Congeneric Modeling of Reliability Using Censored Variables

R. L. Brown
University of Wisconsin

This paper explores the use of Jöreskog's (1970) congeneric modeling approach to reliability using censored quantitative variables, and discusses the compound problem of non-normality and attenuation that occurs when estimating censored continuous variables. Two monte carlo studies were conducted. The first study demonstrated the inappropriateness of using normal theory generalized least-squares (NTGLS) for estimating reliability on censored variables. The second study compared three different estimation procedures —NTGLS, asymptotically distribution free (ADF) estimators, and latent TOBIT estimators—as to their efficiency in estimating individual and composite reliability on censored variables. Results from the studies indicate that problems of non-normality and attenuation must be addressed before accurate reliability estimates may be obtained.   *Index terms: censored variables, congeneric model, covariance modeling, monte carlo study, reliability, TOBIT correlations.*

Of the various procedures that have been proposed to estimate reliability, many have been based on an analysis-of-variance (ANOVA) approach (Hoyt, 1941). These include intraclass correlation and generalizability theory (Cronbach, Rajaratnam, & Gleser, 1963). Though these procedures are frequently used, they require accepting a number of assumptions (Saal, Downey, & Lahey, 1980). One major assumption is that all the measures must have

the same unit of measurement, which is sometimes referred to as "essentially tau equivalent" (Lord & Novick, 1968). In response to this assumption, a covariance modeling procedure that is based on Jöreskog's (1970) general model for the analysis of covariance structures has been proposed (van der Kamp & Mellenbergh, 1976; Werts, Linn, & Jöreskog, 1974).

Although the covariance modeling procedure provides details about the assumptions surrounding the use of ANOVA-based reliability measures, it is limited. One major assumption—and subsequent limitation—is that the observed variables that are provided in the sampled covariance matrix (S) accurately represent variables from a multivariate normally distributed population, if normal theory-based estimators are to be used (e.g., maximum likelihood or generalized least squares). The appropriateness of using procedures based on normal theory for estimates is questionable when analyzing a covariance matrix based on "censored variables" (variables with a high concentration of cases at either or both ends of the distribution; Hald, 1949), due to the compound problem of non-normality and attenuated covariances.

This paper explores the use of Jöreskog's (1970) congeneric modeling approach to reliability using censored quantitative variables, and provides a possible solution to the compound problem of non-normality and attenuated covariances brought about by censored variables.

Congeneric Modeling Approach to Reliability

The ANOVA approach to reliability generally considers a two-way ANOVA model (Werts et al., 1974), in which the reliability of the ith measure (item or test) may be estimated, assuming similar units of measurement, by

$$\hat{\rho}_i = \frac{MS_b - MS_r}{MS_b + (p-1)MS_r} \quad , \tag{1}$$

where MS_b refers to the mean square between measures and MS_r is the interaction (residual) mean square term, over p measures. An estimate of composite reliability may be obtained by

$$\hat{\rho}_c = \frac{MS_b - MS_r}{MS_b} \quad . \tag{2}$$

The congeneric approach to reliability is based on covariance modeling of a data matrix $Y(N \times k)$, denoting scores for a randomly selected group, where N is the number of observations over k measures. It is assumed that the rows of Y are independently distributed, each having a multivariate normal distribution. The congeneric model for a four-indicator reliability model may be written as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} \eta + \begin{bmatrix} \theta_{\varepsilon_1} \\ \theta_{\varepsilon_2} \\ \theta_{\varepsilon_3} \\ \theta_{\varepsilon_4} \end{bmatrix} = \Lambda\eta + \Theta_\varepsilon \quad , \tag{3}$$

where λ and θ_ε are parameters to be estimated for each y, with η representing the true component and Θ_ε the error component. Assuming uncorrelated measurement errors, with variables that have a mean of 0 and a factor variance of 1, the variance-covariance matrix for this model may be defined as

$$\Sigma = \begin{bmatrix} \lambda_1^2 + \theta_{\varepsilon_1} & & & \\ \lambda_1\lambda_2 & \lambda_2^2 + \theta_{\varepsilon_2} & & \\ \lambda_1\lambda_3 & \lambda_2\lambda_3 & \lambda_3^2 + \theta_{\varepsilon_3} & \\ \lambda_1\lambda_4 & \lambda_2\lambda_4 & \lambda_3\lambda_4 & \lambda_4^2 + \theta_{\varepsilon_4} \end{bmatrix} \tag{4}$$

which may be rewritten as

$$\Sigma = \Lambda\Lambda' + \Theta_\varepsilon \quad , \tag{5}$$

where Θ_ε is a $k \times k$ diagonal matrix of error variances, and Λ is a vector of loadings on a single common factor. Using this model, it is possible to test the initial assumption that all measures have the same underlying true score η. For example, if four items are assessed for reliability, the Λ loading matrix and the error variances θ_ε are estimated. For convenience, as previously mentioned, the variance of the true score may be standardized $[\sigma^2(\eta) = 1]$. This approach provides two over-identifying restrictions in the four-variable model (two degrees of freedom), allowing the null hypothesis in Equation 5 to be tested.

If the differences are considered small enough to be regarded as sampling fluctuations, based on the chi-square distribution, item reliability can be estimated as

$$\hat{\rho}_i = \frac{\hat{\lambda}_i^2}{\hat{\lambda}_i^2 + \hat{\theta}_{\varepsilon_i}} \tag{6}$$

with composite reliability estimated as

$$\hat{\rho}_c = \frac{(\Sigma\hat{\lambda}_i)^2}{(\Sigma\hat{\lambda}_i)^2 + \Sigma\hat{\theta}_{\varepsilon_i}} \quad . \tag{7}$$

Subsets of the Congeneric Model

If the assumption of tau equivalence is not met in computing the reliability of a set of measurements, then averaging scores would not be meaningful, and using ANOVA-based procedures would not be appropriate. Jöreskog (1970) provided a procedure based on a subset of the congeneric model to test for essential tau equivalence. This assumes that measures are equivalent if the regression weights λ_i are equal in Λ. The essentially tau-equivalent model tests $H_0: \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$ with $.5k(k + 1) - (k + 1)$ degrees of freedom. A total of $k + 1$ parameters would then be estimated. If the essentially tau equivalent hypothesis is rejected, then ANOVA procedures should be rejected for estimating reliability and/or generalizability coefficients (Werts et al., 1974). Green, Lissitz, and Mulaik (1977) have indicated that coefficient alpha as an estimate of composite reliability may be robust to this violation when k is large. If the hypothesis is not rejected, the parameter estimates may be used to estimate item reliability,

$$\hat{\rho}_i = \frac{\hat{\lambda}^2}{\hat{\lambda}^2 + \hat{\theta}_{\varepsilon_i}} \quad , \tag{8}$$

and composite reliability,

$$\hat{\rho}_c = \frac{(k\hat{\lambda})^2}{(k\hat{\lambda})^2 + \Sigma\hat{\theta}_{\varepsilon i}} \qquad . \tag{9}$$

In addition, the reliabilities estimated from Equation 8 may vary depending on error variances θ_ε. Because Equation 1 assumes that all measures have the same reliability, Werts et al. (1974) suggested testing for equal error variances. This may be tested using another subset of the congeneric model, sometimes referred to as a "parallel model" (Gulliksen, 1968; Jöreskog, 1970). This model implies equality constraints on both the λ_is and the $\theta_{\varepsilon i}$; thus the hypotheses tested are H_{01}: $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$ and H_{02}: $\theta_{\varepsilon_1} = \theta_{\varepsilon_2} = \theta_{\varepsilon_3} = \theta_{\varepsilon_4}$ with $.5k(k + 1) - 2$ degrees of freedom. This model will estimate two parameters. If the essentially tau equivalent hypothesis (H_{01}) is not rejected, but the equality of error variances (H_{02}) is rejected, then ANOVA-based procedures may be used to estimate composite reliability but not individual item reliability. Werts et al. (1974) asserted that if the parallel model is rejected (H_{01} and H_{02}), use of ANOVA-based procedures would underestimate the composite reliability. If the parallel model is not rejected, the reliabilities can be estimated using either ANOVA-based procedures or by

$$\hat{\rho}_i = \frac{\hat{\lambda}^2}{\hat{\lambda}^2 + \hat{\theta}_\varepsilon} \tag{10}$$

with composite reliability as

$$\hat{\rho}_c = \frac{(k\hat{\lambda})^2}{(k\hat{\lambda})^2 + k\hat{\theta}_\varepsilon} \qquad . \tag{11}$$

Tests of these hypotheses amount to tests of the validity of specific assumptions about the model; these tests are not provided in the ANOVA-based reliability estimates (Werts et al., 1974). Although the Jöreskog (1970) procedure seems attractive for answering questions regarding measurement unit equivalency and stability of error variances in reliability models, the procedure has limitations of its own.

Congeneric Model Limitations

Testing H_{01} and H_{02} using covariance modeling depends on a number of conditions. First, sample size is crucial. Monte carlo studies based on normal theory distribution sampling have indicated that when $N < 50$, normal theory maximum likelihood (NTML) and normal theory generalized least-squares (NTGLS) estimation procedures result in model convergence problems, while sample sizes of $N = 25$ result in improper solutions (Anderson & Gerbing, 1984; Boomsma, 1982, 1983, 1985; Gerbing & Anderson, 1987). Minimum sample size recommendations for unbiased estimates are typically $N \geqslant 100$.

Second, distributions with aberrant skewness and kurtosis values seem to result in higher than usual model rejections (Boomsma, 1983) and poorer parameter estimates. This paper concentrates on this second condition.

Censored Continuous Variables

One source of aberrant skewness and kurtosis values in sample variables has developed from analyzing sample truncated normal distribution variables. Hald (1949) has called such variables censored variables, because the population from which the sample variable was drawn is considered normally distributed, and the sample being considered is incomplete. More formally, a continuous censored variable may be defined as:

$$y = c_1 \quad \text{if} \quad y^* \leqslant c_1$$
$$y = y^* \quad \text{if} \quad c_1 < y^* < c_u$$
$$y = c_u \quad \text{if} \quad y^* \geqslant c_u \quad , \tag{12}$$

where c_1 and c_u represent known lower and upper censoring constants, which may take on values $-\infty$ to $+\infty$, with y the sampled value and y^* the population value. Censored variables may thus be considered variables that have limited variability with a large proportion of cases occurring at one or both endpoints of the scale. An example of a censored sample variable would be a test score which has a "floor" or a "ceiling" that produces a large proportion of cases based on an extreme response (e.g., all correct).

Muthén (1985, 1987a) has shown that increases in censoring have a substantial effect on attenuating the correlation/covariance estimates, which will underestimate the reliability when using procedures

based on typical normal theory. This leads to a compound problem of non-normality brought about by attenuated measures of covariance; as Hollis and Muthén (1987) have indicated, both problems must be addressed to provide correct estimates.

Monte Carlo Studies

To demonstrate the inappropriateness of using normal theory estimation procedures while disregarding the attenuation problem brought about by the censored variables, a simulation study was conducted using the NTGLS estimator.

Study 1 Design

In this study NTGLS estimates were obtained from the computer program LISCOMP, Version 1.1 (Muthén, 1987b). A standardized population covariance matrix Σ was created for a four-variable single-factor congeneric model using Equation 5. This model was selected purely for computational convenience. Two separate single-factor model population covariance matrices were created based on the design in Figure 1.

The first matrix (Σ_1) was based on high reliability ($\rho_i = .81$) with $\sigma^2(\eta) = 1$ and $\Lambda_1 = [.9,.9,.9,.9]$.

The second matrix (Σ_2), based on low reliability ($\rho_i = .36$), was established using $\sigma^2(\eta) = 1$ and $\Lambda_2 = [.6,.6,.6,.6]$. Variances in the errors of measurement (uniquenesses) were defined for both models as $\theta_\varepsilon = 1 - \lambda^2$. According to Equation 6, item reliability for the congeneric model in this example may be reduced to λ^2. Each Σ_i matrix was then used to generate 100 above-censored sample covariance matrices (S) for each sample size condition ($N = 25, 50, 100, 400$). Above-censoring was accomplished by assigning all normal distribution values above a specific percentage value to the data value at that percentage. For example, a 50% above-censoring distribution would consist of transforming the upper half of the distribution to the value at the 50th percentile. A mixed censored-variable design was then used with the first two variables normally distributed and the last two variables censored from above at 0% (no censoring), 25%, 50%, and 75%.

Study 2 Design

To demonstrate the influence of the attenuated covariance matrix resulting from the censored variables, Browne's (1982) asymptotically distribution free (ADF) generalized least-squares estimates

Figure 1
Mixed Censored-Variable Design for the Monte Carlo Study
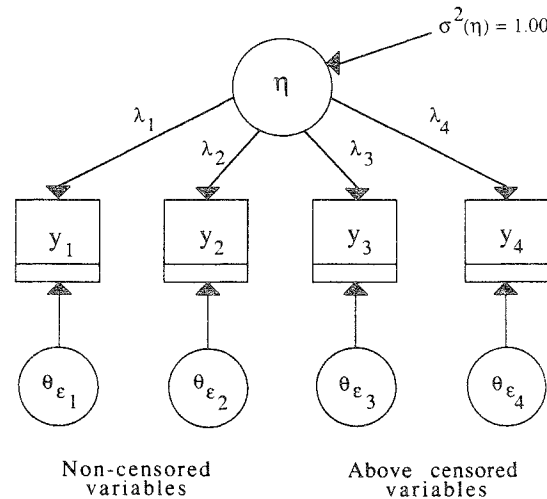
were obtained using LISCOMP. It was hypothesized that the ADF procedure would not be hampered by non-normality, but would still perform poorly due to attenuation. LISCOMP was also used to obtain NTGLS estimates. As a final solution to the compounded problem brought about by the censored variables, Muthén's (1985, 1987a) latent TOBIT estimates were obtained using LISCOMP and were compared to the other estimators. [Based on the rubric of LOGIT and PROBIT analyses, the term TOBIT was coined by Tobin (1958) in referring to his regression approach to censored variables, and is used here in similar context.]

A single-factor congeneric model [$\rho_i = .81$, $\lambda_i = .9$, with $\theta_\varepsilon = 1 - \lambda^2$ and $\sigma^2(\eta) = 1$] with $N = 100$ was considered for the comparison analysis using the mixed censored-variable design (Figure 1). One hundred above-censored sample covariance matrices (S) were generated using the same procedures defined in Study 1. Again, the selection of the single-factor four-variable model was based solely on computational convenience.

Estimation Procedures

Estimates are represented by the general family of fit functions for analysis of covariance structures, as proposed by Browne (1977). The NTGLS procedure may be defined as the fit function

$$F_{GLS} = (s - \sigma)'W^{-1}(s - \sigma) \quad , \qquad (13)$$

where s and σ refer to y based on a Pearson correlation/covariance with a weight matrix $W = \Sigma \times \Sigma$ or $S \times S$, where $\Sigma = \Lambda\Phi\Lambda' + \Theta_\varepsilon$ and $\sigma = \text{Vec}[\Sigma]$ (Jöreskog & Goldberger, 1972).

The fitting function for Browne's (1982) ADF generalized least-squares estimator is

$$F_{ADF} = (s - \sigma)'W^{-1}(s - \sigma) \quad , \qquad (14)$$

where

$$W = s_{ijkl} - s_{ij}s_{kl} \quad , \qquad (15)$$

where s_{ij} and s_{kl} are biased sample covariances, and s_{ijkl} is the fourth-order multivariate cumulant (Kendall & Stuart, 1977).

Muthén's (1985, 1987a) latent TOBIT estimator has the same fitting function as NTGLS, except that instead of y referring to a sample-based Pearson

correlation/covariance, y^* refers to a latent TOBIT correlation/covariance, based on the underlying assumption of individual univariate normality. The TOBIT fitting function is

$$F_{TOBIT} = (s^* - \sigma^*)'W^{-1}(s^* - \sigma^*) \quad , \qquad (16)$$

where s^* and σ^* refer to y^* TOBIT-based correlation/covariances with a weight matrix of $\Sigma^* \times \Sigma^*$ or $S^* \times S^*$, where $\Sigma^* = \Lambda\Phi\Lambda' + \Theta_\varepsilon$ and $\sigma^* = \text{Vec}[\Sigma^*]$ (Muthén, 1985; Muthén & Kaplan, 1985). The TOBIT approach is a two-stage estimation of s^*. The first stage is the univariate estimation of μ and σ^2 based on a normal distribution for censored variables (Gupta, 1952). The second stage estimates the covariances from the bivariate information using maximum likelihood, by holding μ and σ^2 constant at the univariate estimated levels (Muthén, 1985).

Results

The results of Study 1 support previous findings (Anderson & Gerbing, 1984; Boomsma, 1985) that the nonconvergence rate is inversely related to sample size at 0% censoring. Table 1 shows the percent of nonconvergence for the NTGLS estimates for the single-factor model over two levels of model reliability, four levels of sample size, and four levels of variable censoring.

The asymmetry provided in the higher censoring conditions seems to enhance nonconvergence. This effect is greater in the high-reliability model than in the low-reliability model, with the exception of the small-sample ($N = 25$) condition. Table 2 indicates that variable censoring has a dramatic effect on the percent of congeneric model rejections ($\alpha \leq .05$), using NTGLS estimates. The percent of model rejections for small samples (e.g., $N = 25$ and $N = 50$) should be interpreted cautiously.

Table 3 presents the percent of bias in the estimates of individual and composite reliability over the studied conditions for the NTGLS estimates. In the larger samples studied ($N = 100$ and $N = 400$), the percent of bias in the above-censored variables (ρ_3 and ρ_4) relates positively to the amount of variable censoring, with stronger bias occurring in the low-reliability model ($\rho = .36$). The small-sample

Table 1
Percent of Nonconvergent Solutions for NTGLS Estimates
for Low and High Population Reliability Models
(N = 25, 50, 100, 400)

Censoring Level	Population Model Reliability							
	Low (ρ = .36)				High (ρ = .81)			
	25	50	100	400	25	50	100	400
0%	12	0	0	0	13	7	2	1
25%	14	2	0	0	8	12	7	5
50%	17	3	1	0	12	13	6	7
75%	29	16	2	1	36	37	24	19

low-reliability model conditions (ρ = .36, N = 25 and N = 50) produced erratic results, as expected.

The percent of bias in the noncensored continuous variables (ρ_1 and ρ_2) may be considered trivial, with the exception of the small-sample low-reliability model conditions (ρ = .36, N = 25 and N = 50). The bias in ρ_1 and ρ_2 for the high-reliability model tended to be negative, or underestimated when censoring was present, whereas the estimates for the low-reliability model tended to fluctuate more (Table 3). Composite reliability was also underestimated by censoring, but not with the magnitude of the individual values, due to the mixed censored design used in the study.

The results from Study 2 are provided in Table 4, which gives the percent of model rejections for the three estimation procedures (NTGLS, ADF, and TOBIT) across the four levels of above-censoring.

All three procedures similarly reject few congeneric models at 0% censoring, but as censoring increases (increasing the asymmetry in the distri-

butions, as well as the attenuation in the matrix), the percent of rejection increases in both NTGLS and ADF. This is to be expected with the NTGLS estimator. The ADF procedure does not do better because of the attenuation in the covariance matrix used in this mixed design.

Table 5 shows the percent of bias occurring in the estimates of individual and composite reliability. Again, the percent of bias for the censored variables strongly relates to the amount of above-censoring for both NTGLS and ADF estimation procedures. The TOBIT estimates, on the other hand, produced very little bias. The percent of bias occurring for noncensored variables was negligible for the TOBIT and NTGLS estimates, but resulted in slightly larger negative bias with the use of the ADF procedure as censoring of companion variables increased.

Table 5 also indicates a strong positive relationship between the level of above-censoring and the percent of bias in the estimate of composite reli-

Table 2
Percent of Model Rejections Adjusted for Nonconvergence
Based on $\alpha \leq .05$ for Low and High Reliability Models
(N = 25, 50, 100, 400)

Censoring Level	Population Model Reliability							
	Low (ρ = .36)				High (ρ = .81)			
	25	50	100	400	25	50	100	400
0%	0.0	4.0	0.0	7.0	4.6	3.2	2.0	5.1
25%	1.2	3.1	1.0	5.0	6.5	12.5	20.4	80.0
50%	1.2	5.2	3.0	13.0	12.5	46.0	85.1	100.0
75%	0.0	3.6	7.1	25.3	25.0	57.1	92.1	100.0

Table 3
Percent of Bias in NTGLS Estimates of Reliability
(N = 25, 50, 100, 400)

Censoring Level and Estimate[a]	Population Model Reliability							
	Low (ρ = .36)				High (ρ = .81)			
	25	50	100	400	25	50	100	400
0%								
ρ_1	19.89	9.71	−.12	0.00	.81	2.04	1.01	0.00
ρ_2	21.30	4.19	2.04	3.14	1.91	.60	0.00	0.00
ρ_3	−.27	−3.47	−2.30	−2.14	1.91	.17	−.43	0.00
ρ_4	15.51	8.56	5.25	0.00	.81	.60	.60	0.00
ρ_c	6.22	2.22	.59	.12	.39	.25	−.09	0.00
25%								
ρ_1	25.73	7.53	−.12	0.00	−.26	.60	−.43	−1.43
ρ_2	8.88	4.19	2.04	3.14	.81	−.43	−1.43	−1.43
ρ_3	−7.44	−10.66	−10.90	−9.64	−1.98	−3.64	−5.24	−4.60
ρ_4	14.97	2.12	−3.90	−6.94	−1.40	−3.02	−5.24	−4.60
ρ_c	4.80	.34	−1.69	−1.70	−.22	−.46	−.87	−.88
50%								
ρ_1	19.89	9.89	−.12	−2.14	−2.25	−.87	−1.88	−2.88
ρ_2	16.28	4.19	−1.14	0.00	−2.25	−1.32	−1.88	−2.41
ρ_3	−6.70	−19.35	−21.88	−20.58	−5.31	−9.37	−12.28	−11.22
ρ_4	5.08	−5.14	−14.43	−20.58	−8.36	−11.22	−11.22	−10.20
ρ_c	3.47	−1.80	−5.45	−6.03	−1.51	−1.56	−1.88	−1.97
75%								
ρ_1	−10.41	19.38	−2.30	−3.12	−1.64	−1.32	−2.35	−2.35
ρ_2	.85	−.27	−.12	0.00	−2.61	−.26	−2.35	−1.88
ρ_3	38.89	−39.02	−44.44	−43.38	−14.49	−24.72	−24.72	−22.44
ρ_4	17.68	−19.92	−32.48	−43.38	−20.28	−26.77	−22.44	−22.44
ρ_c	−6.97	−6.27	−11.90	−12.64	−2.85	−2.88	−3.30	−3.16

[a]Individual and composite reliability estimates.

ability for both NTGLS and ADF procedures. The percent of bias for the TOBIT estimates of composite reliability was trivial.

Discussion

Study 1 demonstrated the inappropriateness of using the NTGLS estimation procedure to retrieve correct reliability estimates and model fit indices (likelihood ratios) with small sample sizes, as indicated by the general lack of model convergence and the erratic estimates for the small-sample low-reliability model conditions (ρ = .36, N = 25 and N = 50). This may prove to limit the use of congeneric modeling techniques in reliability studies. Although not reported in this paper, increases in improper solutions (e.g., negative θ_ε estimates) due

to small sample sizes would also cause difficulty in estimating reliability. In general, the NTGLS estimations for both individual and composite reliability for above-censored variables were poor,

Table 4
Percent of Model Rejections
Based on $\alpha \leq$ .05

Censoring Level	Estimation Procedure		
	NTGLS[a]	ADF	TOBIT
0%	2.04	7.00	5.00
25%	20.43	26.00	4.00
50%	85.11	71.00	5.00
75%	92.11	58.00	3.00

[a]Percentages adjusted for nonconvergence.

Volume 13 Number 2 June 1989
158 *APPLIED PSYCHOLOGICAL MEASUREMENT*

Table 5
Percent of Bias in
Estimates of Reliability

Censoring Level and Estimate[a]	Estimation Procedure		
	NTGLS	ADF	TOBIT
0%			
ρ_1	1.01	1.01	-.56
ρ_2	0.00	1.01	-1.88
ρ_3	-.43	.60	-1.43
ρ_4	.60	.60	-1.96
ρ_c	-.09	.23	-.50
25%			
ρ_1	-.43	-1.78	-.56
ρ_2	-1.43	-1.78	-1.88
ρ_3	-5.24	-7.46	-2.41
ρ_4	-5.24	-5.89	-1.96
ρ_c	-.87	-1.11	-.50
50%			
ρ_1	-1.88	-4.27	-.56
ρ_2	-1.88	-3.12	-1.88
ρ_3	-12.28	-19.59	-1.43
ρ_4	-11.22	-19.59	-1.88
ρ_c	-1.88	-3.24	-.42
75%			
ρ_1	-2.35	-5.36	-.56
ρ_2	-2.35	-3.64	-1.88
ρ_3	-24.72	-48.22	-1.96
ρ_4	-22.44	-35.90	-4.34
ρ_c	-3.30	-5.22	-.64

[a]Individual and composite reliability estimates.

and can probably be attributed to problems of non-normality and attenuation in the covariance matrix.

Study 2 demonstrated that attempting to deal with the problem of non-normality alone (e.g., using the ADF procedure) was not enough to provide accurate reliability estimates. The attenuation in the correlations/covariances must also be considered. This study demonstrated the superiority of latent TOBIT estimates in dealing with these two problems. If underlying normality exists in censored variables, then latent correlation/covariance-based estimates of reliability, such as TOBIT estimates, may provide an effective estimation procedure for congeneric modeling of reliability, though sample size is still a limiting factor.

Although the TOBIT procedure has definite appeal for estimating reliability with censored vari-

ables, some model limitations exist. Heavy computational demands usually restrict the size of the weight matrix used in the estimation procedure; for example, estimation becomes impractical when the number of y variables exceeds 20 to 25.

References

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49,* 155–173.

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (Part 1, pp. 149–173). Amsterdam: North-Holland.

Boomsma, A. (1983). *On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality.* Unpublished doctoral dissertation, University of Groningen.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika, 50,* 229–242.

Browne, M. W. (1977). Generalized least squares estimates in the analysis of covariance structures. In D. J. Aigner & A. S. Goldberger (Eds.), *Latent variables in socio-economic models.* Amsterdam: North-Holland. [Reprint of article in *South African Statistical Journal, 8,* 1–24, 1974.]

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis.* Cambridge: Cambridge University Press.

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology, 16,* 137–163.

Gerbing, D. W., & Anderson, J. C. (1987). Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate respecifications. *Psychometrika, 52,* 99–111.

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement, 37,* 827–838.

Gulliksen, H. (1968). Methods for determining equivalence of measures. *Psychological Bulletin, 70,* 534–544.

Gupta, A. K. (1952). Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika, 39,* 260–273.

Hald, A. (1949). Maximum likelihood estimation of the parameters of a normal distribution which is truncated

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

at a known point. *Skandinavisk Aktuarietidskrift, 32*, 119.

Hollis, M., & Muthén, B. (1987). *Structural covariance models with categorical data: An illustration involving the measurement of political attitudes and belief systems.* Paper presented at the annual meeting of the American Political Science Association, Chicago.

Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika, 6,* 153–163.

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika, 57,* 239–251.

Jöreskog, K. G., & Goldberger, A. S. (1972). Factor analysis by generalized least squares. *Psychometrika, 37,* 243–260.

Kendall, M., & Stuart, A. (1977). *The advanced theory of statistics, Vol. 1* (4th ed.). New York: Macmillan.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Wesley.

Muthén, B. (1985, July). TOBIT *factor analysis.* Paper presented at the Fourth European Meeting of the Psychometric Society, Cambridge, England.

Muthén, B. (1987a, April). *Advances in factor analysis and structural equation modeling with categorical and other non-normal data.* Paper presented at the meeting of Advances in Factor Analysis, University of Wisconsin, Madison.

Muthén, B. (1987b). LISCOMP: *Analysis of linear structural relations using a comprehensive measurement model* [Computer program]. Mooresville IN: Scientific Software.

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38,* 171–189.

Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin, 88,* 413–428.

van der Kamp, L. J. T., & Mellenbergh, G. J. (1976). Agreement between raters. *Educational and Psychological Measurement, 36,* 311–317.

Werts, C. E., Linn, R. L., & Jöreskog, K. G. (1974). Intraclass reliability estimates: Testing structural assumptions. *Educational and Psychological Measurement, 34,* 25–33.

Acknowledgments

Author's Address

Send requests for reprints or further information to R. L. Brown, Department of Family Medicine and Practice, University of Wisconsin, Madison WI 53715, U.S.A. BITNET address: RLBROWN@WISCMACC.