	b_3		b_1	b_2	b_3		b_1	b_2	b_3	
	-.41	.80	.12	2.29	.64	-.31	.04	1.83	.49	.10	.20	3.40
	(.22)	(.22)	(.07)		(.21)	(.22)	(.07)		(.43)	(.49)	(.12)	
Secondary	c_1	c_2	c_3		c_1	c_2	c_3		c_1	c_2	c_3	
	.78	-.40	.12	2.29	.17	.16	.04	1.83	.64	-.05	.20	3.40
	(.14)	(.11)	(.07)		(.15)	(.11)	(.07)		(.22)	(.20)	(.12)	
Agnostic	d_1	d_2	d_3		d_1	d_2	d_3		d_1	d_2	d_3	
	.38	-.01	.12	2.29	.33	.24	.04	1.83	.59	.27	.20	3.40
	(.07)	(.06)	(.07)		(.07)	(.06)	(.07)		(.11)	(.12)	(.12)	
Skeptical	t_1	t_2	t_3		t_1	t_2	t_3		t_1	t_2	t_3	
	.22	.37	-.05	1.79	.35	-.01	.06	3.83	.53	.29	.10	1.78
	(.08)	(.10)	(.08)		(.07)	(.10)	(.08)		(.13)	(.14)	(.15)	

TABLE 2. MONTE CARLO RESULTS

Unrestricted		Pretest Estimator				
			Conditional		Unconditional	
	(1)		(2)	(3)	(4)	(5)
Mean b_1	.389	Mean b_1^*	–	.499	.291	.343
Mean b_2	.109	Mean b_2^*	.463	–	.208	.154
Mean b_3	.298	Mean b_3^*	.343	.293	.302	.302
Mean $s(b_1)$	.268	Mean $s(b_1^*)$	–	.079	.272	.185
Mean $s(b_2)$	.248	Mean $s(b_2^*)$	.077	–	.252	.151
Mean $s(b_3)$	.081	Mean $s(b_3^*)$	.075	.076	.082	.079
$SD(b_1)$	.264	$SD(b_1^*)$	–	.079	.146	.179
$SD(b_2)$	.247	$SD(b_2^*)$	.079	–	.131	.164
$SD(b_3)$	.079	$SD(b_3^*)$	.078	.078	.077	.078