

Genetic Tobit Factor Analysis: Quantitative Genetic Modeling with Censored Data

Niels G. Waller^{1,3} and Bengt O. Muthén²

Received 26 Oct. 1991—Final 11 Dec. 1991

Parameters of quantitative genetic models have traditionally been estimated by either algebraic manipulation of familial correlations (or familial mean squares), biometric model fitting, or multiple-group covariance structure analysis. With few exceptions, researchers who have used these methods for the analysis of twin data have assumed that their data were multinormal and, consequently, have used normal-theory estimation methods. It is shown that normal-theory methods produce biased genetic and environmental parameter estimates when data are censored. Specifically, with censored data, (1) normal-theory estimates of narrow-sense heritability are either positively or negatively biased, whereas (2) estimates of shared-familial environmental variance are always biased downward. An alternative method for estimating genetic and environmental parameters from censored twin data is proposed. The method is called genetic Tobit factor analysis (GTFA) and is an extension of the Tobit factor analysis model developed by Muthén (Br. J. Math. Stat. Psychol. 42, 241–250, 1989). Using a Monte Carlo design, the performance of GTFA is compared to traditional quantitative genetic methods in both large and small data sets. The results of this study suggest that GTFA is the preferred method for the genetic modeling of censored data obtained from twins.

KEY WORDS: genetic factor analysis; censored twin data; Tobit.

¹ Department of Psychology, University of California, Davis, California 95616.

² University of California, Los Angeles, California 90024.

³ To whom correspondence should be addressed.

INTRODUCTION

Two approaches to estimating behavioral genetic parameters have dominated the field of human quantitative genetics during the last three decades (Eaves *et al.*, 1989). The first, known as the classical method, involves the algebraic manipulation of familial correlations, or familial mean squares, to derive simple estimates of genetic and environmental variance ratios. Falconer's (1981) formula for heritability, which is calculated as twice the difference between the correlations among monozygotic and dizygotic twins, is a well-known example. The second method is biometrical analysis (Mather and Jinks, 1982). Originally used in the area of plant and animal genetics (Mather, 1949), the biometrical approach has been used increasingly in human behavior genetics since it was first introduced to the field by Jinks and Fulker (1970). Biometrical analysis offers several advantages over the classical method since it is framed within the general linear model and, thus, enables researchers to assess rigorously the relative fit of alternative genotype-environmental models (Eaves *et al.*, 1978).

In recent years, a third method, known as covariance structure analysis (Jöreskog and Sörbom, 1979) has also been used increasingly in human quantitative genetic studies (Boomsma *et al.*, 1989). Not surprisingly, the latter method is rapidly gaining popularity since a plethora of genotype-environmental models can be parameterized easily within this framework using widely available covariance structure software such as **LISCOMP** (Muthén, 1987a), **LISREL** (Jöreskog and Sörbom, 1988), **EQS** (Bentler, 1989), and **MX** (Neale, 1991; see also Heath *et al.*, 1989; Neale and Martin, 1989).

With few exceptions, researchers who have used these methods for the quantitative resolution of behavioral variation have assumed that their data were sampled from a multinormal parent distribution. For many data sets in the social sciences the assumption of multivariate normality is untenable. Specifically, many psychological variables have distributions with a large proportion of cases at the extremes of the scale. Such variables are said to be censored (Muthén, 1990; Tobin, 1958). When familial data are censored, that is, when the observed scores for a phenotypic characteristic exhibit marked ceiling or floor effects, genetic and environmental parameters will be biased if estimated by traditional, normal-theory, quantitative genetic methods. (Note that our use of the term *censoring* refers to restricted variance due to floor and ceiling effects. The term censoring is also used in the literature to refer to restricted variance due to differential survival times in event history analysis. The

Tobit estimator, discussed below, is not appropriate for the latter type of censoring.)

Genetic Tobit factor analysis (GTFA), a covariance structure model that exploits the latent information contained in censored data, is proposed as an alternative means of estimating behavior genetic parameters from censored data obtained from identical and fraternal twins. This model is an extension of the Tobit factor analysis model developed by Muthén (1989a). In the following sections we describe the GTFA model and present Monte Carlo results that argue strongly for the use of GTFA when analyzing censored twin data.

The basic theory of covariance structure analysis, as it applies to the analysis of twin data, has been discussed repeatedly in *Behavior Genetics* (see, for example, the recent special issue entitled *Twin Methodology Using LISREL*, edited by Boomsma, Martin, and Neale, 1989). Hence, we do not review this material in detail. However, because most previous behavior genetic applications of covariance structure modeling have been discussed from the perspective of the **LISREL** model and computer program (Jöreskog and Sörbom, 1988), some preliminary discussion is in order since the GTFA model was conceptualized from the perspective of Muthén's (1984, 1987a) **LISCOMP** model and corresponding program. In the following section the **LISCOMP** model is briefly described.

The **LISCOMP** model (Muthén, 1984), like the related **LISREL** model (Jöreskog, 1973), has two basic components: a measurement model and a structural equations model. The measurement model depicts the regression relations among the p observed (measured) dependent variables, y , and the m unobserved (latent) variables, η . One way of conceptualizing the measurement model is as a restricted factor analysis model with m latent factors. These regression relations define a series of linear equations which can be expressed in matrix notation as follows:

$$y = \nu + \Lambda\eta + \epsilon \quad (1)$$

where y is a $p \times 1$ vector of measured variables, ν is a $p \times 1$ vector of intercept terms, Λ is a $p \times m$ vector of factor loadings, η is an $m \times 1$ vector of latent variables, and ϵ is a $p \times 1$ vector of error terms.

The structural equations denote the regression relations among the latent variables, as well as regression relations among the latent variables and the exogenous-observed variables (defined as x variables in the model, i.e., the regressions among the latent factors and the observed predictor variables). These relations can also be expressed in matrix notation as

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta} \quad (2)$$

where $\boldsymbol{\eta}$ is as defined above, $\boldsymbol{\alpha}$ is an $m \times 1$ vector of latent variable means and intercepts, \mathbf{B} is an $m \times m$ matrix of regression weights among the latent variables, \mathbf{x} is a q -vector of observed exogenous variables, $\boldsymbol{\Gamma}$ is an $m \times q$ matrix of regression weights of the m latent variables on the q observed-exogenous variables, and $\boldsymbol{\zeta}$ is the $m \times 1$ vector of equation disturbance terms.

For a given model, the measurement and structural equations define a covariance matrix of a specified form. For example, reexpressing Eq. (2) in reduced form, and substituting this expression for $\boldsymbol{\eta}$ in Eq. (1), defines the endogenous y variables as follows:

$$y = \boldsymbol{\nu} + \boldsymbol{\Lambda}[(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\mathbf{x} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\zeta}] + \boldsymbol{\epsilon} \quad (3)$$

which implies the following expressions for the first- and second-order moments (i.e., the means, variances, and covariances) of the observed data:

$$E(y|x) = \boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\mathbf{x} \quad (4)$$

$$\text{Cov}(y|x) = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})' - \boldsymbol{\Lambda}' + \boldsymbol{\Theta} \quad (5)$$

where $\boldsymbol{\Psi}$ is an $m \times m$ covariance matrix for the latent factors and the residuals in the latent factor relations, and $\boldsymbol{\Theta}$ is a $p \times p$ covariance matrix for the residuals in the measurement relations.

Note that in the **LISCOMP** model, in contrast to the **LISREL** model, the expected first- and second-order moments of the observed-endogenous (y) variables are conditioned on the observed-exogenous (x) variables if they are present in the model. This feature of **LISCOMP** is useful in quantitative genetic applications, where it is often important to partial out the effects of age or sex from the endogenous variables of interest.

THE GENETIC TOBIT FACTOR ANALYSIS MODEL

Figure 1 illustrates a path diagram of the GTFA model (see the Appendix for the **LISCOMP** command file for this model). Note that this model is similar to the multiple-group, univariate behavior genetic models described by Heath *et al.* (1989), with the important exception that age in our model is treated as an observed, rather than a latent variable. In the simplified model the phenotypic variance of a metrical character is putatively caused by three latent influences: (1) a unique environmental factor, U_e ; (2) an additive genetic factor, A ; and (3) a

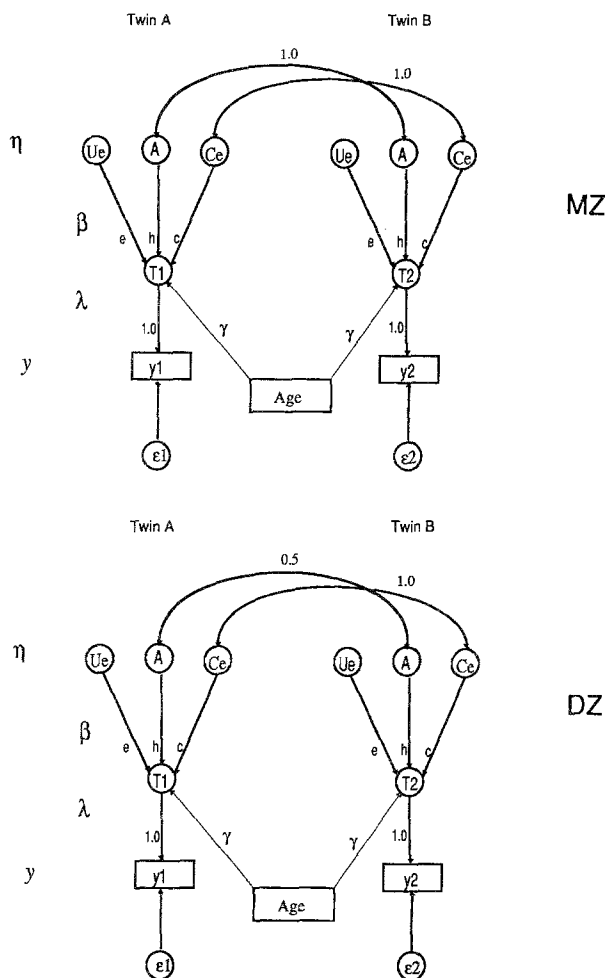


Fig. 1. Genetic Tobit factor analysis model.

common-family environmental factor, C_e (note that the residual factors, ϵ_1 and ϵ_2 are not identified in this model and are, therefore, fixed to zero). Additionally, the observed scores for Twins A and B are regressed on age, which, as noted above, is modeled here as an observed, rather than a latent, causal influence. Furthermore, as described below, the regressions of the observed variables on the factors and age are Tobit regressions.

The Effects of Censoring on Genetic and Environmental Parameter Estimates

A number of researchers have recently argued, sometimes vociferously (Bradley, 1980), that the assumption of normality is often untenable with biometric and/or social science data (Andrews *et al.*, 1972; Ansell, 1973; Boomsma, 1983; Browne, 1982, 1984, 1987; Harlow, 1985; Hill and Dixon, 1982; Muthén, 1984, 1987b; Muthén and Kaplan, 1985; Tanaka, 1984; Tukey and McLaughlin, 1963; Wainer and Thissen, 1976). One form of nonnormality that occurs with sufficient frequency to warrant special attention in quantitative genetic analyses is that of censored data. Data are considered censored when there is limited variability at the end points of a metric distribution (Maddala, 1983; Muthén, 1990; Tobin, 1958). Another way of stating this is that censoring occurs when a large number of cases are located at one or another (or both) end points of a scale. In the behavioral sciences, scales associated with many cases at either the high or the low end of the score distribution are said to exhibit ceiling or floor effects, respectively. This will occur when the difficulty of a test is either too easy or too hard for the specific sample tested; thus censoring results from the *interaction* between the test and the sample characteristics and should not be thought of as arising solely from the psychometric properties of the test. Any test can manifest ceiling or floor effects if administered to the "right" sample. For example, Eaves *et al.* (1989) note that "the neuroticism scale of the Eysenck Personality Questionnaire . . . reveals substantial 'ceiling' and 'floor' effects because the collection of binary items does not discriminate as effectively in the tails of the distribution as its center" (p. 256).

One possible consequence of censoring that has received scant attention in the psychometric literature is that genetic and environmental parameters estimated from censored data are systematically biased if the analyses are conducted via traditional behavior genetic methods. This is easily shown by observing the effects of censoring on the Pearson product moment correlation.

With two variables many combinations of censoring are theoretically possible. Each variable can be uncensored, censored from above, below, or doubly censored, and the degree of censoring need not be the same for both variables. Formulae for deriving the moments of the censored and/or truncated bivariate normal distribution, which can be used to investigate the effects of censoring on simple correlations, have recently been published by Muthén (1990).

Table I illustrates the effects of censoring on the Pearson product-moment correlation. The data in Table I represent the *percentage* of the

Table I. Percentage of Latent Correlation Retained Under Nine Censoring Conditions^a

R_1	Percentage censoring from below								
	10	20	30	40	50	60	70	80	90
.95	.997	.994	.990	.986	.981	.974	.966	.953	.931
.90	.994	.988	.981	.973	.964	.952	.936	.913	.874
.85	.992	.983	.973	.962	.948	.931	.908	.876	.822
.80	.990	.979	.966	.951	.933	.911	.882	.842	.774
.75	.988	.974	.959	.940	.919	.892	.858	.809	.729
.70	.986	.970	.952	.930	.905	.874	.834	.778	.686
.65	.984	.966	.945	.920	.891	.856	.811	.748	.646
.60	.983	.962	.938	.911	.878	.839	.788	.719	.608
.55	.981	.958	.932	.902	.866	.822	.766	.691	.572
.50	.980	.955	.926	.892	.853	.805	.745	.664	.538
.45	.978	.951	.920	.883	.840	.789	.724	.637	.505
.40	.977	.948	.914	.874	.828	.773	.703	.612	.474
.35	.975	.944	.908	.866	.816	.757	.683	.587	.444
.30	.974	.941	.902	.857	.804	.741	.663	.562	.415
.25	.973	.938	.896	.848	.792	.725	.644	.538	.388
.20	.972	.935	.891	.840	.780	.710	.624	.515	.362
.15	.971	.932	.885	.831	.769	.695	.605	.493	.337
.10	.970	.929	.880	.823	.757	.679	.586	.470	.313
.05	.969	.926	.874	.815	.745	.664	.568	.449	.291

^a R_1 —latent, unobserved correlation.

latent (i.e., uncensored) correlations retained under nine censoring conditions when both variables have been censored from below from between 10 and 90%. Note that the effects of censoring are not constant across the absolute range of the correlations.

A somewhat crude, though common method of deriving genetic and environmental parameter estimates is to examine weighted differences between the monozygotic (MZ) and the dizygotic (DZ) twin intraclass correlations (e.g., using Falconer's formula for heritability). Using simple formulas for estimating genetic and environmental variance ratios and the data reported in Table I will allow us to examine the effects of censoring on genetic and environmental parameter estimates.

Define R_{mz} and R_{dz} as the latent (i.e., nonobserved) intraclass correlations between n pairs of MZ and DZ twins, respectively. Furthermore, define a and b as the entries in Table I that correspond to these latent correlations at the appropriate censoring levels. Observed genetic and environmental parameter values: h^{2*} (narrow-sense heritability), c^{2*}

(shared-familial environmentality), and e^{2*} (unique environmentality) are derived as followed:

$$h^{2*} = 2(aR_{mz} - bR_{dz}) \quad (6)$$

$$c^{2*} = 2bR_{dz} - aR_{mz} \quad (7)$$

$$e^{2*} = 1 - (h^{2*} + c^{2*}) \quad (8)$$

For example, if the latent, uncensored MZ and DZ correlations are .70 and .45, respectively, and the data are 60% censored, then $h^{2*} = 2(.874 * .70 - .789 * .45) = .514$; $c^{2*} = 2(.789 * .45) - (.874 * .70) = .099$; and $e^{2*} = 1 - (.514 + .099) = .387$. Compare these estimates to the unbiased parameter values: $h^2 = .50$, $c^2 = .20$, and $e^2 = .30$. Notice that in the current case h^{2*} and e^{2*} are positively biased, whereas c^{2*} is negatively biased.

Now consider a second example in which the latent MZ and DZ correlations are .50 and .30, respectively, and again, the data are 60% censored. The observed genetic and environmental parameter estimates are $h^{2*} = 2(.805 * .50 - .741 * .30) = .360$; $c^{2*} = 2(.741 * .30) - (.805 * .50) = .042$; and $e^{2*} = 1 - (.360 + .042) = .598$. The corresponding unbiased estimates are $h^2 = .40$, $c^2 = .10$, and $e^2 = .50$.

In the second example h^2 is *negatively* biased, whereas in the first case it was positively biased. The shift in bias is a function of the relative magnitudes of the twin correlations in the two examples, a fact which raises an important point. That is, the bias in h^2 will often be domain specific. For example, in the cognitive abilities domain, where the MZ and DZ correlations are similar to those in the first example, h^2 may be positively biased if the data are censored, whereas in the personality domain (Tellegen *et al.* 1988), where the twin correlations are more like those of the second example, h^2 will be negatively biased. c^2 will usually be biased downward. In many cases the bias in c^2 will be so extreme that conventional behavior genetic methods will suggest that c^2 is not statistically distinguishable from zero. Plomin and Daniels (1987) have recently concluded that c^2 effects are rarely appreciable for personality and attitudinal variables. Could censoring have played a role in these findings? In the following section we describe the Tobit estimator and review more formally the effects of censoring on genetic and environmental parameter estimates using Monte Carlo simulations.

The Tobit Estimator

To date, most investigators working with covariance structures have relied on normal-theory estimators, such as normal-theory maximum

likelihood (NTML) or normal-theory generalized least squares (NTGLS). These estimators are appropriate when the distribution of the observed data, y , is multivariate normal. Browne (1982, 1984, 1987), Muthén (1987a,b), and others have noted that the assumption of a multinormal y is often untenable with social science data [see Grayson (1990) for a discussion of the importance of scale in quantitative genetics research].

In the social sciences, constructs are often measured on 7-, 5-, 3-, or even 2-point scales. The effects of such response categorization (Peters and Van Voorhis, 1940, p. 393) or scale transformations (Grayson, 1990) are well-known and may include, among other things, severe skewing of the latent distribution if the response cutpoints are not judiciously chosen (i.e., given sample characteristics).

Muthén (1984, 1987a,b) has developed estimators for single- and multiple-group covariance structure models that assume that the underlying, *latent* response variables, y^* , are continuous, normally distributed. The observed, y , distributions may be nonnormal, however, because of crude measurement. In this approach, latent correlations (or covariances), such as tetrachorics (Divgi, 1979), polychorics (Olsson, 1979b), polyserials (Olsson *et al.*, 1982), and/or Tobit (Muthén, 1990; Tobin, 1958) correlations, are used during parameter estimation rather than more traditional product moments. Normal-theory estimators cannot be used with y^* models because the latent correlations (or covariances) have larger sampling errors than the corresponding product-moment correlations.

Noting limitations of normal-theory methods, Browne (1982, 1984) developed two estimators that impose less stringent distributional requirements than either the NTML or the NTGLS estimators. Browne called his estimators elliptical (ELP) and asymptotic distribution free (ADF). Elliptical estimators are appropriate when all univariate distributions have equal (though not necessarily zero) kurtosis and zero skew, whereas ADF estimators, being more general, are appropriate regardless of the distributional properties of the data *as long as the data conform to the assumptions of the general linear model*. An important assumption of the general linear model is that the model residuals have zero expectation.

ELP and ADF estimators, although useful for many data types, are not appropriate for censored data. When data are censored, the residuals at the scale end points cannot have zero expectation, "only positive or negative residuals are possible, depending on whether the censoring is from below or above" (Muthén, 1989a, p. 242). The Tobit estimator was designed by Muthén specifically for the case of censored data. Initial studies of this estimator suggest that when the data exhibit moderate to extreme censoring, the Tobit estimator yields more accurate model output

(e.g., less biased parameter values, standard errors, etc.) than either the NTML, the NTGLS, or the ADF estimators (Muthén, 1989a).

Assuming a y^* model, let s be defined as the vector of nonredundant elements of $\text{Cov}(y^*)$, and σ as the vector of nonredundant elements of Σ , the model-implied covariance matrix. Furthermore, let F represent a weighted least-squares function to be minimized with respect to the model parameters. Define F as

$$F = (s - \sigma)'W^{-1}(s - \sigma) \quad (9)$$

F is the sum of the weighted squared differences between the observed and the predicted covariances (or correlations) of the model. Let W be a consistent estimator of the asymptotic covariance matrix of s (see Browne, 1982). The choice of W will determine the type of estimator that results from Eq. (9). When $W = I$, the identity matrix, the above expression results in the unweighted least-squares estimator. If the observed variables are normal, and s is a sample covariance, setting W to a function of S , the observed covariance matrix, results in the NTGLS estimator; setting W to a function of Σ , the model implied covariance matrix, results in the NTML estimator.

In Tobit factor analysis the elements of s are estimated in a two-step procedure. First, the mean and variance of the censored variables are calculated by maximum likelihood (Cohen, 1950, 1955; Des Raj, 1953; Gupta, 1952). Second, maximum-likelihood estimates of the Tobit covariances [or correlations (Muthén, 1990)] are obtained from bivariate information, holding the means and variances from the first step fixed at their estimated values. W is obtained by summing first-order derivative products of the univariate and bivariate log likelihood equations from steps one and two (for details see Muthén, 1984). In the following section the Tobit estimator is compared to normal-theory and asymptotic distribution free estimators in a univariate behavior genetic design.

Study Design

Three estimators are compared for their ability to recover known genetic and environmental parameter values from a multiple-group covariance structure model: (1) NTML (Jöreskog, 1973), (2) Browne's (1982, 1984) asymptotic distribution free generalized least squares (ADF), and (3) Muthén's (1989a) Tobit estimator (Tobit). Because both the ADF and the Tobit estimators use sample information when constructing the model weight matrix, the performance of these estimators is compared under varying sample sizes.

The question of "How big [a sample] is big enough?" (Tanaka,

1987) for covariance structure models is difficult to answer. Previous Monte Carlo studies have looked at sample size requirements for a handful of factor analysis models (Boomsma, 1983; Geweke and Singleton, 1980; Tanaka, 1984), though none of these studies included multiple-group models in their designs. Thus, it is ambiguous how the findings from these earlier investigations apply to the types of models used in the analysis of twin data.

All three estimators included in the present study were developed using results from asymptotic (i.e., large-sample) statistical theory. Consequently, when data meet the appropriate distributional requirements of the estimator, and when the sample size is sufficiently large, each estimator produces parameter estimates with desirable statistical properties (e.g., consistency, unbiasedness, asymptotic efficiency). Model estimation may break down, however, in finite samples.

In single-group analyses a common rule of thumb for determining sample size requirements is to have a ratio of 10:1 of subjects to parameters. This rule seems to work well in factor analysis models (Tanaka, 1987) when the data are multinormal and when parameters are estimated using normal-theory methods (e.g., NTML, NTGLS). When data are not multinormal, however, or when alternative estimators such as the ADF and Tobit estimator are used, a higher ratio of subjects to parameters may be needed. Tanaka (1984) found that with the ADF estimator, larger samples were needed to obtain accurate parameter standard errors, test statistics, and other model output. This finding suggests that, in addition to parameter estimates, other aspects of model output need to be considered when evaluating sample size requirements.

A review of the covariance structure literature (Tanaka, 1984; Muthén, 1989a) suggests that at least four components of model output should be evaluated at each sample size: (1) parameter estimates, (2) theoretical estimates of standard errors (computed from the model information matrix), (3) empirical estimates of standard errors (computed over replications), and (4) accuracy of the test statistic. Each of these components was examined in the present investigation.

Sample size was varied such that each model was investigated using a "large sample" and a "small sample." The large sample contained simulated data representing 1000 MZ and 1000 DZ twin pairs. This sample contained considerably more data than found in the average twin study, though it by no means exceeds the limits of the larger registries (e.g., Floderus-Myrhed *et al.*, 1980,).

In the large sample the ratio of cases to parameters was 333:1 in each zygosity group. Hence, results from asymptotic theory are likely to be valid at this sample size. The small sample was composed of simulated

data representing 100 MZ and 100 DZ twin pairs. The ratio of cases to parameters at this sample size is approximately 33:1. It is unclear whether asymptotic theory applies at this more modest sample size. Because many twin studies have used even fewer subjects, however, it seemed prudent to observe the behavior of the models under this condition as well.

To study the effects of censoring, data were generated under four conditions. These conditions ranged from no censoring to extreme censoring. When comparing alternative estimators it is important to investigate a range of censoring conditions to determine at which point normal theory methods provide biased results. The four censoring conditions studied included 0, 20, 40, and 60% censoring from below. These conditions simulate zero, mild, moderate, and extreme censoring.

We hypothesized that under 0% censoring all estimators would produce consistent estimates of model parameters and that the NTML estimates would have standard errors with the smallest finite sampling variability (i.e., the NTML estimates would exhibit asymptotic efficiency). Under increasing conditions of censoring, the Tobit estimator should outperform the NTML and ADF estimators. The study design is a 3 estimators \times 4 censoring conditions \times 2 samples size factorial design. Each condition was replicated 500 times to investigate the sampling behavior of the model parameters, standard errors, and test statistic, and each replicate was analyzed by the three estimators.

In the genetic and environmental analysis of univariate twin data, only a small number of models are of potential interest. In the present study, a simple model that included additive genetic effects (A), within-family, unique environmental effects (U_e), and between-family, common environmental effects (C_e) was used for all simulations (age was not modeled in the simulations). The known genetic and environmental parameter values for this model were as follows $h^2 = .50$, $c^2 = .10$, and $e^2 = .40$. These values were chosen since they are typical of those found in many behavior genetic studies of personality (e.g., Tellegen *et al.*, 1988), leisure-time and occupational interests (Waller *et al.*, 1990). Although it has been difficult to document common family environmental effects in personality data (Plomin and Daniels, 1987), inclusion of c^2 in our model was deemed important because of the previously noted effect of censoring on this parameter.

All data were generated and analyzed with the data simulation and model estimation facilities of the LISCOMP (Muthén, 1987a) computer program. Data were initially sampled from a multinormal parent distribution satisfying the aforementioned parameter values. For the censoring conditions, the univariate responses were subsequently censored in ac-

cordance with the study design. Covariances, rather than correlations, were analyzed throughout (see Cudeck, 1989, for a general discussion).

Several hypotheses were tested. (1) Under increasing conditions of censoring it was predicted that the NTML test statistic would depart from the expected χ^2 distribution and tend toward inflated values. The ADF test statistic should be less affected by moderate levels of censoring since the ADF estimator was developed expressly to handle nonnormal continuous data. When censoring becomes extreme, however, the ADF test statistic should also tend toward inflated values, as the linear model no longer holds when the residuals are restricted in range. The Tobit test statistic should remain valid under all censoring conditions. (2) Under all censoring conditions the NTML and ADF estimators will produce biased parameter estimates. Specifically, given the relative magnitude of the MZ and DZ correlations resulting from the model parameters, estimates of h^2 and c^2 will be negatively biased, whereas estimates of e^2 will be positively biased. Bias is expected to increase under moderate to extreme censoring. The Tobit estimator should produce the least biased parameter estimates under all conditions. (3) Theoretical estimates of parameter standard errors will be incorrect for NTML and ADF estimation under moderate to extreme censoring because of the violation of the linear model. With the Tobit estimator, the theoretical standard errors will be close to the empirically derived standard errors.

RESULTS

The first simulation examined the behavior of the NTML, ADF, and Tobit test statistics under varying degrees of censoring in the large sample. With multinormal data, all three test statistics should be distributed as a central χ^2 variate since our model is known to hold in the population. The expected value and variance of a central χ^2 variate are equal to the model degrees of freedom and twice the model degrees of freedom, respectively (Hogg and Tanis, 1983, p. 191). In our model three parameters (h , c , e) are estimated from six summary statistics (two variances and one covariance in each zygosity group); hence, the model has 3 degrees of freedom.

Table II reports a summary of findings, with respect to the relative performance of the test statistics of the three estimators. The *reject proportion* is the proportion of replications for which the model is rejected at the 5% level. Specifically, in the present study, it is the proportion of replications of 500 in which the test statistic exceeds 7.815, the $\alpha = .05$ critical value for a central χ^2 distribution with 3 degrees of freedom.

Table II. Chi-Square and Reject Proportions for All Cases of Censoring^a

Percentage censoring	NTML	ADF	Tobit
0%			
Mean	2.958	2.974	2.946
Variance	5.527	5.643	5.523
Reject proportion	.046	.040	.046
20%			
Mean	2.841	3.108	3.032
Variance	5.185	6.426	5.823
Reject proportion	.046	.052	.048
40%			
Mean	4.303	2.871	2.948
Variance	11.243	5.258	5.802
Reject proportion	.124	.036	.050
60%			
Mean	9.058	3.271	3.077
Variance	52.056	6.807	6.896
Reject proportion	.448	.068	.050

^a Degrees of freedom = 3. $N = 1000$ MZ pairs and 1000 DZ pairs.

If the model holds, and the estimators are performing correctly, the test statistic should exceed this value approximately 5% of the time due to chance alone.

Inspection of Table II reveals that under 0% censoring, the mean NTML, ADF, and Tobit test statistics, averaged over 500 replications, are close to their expected value of 3.0. The variances of the test statistics are also close to, but slightly lower than, their expected value of 6.0. These results suggest that under 0% censoring, the NTML, ADF, and Tobit test statistics are indeed distributed as central χ^2 variates at this sample size. Furthermore, under mild censoring (e.g., up to 20%), all three indices can still be relied upon as valid indicators of model fit. At moderate to extreme censoring, however, this is no longer true. At 40% censoring the NTML test statistic rejects the true model approximately 2.5 times too often, whereas at 60% censoring the true model is rejected almost nine times too frequently. The ADF test statistic performs better under these conditions, though it quickly becomes too large as censoring becomes extreme. The Tobit test statistic performs well under all conditions.

Our results suggest that under moderate censoring the NTML test statistic is an unreliable indicator of model fit. The ADF test statistic performs better than the NTML index but breaks down as the degree of

censoring increases. These findings are relevant when considering other fit indices such as the Akaike (1974), Schwarz (1978), or Cudeck-Browne (1983), which are themselves functions of χ^2 .

With multinormal data the NTML, ADF, and Tobit estimators are expected to produce consistent estimates of model parameters. Additionally, the NTML estimates have the smallest asymptotic sampling variability of all consistent estimates [i.e., by the Rao-Cramer lower bound (Hogg and Tanis, 1983, p. 332)]. Prior studies have found that GLS estimates exhibit small negative bias in finite samples (Browne, 1982; Tanaka, 1984). Thus we expect to find some bias in the ADF and Tobit estimates since both of these estimators belong to the class of GLS estimators. [We note that although GLS and NTML estimates are asymptotically equivalent (Browne, 1974), they will not necessarily be the same in finite samples.]

Table III reports average parameter estimates and average bias for four censoring conditions in the large sample. Several features of this table deserve comment. First, note that the estimates are reported in the metric of the observed data. Later these results will be reported in a standardized metric (i.e., expressing h^2 , c^2 , and e^2 as proportions of total variance). Second, note that with 0% censoring the NTML, ADF, and Tobit estimates are close to their true values, although the Tobit estimate of c is slightly too small.

In general, as censoring becomes more extreme, the NTML and ADF estimates become increasingly biased downward. For example, under moderate to extreme censoring, estimates of h are attenuated by a factor of 35 to 55%, whereas estimates of c are too small by as much as 50 to 75%, and e by as much as 45%. The Tobit estimates of h and e , however, are close to their true values under all conditions, though under extreme censoring c is negatively biased by about 10%. The bias in c , however, is more than three times as large under mild censoring with the NTML and ADF estimators than under extreme censoring with the Tobit estimator.

Table IV reports average estimates of h^2 , c^2 , and e^2 , and average bias for the four censoring conditions in the large sample. Comparison of Tables III and IV reveals that parameter bias is metric dependent. In the metric of the observed data all parameters are negatively biased under varying degrees of censoring. In the standardized metric this is no longer true. In all cases bias is most extreme with the NTML or ADF estimators and least extreme with the Tobit estimator.

The ratio of a parameter estimate to its standard error is a frequently used guide for determining parameter significance [though it is often a misleading one, see Neale *et al.* (1989)]. It is important, therefore, to

Table III. Estimates of Genetic and Environmental Paths and Percentage Bias for Four Censoring Conditions^a

Parameter	Estimator (% censoring)											
	NTML				ADF				Tobit			
	0	20	40	60	0	20	40	60	0	20	40	60
<i>e</i> .632 (.0)	.632 (.0)	.542 (-14.2)	.451 (-28.6)	.350 (-44.6)	.632 (.0)	.541 (-14.4)	.450 (-28.8)	.349 (-44.8)	.630 (-.3)	.628 (-.6)	.627 (-.8)	.624 (-1.3)
<i>h</i> .707 (-.3)	.705 (-.3)	.584 (-17.4)	.462 (-34.6)	.331 (-53.2)	.704 (-.4)	.583 (-17.5)	.460 (-34.9)	.329 (-53.5)	.704 (-.4)	.705 (-.2)	.700 (-1.0)	.694 (-1.8)
<i>c</i> .316 (-3.8)	.304 (-3.8)	.220 (-30.4)	.148 (-53.2)	.082 (-74.1)	.303 (-1.0)	.218 (-31.0)	.146 (-53.8)	.081 (-74.4)	.299 (-5.4)	.289 (-9.5)	.287 (-9.2)	.289 (-8.5)

^a *N* = 1000 MZ pairs and 1000 DZ pairs.

Table IV. Average Estimates and Bias for h^2 , c^2 , and e^2 for Four Censoring Conditions^a

Parameter	NTML						ADF						Tobit					
	Estimator (% censoring)																	
	0	20	40	60	0	20	40	60	0	20	40	60	0	20	40	60	0	20
e^2 .400	.404 (1.0)	.430 (7.5)	.464 (16.0)	.513 (28.3)	.405 (1.3)	.430 (7.5)	.465 (16.3)	.515 (28.8)	.404 (1.0)	.405 (1.3)	.407 (1.8)	.408 (2.0)	.405 (1.3)	.407 (1.8)	.408 (2.0)			
h^2 .500	.503 (.6)	.499 (-.2)	.486 (-2.8)	.459 (-8.2)	.502 (.4)	.500 (.0)	.486 (-2.8)	.457 (-8.6)	.504 (.8)	.510 (2.0)	.508 (1.6)	.504 (.8)	.504 (.8)	.508 (1.6)	.504 (.8)			
c^2 .100	.093 (-7.0)	.071 (-29.0)	.050 (-50.0)	.028 (-72.0)	.093 (-7.0)	.070 (-30.0)	.049 (-51.0)	.028 (-72.0)	.091 (-9.0)	.085 (-15.0)	.085 (-15.0)	.087 (-13.0)	.085 (-15.0)	.085 (-15.0)	.087 (-13.0)			

^a N = 1000 MZ pairs and 1000 DZ pairs.

study the effects of censoring on the estimation of parameter standard errors, since extreme standard error bias may lead researchers to accept erroneous models. Table V reports average estimates of the theoretical and empirical standard errors for the four censoring conditions and three estimators in the large sample.

The empirical standard errors are defined as the standard deviations of the observed parameter estimates over the 500 replications. The theoretical standard errors are computed from the model information matrix which is obtained during parameter estimation (Browne, 1982). When the estimated and "true" model are equivalent, the empirical and theoretical standard errors should be in close agreement (Browne, 1982).

Table V. Theoretical and Empirical Sampling Variability of Three Estimators in Large Samples^a

Parameter	NTML	ADF	Tobit
0%			
<i>e</i>	.014 ^b	.014	.020
	.015 ^c	.015	.019
<i>h</i>	.029	.045	.045
	.045	.045	.053
<i>c</i>	15.441	25.294	9.523
	.094	.095	.107
20%			
<i>e</i>	.012	.013	.023
	.013	.013	.022
<i>h</i>	.026	.042	.048
	.038	.038	.058
<i>c</i>	47.050	111.588	68.224
	.098	.099	.121
40%			
<i>e</i>	.010	.012	.027
	.013	.013	.024
<i>h</i>	.023	.039	.054
	.036	.036	.067
<i>c</i>	146.164	209.756	246.588
	.089	.090	.144
60%			
<i>e</i>	.007	.012	.034
	.012	.011	.031
<i>h</i>	.019	.037	.065
	.031	.030	.082
<i>c</i>	835.172	1361.739	101.261
	.076	.075	.160

^a N = 1000 MZ pairs and 1000 DZ pairs.

^b Mean of estimated standard errors.

^c Empirical standard deviation of estimates.

The data reported in Table V illustrate that in the absence of censoring, the theoretical and empirical standard errors are indeed similar for h and e . With c , however, the theoretical standard errors are more variable than their empirical counterparts by a factor of 100 with the Tobit estimator and by factors of 164 and 260 with the NTML and ADF estimators, respectively. Closer inspection of the model output revealed that when c deviates below its expected value, and is estimated close to zero, the second derivative of the fit function with respect to c becomes increasingly small and tends toward zero. Hence when the expected matrix of second-order partial derivatives is inverted to obtain the theoretical standard errors the variance of this parameter becomes increasingly large—and sometimes huge. These results suggest that regardless of whether the NTML, ADF, or Tobit estimators are used, researchers should refrain from relying on theoretical standard errors when deciding on whether shared-familial environmental effects explain significant variance in their models. Neale *et al.* (1989, p. 43–44) provide additional reasons for avoiding the use of theoretical standard errors in behavior genetic applications.

Small-Sample Behavior of the Estimators

Previously we noted that the NTML, ADF, and Tobit estimators were developed using asymptotic statistical theory and that these methods may provide biased results in small samples. We also noted that the meaning of “small” was unclear in the context of genotype–environment covariance structure models. All previous simulations were repeated at the small sample size to investigate the feasibility of small sample genetic modeling with censored data.

Table VI reports the behavior of the NTML, ADF, and Tobit test statistics under the four censoring conditions at the small sample size. Note that in the absence of censoring, the expected test statistics are invariably too large with all three estimators. The variances of the test statistics are also greater than their expected values, and the true model is rejected far more than expected by chance alone. As censoring increases, the test statistics deviate farther from the expected χ^2 distribution with both the NTML and the ADF estimators. For example, under extreme censoring the NTML reject frequency is approximately 10 times larger than expected; the ADF test statistic is better behaved, though it rejects a true model more than twice as often as expected under moderate to extreme censoring. Surprisingly, the Tobit test statistic performs better as censoring increases. For example, under moderate to extreme censoring, but not under mild censoring, the Tobit reject frequencies are at the

Table VI. Chi-Square and Reject Proportions for All Cases of Censoring^a

Percentage censoring	NTML	ADF	Tobit
0%			
Mean	3.263	3.485	3.242
Variance	7.022	8.602	9.582
Reject proportion	.074	.068	.084
20%			
Mean	3.172	3.736	3.244
Variance	6.385	8.831	7.022
Reject proportion	.052	.088	.062
40%			
Mean	5.060	3.821	3.040
Variance	16.971	10.410	6.700
Reject proportion	.196	.112	.050
60%			
Mean	9.696	4.115	3.124
Variance	54.555	10.070	10.468
Reject proportion	.498	.112	.050

^a Degrees of freedom = 3. $N = 100$ MZ pairs and 100 DZ pairs.

expected 5% level. The variances of the Tobit test statistic, however, are invariably too large.

Table VII reports the average parameter estimates and average bias for the four censoring conditions at the small sample size. Inspection of these data reveal that under all censoring conditions the parameter estimates are negatively biased. The size of the bias is comparable to that found at the large sample size. Standardized parameter values are reported in Table VIII. Note that as censoring increases, the NTML and ADF estimates become increasing discrepant from the true values. For example, the NTML and ADF estimates of c^2 , under extreme censoring, are only half as large as the population value of this parameter. The Tobit estimates, however, are close to their theoretical values under all conditions. Thus it appears that the Tobit estimator can be used to obtain accurate parameter estimates with censored twin data, even in relatively small samples. These findings do not, however, support the use of Tobit theoretical standard errors, which, as we have seen, are misleading even in large samples.

Table IX compares the small sample, theoretical and empirical standard errors for the four censoring conditions. These data confirm our suspicion that the theoretical standard errors will be much larger than the observed variation of the parameter estimates. This is especially true for

Table VII. Estimates of Genetic and Environmental Paths and Percentage Bias for Four Censoring Conditions^a

Parameter	Estimator											
	NTML						ADF					
	0	20	40	60	0	20	40	60	0	20	40	60
<i>e</i> .632	.625 (-1.1)	.537 (-15.0)	.450 (-28.8)	.348 (-44.9)	.622 (-1.6)	.533 (-15.7)	.444 (-29.7)	.336 (-46.8)	.599 (-5.2)	.592 (-6.3)	.590 (-6.6)	.554 (-12.3)
<i>h</i> .707	.667 (-5.6)	.546 (-22.8)	.418 (-40.9)	.294 (-58.4)	.652 (-7.8)	.533 (-24.6)	.398 (-43.7)	.270 (-61.8)	.657 (-7.1)	.643 (-9.0)	.616 (-12.9)	.597 (-15.5)
<i>c</i> .316	.286 (-9.5)	.220 (-30.4)	.116 (-63.3)	.105 (-66.8)	.278 (-12.0)	.215 (-32.0)	.164 (-48.1)	.095 (-69.9)	.273 (-13.6)	.277 (-12.3)	.294 (-7.0)	.276 (-12.6)

^a *N* = 100 MZ pairs and 100 DZ pairs.

Table VIII. Estimates of h^2 , c^2 , and e^2 and Average Percentage Bias for Four Censoring Conditions at the Small Sample Size^a

Parameter	Estimator (% censoring)											
	NTML				ADF				Tobit			
	0	20	40	60	0	20	40	60	0	20	40	60
e^2 .400	.426 (6.5)	.454 (13.5)	.499 (24.8)	.554 (38.5)	.435 (8.8)	.463 (15.8)	.516 (29.0)	.579 (44.8)	.415 (3.8)	.417 (4.3)	.428 (7.0)	.415 (3.8)
h^2 .500	.485 (-10.4)	.470 (-6.0)	.432 (-13.6)	.396 (-20.8)	.478 (-4.4)	.461 (-7.8)	.414 (-17.2)	.374 (-25.2)	.499 (-.2)	.491 (-1.8)	.466 (-6.8)	.482 (-3.6)
c^2 .100	.089 (-11.0)	.077 (-23.0)	.068 (-32.0)	.050 (-50.0)	.087 (-13.0)	.076 (-24.0)	.070 (-30.0)	.046 (-54.0)	.086 (-14.0)	.092 (-8.0)	.106 (-6.0)	.103 (3.0)

^a $N = 100$ MZ pairs and 100 DZ pairs.

Table IX. Theoretical and Empirical Sampling Variability of Three Estimators Under Four Censoring Conditions in Small Samples^a

Percentage censoring	NTML	ADF	Tobit
0%			
<i>e</i>	.044 ^b	.041	.066
	.044 ^c	.046	.067
<i>h</i>	12.179	282.788	111.782
	.144	.148	.174
<i>c</i>	4637.005	10,411.059	7527.523
	.226	.232	.240
20%			
<i>e</i>	.037	.039	.076
	.040	.041	.074
<i>h</i>	35.359	32.086	785.506
	.119	.124	.193
<i>c</i>	4049.517	7320.554	9995.920
	.186	.185	.240
40%			
<i>e</i>	.031	.036	5.907
	.038	.040	.094
<i>h</i>	119.609	518.474	408.253
	.121	.121	.235
<i>c</i>	12910.780	13598.697	5117.869
	.158	.156	.257
60%			
<i>e</i>	.024	.033	510.230
	.038	.039	.133
<i>h</i>	291.253	333.980	1712.554
	.102	.101	.262
<i>c</i>	7179.109	10162.640	5997.807
	.116	.111	.268

^a $N = 100$ MZ pairs and 100 DZ pairs.^b Mean of estimated standard errors.^c Empirical standard deviation of estimates.

c, and less so for *h*. Even the Tobit standard errors are larger than the parameter standard deviations.

DISCUSSION

Monte Carlo simulations were used to demonstrate that genetic and environmental parameter values obtained from censored twin data will be biased if normal-theory methods are used for parameter estimation. The direction and magnitude of the bias in the additive genetic and nonshared environmental parameters are a function of the relative sizes of the MZ and DZ correlations. For c^2 , the shared-familial environmental

parameter, bias is always negative and often extreme with censored data. Under moderate censoring c^2 is sufficiently biased that researchers are apt to conclude erroneously that it is not significantly distinguishable from zero—even when there are relatively strong shared-familial environmental effects operating on the behavior.

An alternative method called genetic Tobit factor analysis was proposed for the analysis of censored twin data. The method was shown to be a logical extension of Muthén's (1989a) Tobit factor analysis model. We explored the behavior of GTFA parameter estimates using Monte Carlo simulations. The basic finding of these simulations is that, with censored data, the Tobit parameter estimates are significantly less biased than either normal-theory maximum-likelihood or asymptotic distribution free estimates. Under all conditions of censoring, the empirical distribution of the Tobit test statistic closely follows the χ^2 distribution, as predicted by theory. Parameter standard errors, however, are inaccurate regardless of which estimator is used.

The findings of this study point to the need for changing our reporting habits in the field of behavior genetics. It is not sufficient to report only the first- and second-order moments of the data; third- and higher-order moments are often equally important in determining model outcome.

APPENDIX: MODEL SETUP FOR FIG. 1

```

TI UNIVARIATE GENETIC TOBIT FACTOR ANALYSIS
DA NO=310 IY=2 IX=1 NGR=2 VT=OT MA=CM
TE
-1 -1
CE
0 0
LA
'TWIN1' 'TWIN2' 'AGE'
MO MO=SE P2 P3 NE=8 LY=FI PS=FI TE=FI BE=FI GA=FI
VA 1.0 LY(1,1) LY(2,5)
VA 1.0 PS(2,2) PS(3,3) PS(4,4) PS(6,6) PS(7,7) PS(8,8)
VA 1.0 PS(7,3) PS(8,4)
VA 0.0 TE(1,1)-TE(2,2)
FR BE(1,2) BE(1,3) BE(1,4) BE(5,6) BE(5,7) BE(5,8)
VA .5 BE(1,2) BE(1,3) BE(1,4) BE(5,6) BE(5,7) BE(5,8)
EQ BE(1,2) BE(5,6)
EQ BE(1,3) BE(5,7)
EQ BE(1,4) BE(5,8)
FR GA(1,1) GA(5,1)
EQ GA(1,1) GA(5,1)
LL
'T1' 'E1' 'A1' 'C1' 'T2' 'E2' 'A2' 'C2'
OU WF ALL TO
RA FO UN='MZdata'

```


(Data format for group 1 goes here)
 TI GTFA
 DATA NO=244 IY=2 IX=1 NG=2 VT=OT MA=CM
 MO MO=SE P2 P3 NE=8 LY=FI PS=FI TE=FI GA=FI
 VA 1.0 LY(1,1) LY(2,5)
 VA 1.0 PS(2,2) PS(3,3) PS(4,4) PS(6,6) PS(7,7) PS(8,8)
 VA .5 PS(7,3)
 VA 1.0 PS (8,4)
 VA 0.0 TE(1,1)-TE(2,2)
 FR BE(1,2) BE(1,3) BE(1,4) BE(5,6) BE(5,7) BE(5,8)
 VA .5 BE(1,2) BE(1,3) BE(1,4) BE(5,6) BE(5,7) BE(5,8)
 EQ BE(1,1,2) BE(2,1,2) BE(2,5,6)
 EQ BE(1,1,3) BE(2,1,3) BE(2,5,7)
 EQ BE(1,1,4) BE(2,1,4) BE(2,5,8)
 FR GA(1,1) GA(5,1)
 EQ GA(1,1,1) GA(2,1,1) GA(2,5,1)
 LL
 'T1' 'E1' 'A1' 'C1' 'T2' 'E2' 'A2' 'C2'
 OU WF ALL TO
 RA FO UN='DZdata'
 (Data format for group 2 goes here)

ACKNOWLEDGMENTS

This research was conducted as part of a Ph.D. dissertation by the first author completed at the University of Minnesota. The authors thank Drs. Auke Tellegen, Robert Cudeck, David T. Lykken, Lon Cardon, and Joanne Meyer for comments on previous versions. Special thanks go to Ginger Nelson for help in preparing Table I.

REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Auto. Control* AC 19:716-723.
- Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972). *Robust Estimates of Location Survey and Advances*, Princeton University Press, Princeton, NJ.
- Ansell, M. J. G. (1973). Robustness of location estimators to asymmetry. *Appl. Stat.* 22:249-254.
- Bentler, P. M. (1989). *EQS, a Structural Equations Program*, BMDP Statistical Software, Los Angeles.
- Boomsma, A. (1983). *On the Robustness of LISREL (Maximum Likelihood Estimation) Against Small Sample Size and Nonnormality*, Unpublished doctoral dissertation, University of Groningen, Groningen, The Netherlands.
- Boomsma, D. I., Martin, N. G., and Neale, M. C. (1989). Genetic analysis of twin and family data: Structural modeling using LISREL. *Behav. Genet.* 19:whole issue.
- Bradley, J. W. (1980). Nonrobustness in z, t, and F tests at large sample sizes. *Bull. Psychonom. Soc.* 16:333-336.
- Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *So. Afr. Stat. J.* 8:1-24.
- Browne, M. W. (1982). Covariance structures. In Hawkins, D. M. (ed.), *Topics in*

- Applied Multivariate Analysis*, Cambridge University Press, Cambridge, pp. 72–141.
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *Br. J. Mat. Stat. Psychol.* **37**:62–83.
- Browne, M. W. (1987). Robustness of statistical inference in factor analysis and related models. *Biometrika* **74**(2):375–384.
- Cohen, A. C. (1950). Estimating the means and variance of normal populations from singly and doubly truncated samples. *Ann. Mat. Stat.* **21**:557–569.
- Cohen, A. C. (1955). Restriction and selection in samples from bivariate normal distributions. *J. Am. Stat. Assoc.* **50**:884–893.
- Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychol. Bull.* **105**:317–327.
- Cudeck, R., and Browne, M. W. (1983). Cross-validation of covariance structures. *Multivar. Behav. Res.* **18**:147–167.
- Des Raj (1953). On estimating the parameters of bivariate normal populations from double and singly linearly truncated samples. *Sankhya* **12**:277–290.
- Divgi, D. R. (1979). Calculation of the tetrachoric correlation coefficient. *Psychometrika* **44**:169–172.
- Eaves, L. J., Last, K. A., Young, P. A., and Martin, N. G. (1978). Model-fitting approaches to the analysis of human behavior. *Heredity* **41**:249–320.
- Eaves, L. J., Eysenck, H. J., and Martin, N. G. (1989). *Genes, Culture and Personality*, Academic Press, London.
- Falconer, D. S. (1981). *Introduction to Quantitative Genetics*, 2nd ed., Longman, London.
- Floderus-Myrhed, B., Pedersen, N., and Rasmuson, I. (1980). Assessment of heritability for personality, based on a short-form of the Eysenck Personality Inventory: A study of 12,898 twin pairs. *Behav. Genet.* **10**:153–162.
- Geweke, J. F., and Singleton, K. J. (1980). Interpreting the likelihood ratio statistic in factor models when sample size is small. *J. Am. Stat. Assoc.* **75**:133–137.
- Grayson, D. A. (1990). The matter of scale. *Behav. Genet.* **20**:81–89.
- Gupta, A. K. (1952). Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika* **39**:260–273.
- Harlow, L. (1985). *Behavior of Some Elliptical Theory Estimators with Non-normal Data in a Covariance Structures Framework: A Monte Carlo Study*, Unpublished doctoral dissertation, University of California, Los Angeles.
- Heath, A. C., Neale, M. C., Hewitt, J. K., Eaves, L. J., and Fulker, D. W. (1989). Testing structural equation models for twin data using LISREL. *Behav. Genet.* **19**:9–35.
- Hill, M., and Dixon, W. J. (1982). Robustness in real life: A study of clinical laboratory data. *Biometrics* **38**:377–396.
- Hogg, R. V., and Tanis, E. A. (1983). *Probability and Statistical Inference*, Macmillan, New York.
- Jinks, J. L., and Fulker, D. W. (1970). Comparison of the biometrical, genetical, MAVA, and classical approaches to the analysis of human behavior. *Psychol. Bull.* **73**:311–349.
- Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In Goldberger, A. S., and Duncan, O. D. (eds.), *Structural Equation Models in the Social Sciences*, Seminar Press, New York, pp. 85–112.
- Jöreskog, K. G., and Sörbom, D. (1979). *Advances in Factor Analysis and Structural Equation Models*, Abt Books, Cambridge, MA.
- Jöreskog, K. G., and Sörbom, D. (1988). *LISREL VII: Analysis of Linear Structural Relationships*, Scientific Software, Mooresville, IN.
- Maddala, G. S. (1983). *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.

- Mather, K. (1949). *Biometrical Genetics: The Study of Continuous Variation*, Dover Publications Inc., Dover.
- Mather, K., and Jinks, J. L. (1982). *Biometrical Genetics*, 3rd ed., Chapman and Hall, London.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* 49:115–132.
- Muthén, B. (1987a). *LISCOMP. Analysis of Linear Structural Equations with a Comprehensive Measurement Model*, User's Guide, Scientific Software, Mooresville, IN.
- Muthén, B. (1987b). Multiple-group structural modeling with non-normal continuous variables. Presented at the American Statistical Association meeting in Philadelphia, PA (1984).
- Muthén, B. (1989a). Tobit factor analysis. *Br. J. Math. Stat. Psychol.* 42:241–250.
- Muthén, B. (1989b). Latent variable modeling in heterogeneous populations. *Psychometrika* 54:557–585.
- Muthén, B. (1990). Moments of the censored and truncated bivariate normal distribution. *Br. J. Math. Stat. Psychol.* 43:131–143.
- Muthén, B., and Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *Br. J. Math. Stat. Psychol.* 38:171–189.
- Neale, M. C. (1991). *MX Statistical Modeling*, Unpublished program manual.
- Neale, M. C., and Martin, N. G. (1989). The effects of age, sex, and genotype on self-report drunkenness following a challenge dose of alcohol. *Behav. Genet.* 19:63–78.
- Neale, M. C., Heath, A. C., Hewitt, J. K., Eaves, L. J., and Fulker, D. W. (1989). Fitting genetic models with LISREL: Hypothesis testing. *Behav. Genet.* 19:37–49.
- Olsson, U. (1979a). On the robustness of factor analysis against crude classification of the observations. *Multivar. Behav. Res.* 14:485–500.
- Olsson, U. (1979b). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika* 44:443–460.
- Olsson, U., Drasgow, F., and Dorans, N. J. (1982). The polyserial correlation coefficient. *Psychometrika* 47:337–347.
- Peters, C. C., and Van Voorhis, W. R. (1940). *Statistical Procedures and Their Mathematical Bases* McGraw-Hill, New York.
- Plomin, R., and Daniels, D. (1987). Why are children in the same family so different from one another? *Behav. Brain Sci.* 10:1–59.
- Schwarz, G. (1978). Estimating the dimensions of a model. *Ann. Stat.* 6:461–464.
- Tanaka, J. S. (1984). *Some Results on the Estimation of Covariance Structure Models*, PhD. thesis, University of California, Los Angeles.
- Tanaka, J. S. (1987). “How big is big enough?”: Sample size and goodness of fit in structural equation models with latent variables. *Child Dev.* 58:134–146.
- Tellegen, A., Lykken, D. T., Bouchard, T. J., Jr., Wilcox, K. J., Segal, N. L., and Rich, S. (1988). Personality similarity in twins reared apart and together. *J. Person. Soc. Psychol.* 54:1031–1039.
- Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometric* 26:24–36.
- Tukey, J. W., and McLaughlin, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization. *Indian J. Stat.* 25:331–351.
- Wainer, H., and Thissen, D. (1976). Three steps toward robust regression. *Psychometrika* 41:9–34.
- Waller, N. G., Kojetin, B. A., Bouchard, T. J., Lykken, D. T., and Tellegen, A. (1989). Genetic and environmental influences on religious interests, attitudes, and values: A study of twins reared apart and together. *Psychol. Sci.* 1:138–142.

- Waller, N. G., Lykken, D. T., and Tellegen, A. (1990). occupational interests, leisure time interests and personality: Three domains or one? Findings from the Minnesota Twin Registry. In Dawis, R. V., and Lubinski, D. (eds.), *Assessing Individual Differences in Human Behavior: New Methods, Concepts and Findings*, University of Minnesota Press, Minneapolis.

Edited by N. G. Martin