

UCLA

Department of Statistics Papers

Title

On Measures of Explained Variance in Nonlinear Structural Equations Models

Permalink

<https://escholarship.org/uc/item/13q0q1vv>

Authors

Bentler, Peter M

Raykov, Tenko

Publication Date

1998-03-05

March 5, 1998

On Measures of Explained Variance
in Nonrecursive Structural Equation Models

Peter M. Bentler

University of California, Los Angeles

Tenko Raykov

University of Melbourne

Address correspondence regarding editorial and publication matters to Professor Peter M. Bentler at Franz Hall, Box 951563, UCLA, Los Angeles, CA 90095-1563. Bentler also can be reached via Phone: 310 825 2893, Fax: 310 206 4315, email: bentler@ucla.edu

Authors' Note: This research was carried out while T. Raykov was visiting the Department of Psychology of UCLA and supported by grant DA01070. Send correspondence relating to this manuscript to Peter M. Bentler, Franz Hall, Box 951563, UCLA, Los Angeles, CA 90095-1563.

On Measures of Explained Variance in Nonrecursive Structural Equation Models

Abstract

Following up and correcting prior work by Teel, Bearden, and Sharma (1986) in this journal, a general approach to variance explained in latent dependent variables of nonrecursive linear structural equation models is given. A new method of its estimation, easily implemented in EQS or LISREL, is described and illustrated.

Introduction

It is imperative in marketing and behavioral and management research to summarize the ability of the variables in a model to explain critical dependent variables. In single-equation regression, multivariate regression, and multiple-equation structural equation models, in which one or more dependent variables are predicted from a set of predictors, common practice is to summarize the predictability of a dependent variable with the squared multiple correlation $R^{\#}$. Jain (1994) provides a good introduction to regression analysis in marketing, and the $R^{\#}$ in this context. While alternative measures are available, this is a basic and convenient summary measure: it is generally recognized that better models have higher $R^{\#}$ values, especially when prediction is a key issue as it often is in marketing research. Of course, there may be technical issues in its application, such as adjustment for bias due to the size of the predictor set, which are beyond the current scope to discuss.

Similar $R^{\#}$ -like measures have been used to summarize predictability in latent variable structural equation models in which some of the predictors or dependent variables are unmeasured constructs. Although the presence of latent variables change some aspects of interpretation, it would seem that essentially no new principles should be involved in summarization of prediction. In particular, reported $R^{\#}$ values should describe the extent of predictability of any dependent variable, whether latent or observed. However, in a pioneering paper in this journal, Teel, Bearden and Sharma

(1986) pointed out that measurement of explained variance is not at all straightforward when multiple equation models contain feedback loops among variables or have disturbances that are correlated across equations. Models with reciprocal causation, often known as nonrecursive structural equation models, have been of interest to research in the social and behavioral sciences, including econometrics and marketing (e.g., Schaubroeck 1990; Hayduk 1996). A recent discussion of reciprocal causation can be found in Sharma (1996, Ch. 14), using the Shimp and Kavas (1984) paper on the theory of reasoned action applied to coupon usage.

Teel et al. described three alternative measures of explained variance for single equations in nonrecursive structural equation models, based on a LISREL (Jöreskog and Sörbom 1993) approach, a SAS econometric approach, and their own adjusted LISREL approach they consider equivalent to a predicted/observed $R^{\#}$ approach. In their examples, these different approaches gave radically different values for explained variance, and hence it is critical that the optimally justified choice of measure is adopted. The authors do not especially favor the LISREL approach, involving subtraction of residual variance from total variance in endogenous constructs. They show that this expression contains irrelevant extraneous variance. More recently, the LISREL approach also was questioned by Hayduk (1996), who noted "I used to think that no acceptable model could contain a concept having a negative $R^{\#}$, but I was wrong!" (pp. xx-xxi). Actually, it does seem correct that negative $R^{\#}$, as well as values larger than 1.0, are not meaningful, since neither values could be consistent with the usual interpretation of $R^{\#}$ as the proportion of variance in a dependent variable accounted for by the predictors. The econometric approach uses only exogenous variables as predictors, i.e., it ignores endogenous variable predictors of dependent variables, and thus does not correspond to the actual equations of the model. Hence it is not desirable either. Teel et al. conclude that a modified LISREL approach is best. They state "The square of the correlation between the observed and predicted values of a dependent variable in a single equation is

most consistent with the traditional definition of explained variance. This is also the recommended measure of explained variance for regression equations that are linear in the parameters and nonlinear in the variables...This seems to be the most desirable estimate for most marketing research applications" (p. 167). We agree with this conceptual assessment. Unfortunately, the authors' own $R^{\#}$ -like index reflecting variance accounted for seems inappropriate for nonrecursive models. It is narrow in definition, does not extend easily to any nonrecursive model, and is hard to compute in a complex model. In addition, implausible and unnecessary restrictions were imposed in its derivation. And finally, their examples give $R^{\#}$ values that are not consistent with their conceptual definition.

The present article has a three-fold goal. It clarifies a logical measure of proportion of explained variance in a latent dependent variable that in our view is more appropriate for nonrecursive and other linear structural equation models. A method for estimation of this measure is then presented, which can be routinely used with widely circulated structural modeling software, such as EQS (Bentler 1995) or LISREL (Jöreskog and Sörbom 1993). Also, the general formula of the $R^{\#}$ -like index by Teel et al. (1986) is presented.

A General Proposition

Consider an arbitrary structural equation such as
$$C = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon$$
 which is purposely written in the form of a standard regression equation. In this typical situation, the residual ϵ is independent of the predictor variables X_j , and the population squared multiple correlation R^2 can be defined as the squared product moment correlation $R^2 = \text{Corr}^2(C, \hat{C})$, where $\hat{C} = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ is the predicted value of C based on the known parameter values. In practice, one only has a sample from the population, say C and X_j , and an optimal parameter estimator $\hat{\beta}_j$ computed, for example, by least squares. Then the predicted value of C is $\hat{C} = \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k$, and we can define the sample squared multiple correlation as $R^{\#} = \text{Corr}^2(C, \hat{C})$

This is the definition for R^2 given, for example, by Jennrich (1995). It is sometimes called the coefficient of determination or multiple determination. Of course there are other definitions, such as the ratio of regression sum of squares to total sum of squares, that are in the standard linear model setup identical to the previous definition (see e.g., Fox 1997, Ryan 1997).

We propose to use the same squared-correlation definition for any structural equation in a linear structural equation model, whether the variables η and ξ are measured or latent variables, whether the model is a single-equation model or a multiple equation model, whether the residual ϵ is independent of the other predictors in the equation or not, whether residuals ϵ from different equations are independent across equations or not, whether or not the model is nonrecursive, and irrespective of the method of parameter estimation. This idea is consistent with that of Teel et al (1986). In addition to this abstract idea, and in contrast to Teel et al., we propose a simple and general way to permit estimation of $R^2 = \text{Corr}^2(\hat{C}, \hat{Y})$ for any model. To anticipate the idea, following Raykov (1997) the predicted variable \hat{C} is defined within the model structure by an additional equation involving a phantom latent variable, and the correlation of this latent variable with the dependent variable can be obtained as extra output from any structural modeling program. For example, in EQS a structural equation in which a latent variable F1 is predicted from other latent variables F2, F3... is of the form $F1 = \beta_2 F2 + \beta_3 F3 + \dots + D1$; where β is a coefficient to be estimated and D1 is the residual in the model. Then adding the dummy equation $F101 = F1 \cdot D1$, and asking the program to print the model-based correlations among all F's will give the correlation between $F101 = \beta_2 F2 + \beta_3 F3 + \dots$, the arbitrarily named predicted variable, and F1, the dependent variable. This gives the relevant multiple correlation, whose square is R^2 . Next we become more technical, and relate our approach to that of Teel et al.

Teel, Bearden, and Sharma's Index

Teel et al. (1986, p. 165) considered the following nonrecursive model displayed in Figure 1 (identical to their Figure 1)

Insert Figure 1 about here

Because of the restrictions they impose, the model in Figure 1 is actually a model relating observed variables only, and the latent variables that are shown are exactly equal to their indicators, that is, this is not a latent variable model. Thus the diagram and equations are unnecessarily complex, but we adopt their notation since it is consistent with their work, it is a typical notation for LISREL users, and it permits conceptualization in a true latent variable model by the addition of further indicators of the latent variables. Assume that all variables in the figure have zero means. Teel et al. proposed an $R^{\#}$ -like index, here denoted as $R_{>=}^{\#}(\xi_1)$, of explained variance for a latent variable, ξ_1 , involved in a feedback loop or having a disturbance term correlated with that term of another latent variable. Their proposed index is a modification of a basic index as defined in LISREL. They defined LISREL's $R^{\#}(\xi_1)$ for a latent dependent variance freed from its associated disturbance variance:

$$(1) \quad R^{\#}(\xi_1) = E(\xi_1^2) - \sigma_{\epsilon_1}^2,$$

where $E(\cdot)$ stands for expectation. In this model, ξ_1 is involved in a reciprocal causation relationship with another latent variable ξ_2 :

$$(2) \quad \begin{aligned} \xi_1 &= \gamma_{12}\xi_2 + \beta_{11}\xi_1 + \epsilon_1 \quad \text{and} \\ \xi_2 &= \gamma_{21}\xi_1 + \beta_{22}\xi_2 + \epsilon_2 \end{aligned}$$

and $\sigma_{\epsilon_1}^2$ is the variance of the disturbance term ϵ_1 pertaining to ξ_1 . To obtain an expression of $R^{\#}(\xi_1)$ in terms of model parameters, Teel et al. (1986) used a series of algebraic reformulations leading to their Equation (7) (p. 165). For completeness of this discussion, some of their equations are reproduced here with the same number and a "#" prefix:

$$(7) \quad R^{\#}(\cdot) = \sigma_{\#}^2 + 2\sigma_{\#} \sigma_{\zeta_0} r_{\zeta_0} + \sigma_{\zeta_0}^2 + 2\sigma_{\#} \sigma_{\zeta_1} r_{\zeta_1},$$

where the symbol \langle stands for correlation between its subindexed variables. Formula (7) is the basis of numerical examples provided subsequently in Teel et al. (1986).

Equation (7) by Teel et al. is in general incorrect to the extent that it is based on unwarranted restrictions imposed by its authors (p. 165). One of these constraints is

$$(3) \quad E(\zeta_{\#}^2) = 1.$$

This is an unnecessary restriction in the model under consideration since $\zeta_{\#}$ is a dependent variable--rather than an independent variable--that need not have a variance of 1. In general, the variance of ζ_2 is a function of the variances of the variables that predict it, which presumably could be anything. That is, assumption (3) made by Teel et al. is in general incorrect and therefore represents a model misspecification that makes Equation (7) using it incorrect as well. In addition, to arrive at (7) Teel et al. (1986) used the restrictive relationships

$$(4) \quad E(\zeta_0^2) = r_{\zeta_0}^2 \text{ and}$$

$$E(\zeta_1^2) = r_{\zeta_1}^2$$

The first equation in (4) is true only if $E(\zeta_0^2) = 1$, which need not hold in general; the second equation in (4) is true only if $E(\zeta_1^2) = 1$, which like the preceding and above constraint, $E(\zeta_0^2) = 1$, is incorrect in general. Since both equations (4) used by Teel et al. in deriving (7) are incorrect in the general case, so is formula (7) of Teel et al. for their $R^{\#}$ -like index.

To obtain a generally valid expression for Teel et al.'s index $R^{\#}(\cdot)$ we use their Equation (5) reproduced next, for completeness of this discussion:

$$(5) \quad R^{\#}(\cdot) = \sigma_{\#}^2 E(\zeta_{\#}^2) + 2\sigma_{\#} \sigma_{\zeta_0} r_{\zeta_0} E(\zeta_0^2) + \sigma_{\zeta_0}^2 E(\zeta_0^2) + 2\sigma_{\#} \sigma_{\zeta_1} r_{\zeta_1} E(\zeta_1^2) \\ + 2\sigma_{\#} \sigma_{\zeta_2} r_{\zeta_2} E(\zeta_2^2) + \sigma_{\zeta_2}^2 E(\zeta_2^2) + \sigma_{\zeta_3}^2 E(\zeta_3^2) \\ \propto \sigma_{\#}^2 E(\zeta_{\#}^2) + 2\sigma_{\#} \sigma_{\zeta_0} r_{\zeta_0} E(\zeta_0^2) + \sigma_{\zeta_0}^2 E(\zeta_0^2) + 2\sigma_{\#} \sigma_{\zeta_1} r_{\zeta_1} E(\zeta_1^2) + \sigma_{\zeta_1}^2 E(\zeta_1^2) + \sigma_{\zeta_2}^2 E(\zeta_2^2) + \sigma_{\zeta_3}^2 E(\zeta_3^2)$$

since in the present model ζ_{it} and η_{it} are uncorrelated (see Figure 1). Given that variable means are assumed to be zero, Equation (5) is equivalent to

$$(6) \quad R^2(\zeta_{it}) = \frac{\text{Cov}(\zeta_{it}, \hat{\zeta}_{it})}{\text{Var}(\zeta_{it})} = \frac{\text{Cov}(\zeta_{it}, \beta_1 \eta_{it})}{\text{Var}(\zeta_{it})}$$

where $\text{Var}(\cdot)$ denotes variance and $\text{Cov}(\cdot, \cdot)$ covariance. For the covariances in (6), repeated use of the latent variable equations (2) underlying the model leads to

$$(7) \quad \begin{aligned} \text{Cov}(\zeta_{it}, \eta_{it}) &= \beta_1 \text{Cov}(\zeta_{it}, \eta_{it}) \\ \text{Cov}(\zeta_{it}, \eta_{it}) &= \beta_1 \text{Cov}(\zeta_{it}, \eta_{it}) \end{aligned}$$

Equations (7) yield from (6) the generally valid, corrected version of Teel et al.'s (1986) modified LISREL R^2 -like index:

$$(8) \quad R^2(\zeta_{it}) = \frac{\text{Cov}(\zeta_{it}, \hat{\zeta}_{it})^2}{\text{Var}(\zeta_{it}) \text{Var}(\hat{\zeta}_{it})}$$

Teel et al. consider the LISREL expression, corrected here to (8), to be inappropriate for nonrecursive models because it contains common variance due to ζ_{it} and η_{it} . Specifically, they propose to remove the covariance term $2\text{Cov}(\zeta_{it}, \eta_{it})$ from (6).

Again using (7), a corrected version of their proposed measure is thus

$$(9) \quad R^2_{>=}(\zeta_{it}) = \frac{\text{Cov}(\zeta_{it}, \hat{\zeta}_{it})^2}{\text{Var}(\zeta_{it}) \text{Var}(\hat{\zeta}_{it})}$$

They interpret this expression as representing the squared correlation between ζ_{it} and its predicted value $\hat{\zeta}_{it} = \beta_1 \eta_{it}$. Unfortunately, Equation (9) (or their special case of it) does not have this interpretation. A correct expression for this relation is given below.

Even in this simple model, the LISREL index $R^2(\zeta_{it})$ and Teel et al.'s modification of it to $R^2_{>=}(\zeta_{it})$ are nasty expressions. While a LISREL run automatically will generate $R^2(\zeta_{it})$, modifying it using the Teel et al. approach to yield $R^2_{>=}(\zeta_{it})$ even in this simple model requires a difficult derivation. One cannot expect marketing researchers to do a derivation such as this for arbitrarily complex models. Another approach is needed.

A Measure of Proportion Explained Latent Variance

In agreement with Teel et al. (1986) and also Hayduk (1996), we believe that the LISREL-type $R^{\#}$ -like index is not appropriate with nonrecursive models. It does not represent an appropriate measure of the proportion explained variance for a latent variable that is involved in reciprocal causation relationships or that has disturbance terms correlated with other latent variables' disturbances. This is because with such models and variables, the latent disturbance term is not necessarily uncorrelated with the latent predictors in an equation. As a result, dependent latent variance does not partition simply into the sum of variances of the latent predictor compound and a disturbance term as is true for classical linear regression. Specifically, for Teel et al.'s model in Figure 1, due to the relatedness of ζ_1 and ζ_2 the first of Equations (2) entails

$$(10) \quad \text{Var}(\zeta_1) = E(\zeta_1^2) = \text{Var}(\gamma_1 \zeta_2 + \delta_1 \epsilon_1) = \gamma_1^2 \text{Var}(\zeta_2) + 2\text{Cov}(\gamma_1 \zeta_2 + \delta_1 \epsilon_1, \delta_1 \epsilon_1).$$

Hence

$$(11) \quad R^{\#}(\zeta_1) = E(\zeta_1^2) - \gamma_1^2 \text{Var}(\zeta_2) - 2\text{Cov}(\gamma_1 \zeta_2 + \delta_1 \epsilon_1, \delta_1 \epsilon_1)$$

(cf. Equation (1)). Thus, $R^{\#}(\zeta_1)$ cannot represent that part of the variance of the dependent variable ζ_1 that is explained in terms of the variance of its predictors ζ_2 and ϵ_1 . In fact when $2\text{Cov}(\gamma_1 \zeta_2 + \delta_1 \epsilon_1, \delta_1 \epsilon_1)$ is negative and sufficiently small, $R^{\#}(\zeta_1) < 0$, as follows from (11). Simply deleting the covariance term in (11) in the manner proposed by Teel et al. would not provide a fundamental definition of variance explained.

To avoid this limitation, use can be made of the standard definition of the $R^{\#}$ goodness-of-fit measure in conventional regression analysis as noted above and as was suggested but not accurately implemented by Teel et al.. In such a view, $R^{\#}$ is the squared correlation between an observed response and that optimally predicted by the predictors in its equation. This squared correlation equals the proportion of variance in the dependent variable explained in terms of the predictors. Thus, we propose that in nonrecursive structural equation models the proportion of explained variance in a latent

dependent variable be *defined* as the squared correlation between that variable and its prediction under the fitted model. Specifically, for the Teel et al.'s (1986) model in Figure 1, the proportion of explained variance of ζ_1 in terms of its predictors is defined as

$$(12) \quad P^{\#}(\zeta_1) = \text{Corr}^{\#}(\zeta_1, \hat{\zeta}_1) = \text{Corr}^{\#}(\zeta_1, \sum_{\#} \beta_{\#} \zeta_{\#} + \beta_0 \mathbf{0}_N),$$

as had also been proposed by Teel et al. in their Eq. (9). Here and below the carat or hat defines a predicted dependent variable score $\hat{\zeta}_1 = \sum_{\#} \beta_{\#} \zeta_{\#} + \beta_0 \mathbf{0}_N$. This definition holds in the population when the parameters $\beta_{\#}$ and β_0 are known. In practice, these coefficients also must be estimated from sample data; then, estimators $\hat{\beta}_{\#}$ and $\hat{\beta}_0$ replace $\beta_{\#}$ and β_0 , and (12) defines a sample coefficient. Similarly the proportion explained variance in ζ_2 is defined as

$$(13) \quad P^{\#}(\zeta_2) = \text{Corr}^{\#}(\zeta_2, \hat{\zeta}_2) = \text{Corr}^{\#}(\zeta_2, \beta_{21} \zeta_1 + \beta_{22} \mathbf{0}_N).$$

These equations are not operational, and created a problem for Teel et al., whose implementation of formula for (12), for example, was incorrect. Substitution and simplification for this model yields an operational version of (12) as

$$(14) \quad P^{\#}(\zeta_1) \approx \frac{\beta_{11}^2 \sigma_{\zeta_1}^2}{\beta_{11}^2 \sigma_{\zeta_1}^2 + \beta_{12}^2 \sigma_{\zeta_2}^2 + \beta_0^2 \sigma_{\mathbf{0}_N}^2}, \quad \frac{\beta_{21}^2 \sigma_{\zeta_1}^2}{\beta_{21}^2 \sigma_{\zeta_1}^2 + \beta_{22}^2 \sigma_{\zeta_2}^2 + \beta_0^2 \sigma_{\mathbf{0}_N}^2} < \beta$$

where

$$(15) \quad \sigma_{\zeta_1}^2 \approx \beta_{11}^2 \sigma_{\zeta_1}^2 + \beta_{12}^2 \sigma_{\zeta_2}^2 + \beta_0^2 \sigma_{\mathbf{0}_N}^2, \quad \sigma_{\zeta_2}^2 \approx \beta_{21}^2 \sigma_{\zeta_1}^2 + \beta_{22}^2 \sigma_{\zeta_2}^2 + \beta_0^2 \sigma_{\mathbf{0}_N}^2$$

A similar operational version of (13) can be given. However, it is clear that such formulas are far too complicated for routine use, even in this very basic model. More complex models would yield equations for $P^{\#}$ that are harder to derive and more impractical to apply than (14). A simpler and more general approach would be highly desirable.

In the case of a general nonrecursive model, the proportion of explained variance in an arbitrary latent dependent variable ζ_k requires explicit reference to the full model equation for ζ_k , such as:

$$(16) \quad \zeta_k = \beta_{k1} \zeta_1 + \dots + \beta_{kp} \zeta_p + \beta_{k1} \mathbf{0}_1 + \dots + \beta_{kq} \mathbf{0}_q + \epsilon_k,$$

where ζ_1, \dots, ζ_p represent the set of all dependent latent variables influencing ζ_k . These may be involved in reciprocal causation relationships among themselves and/or with ζ_k , and/or their disturbance terms $\epsilon_1, \dots, \epsilon_p$ may be interrelated or correlated with the corresponding term ϵ_k of ζ_k . As usual, $\theta_1, \dots, \theta_q$ represent the set of all independent latent predictors of ζ_k . Obviously, our definition is meant to be general, and any or all of the variables except the residual ϵ_k appearing in Equation (16) may as well be observed variables. (In a LISREL context, this may require the use of dummy latent variables as occurs in the Figure 1 example). The present general definition of proportion explained variance in a latent dependent variable yields:

$$(17) \quad P^\#(\zeta_k) = \text{Corr}^\#(\zeta_k, \hat{\zeta}_k) = \text{Corr}^\#(\zeta_k, \gamma_{k1}\zeta_1 + \dots + \gamma_{kp}\zeta_p + \beta_{k1}\theta_1 + \dots + \beta_{kq}\theta_q + \epsilon_k).$$

We note that this definition of the proportion of explained latent variance as the square of a correlation coefficient differs from another $R^\#$ -like measure for nonrecursive models introduced by Hayduk (1996, Ch. 3). Hayduk's approach follows Teel et al. in recognizing that the LISREL definition is inadequate. He modifies formula (1) by adjusting the residual variance $\sigma_{\epsilon_k}^2$ associated with a dependent variable by a function of L , where L is the product of the coefficients comprising a loop, yielding a loop-adjusted $R^\#$. Not much is known about the properties of this coefficient, and so we do not pursue it further here.

Estimation Method

Estimation of the explained variance in any dependent latent variable ζ_k , $P^\#(\zeta_k)$, is possible via inclusion into the model of an auxiliary construct

$$(18) \quad A = \gamma_{k1}\zeta_1 + \dots + \gamma_{kp}\zeta_p + \beta_{k1}\theta_1 + \dots + \beta_{kq}\theta_q + \epsilon_k \cdot \epsilon_k.$$

Within typical structural equation modeling (SEM) methodology, this inclusion can be accomplished using a three-step procedure (Raykov, 1997). First, a phantom latent variable is introduced into the model. It can be defined in terms of paths from all variables involved in the defining equation (16) for ζ_k as shown in the middle part of

(18), or equivalently in terms of the difference between the dependent variable and its residual as shown in the right part of (18). When the structural modeling program EQS is utilized to estimate latent explained variance, we may call the auxiliary construct F101 (to be consistent with EQS's notation of F for factor, and picking an arbitrary factor number not already in the model); in LISREL, it would be another ζ variable, say ζ_{p+1} . Thus we may write

$$(19a) \quad \zeta_{p+1} = F101 = b_1 \zeta_1 + \dots + b_p \zeta_p + c_1 \mathbf{0}_1 + \dots + c_q \mathbf{0}_q.$$

When Equation (16) is the originally defining equation, and (19a) is used to define the phantom variable, it is clear that the coefficients of (16) and (19a) must be equated. That is, the following set of restrictions are imposed upon these paths (latent partial regression coefficients) b_1, \dots, b_p and c_1, \dots, c_q , insuring that ζ_{p+1} or F101 equals the auxiliary construct A:

$$(20) \quad b_1 = \gamma_{k1}, \dots, b_p = \gamma_{kp}, c_1 = \beta_{k1}, \dots, c_q = \beta_{kq}.$$

In the final step, the model so extended and restricted is fitted to the data, and an estimate of the critical correlation $\text{Corr}(\zeta_k, \zeta_{p+1})$ or $\text{Corr}(\zeta_k, F101)$ is obtained as an "external model parameter" (see next paragraph; Raykov, 1997). Its square is the estimate of explained latent dependent variance, $P^{\#}(\zeta_k)$, as seen from Equation (7)

As shown in Equation (18), an equivalent way to define the auxiliary construct is to use

$$(19b) \quad \zeta_{p+1} = F101 \propto \zeta_k \cdot \gamma_{k1}.$$

In this case, no new coefficients are introduced, and hence the equalities (20) do not need to be imposed. This is the easier approach within EQS, but it is not so direct within LISREL. If EQS is used for fitting such a model, the estimate of $\text{Cov}(\zeta_k, F101)$ and $\text{Corr}(\zeta_k, F101)$ are obtained by adding into the input file (e.g., Bentler, 1995) the following section

/PRINT

COVARIANCES=YES; CORRELATIONS = YES;

The estimates of the variances of ζ_k and F101, and their covariances with each other and other Vs and Fs in a model, are obtained in the relevant parts of the output section "Model Covariance Matrix for Measured and Latent Variables". The estimate of $\text{Corr}(\zeta_k, F101)$ is then provided by EQS as the pertinent entry (at the crossing of the column and row for ζ_k and F101, respectively) in the output section "Model Correlation Matrix for Measured and Latent Variables". If LISREL is used for model fitting, the estimate of $\text{Corr}(\zeta_k, \zeta_{p+1})$ is obtained as the pertinent entry in the output section "Correlation Matrix of Eta" (with the comprehensive submodel 3B; Jöreskog & Sörbom, 1993; see Raykov, 1997).

Data Application

Here we use the two numerical examples provided by Teel et al. (1986) and their model in the above Figure 1. As noted previously, this model involves no true latent variables, and hence we illustrate our method without such latent variables. Inspection of Figure 1 showing the Teel et al. setup verifies that their errors have no variance. Hence they do not exist, and the factors θ_1 , θ_2 , ζ_1 , and ζ_2 are identical to their indicators λ_1 , λ_2 , γ_1 , and γ_2 respectively. The diagram in Figure 2 represents their reciprocal causation model as a measured variable model with two additional phantom variables F1 and F2, using defining equations analogous to (19b).

Insert Figure 2 about here

In Figure 2, the exogenous variables are V3 and V4, and the endogenous variables involved in the nonrecursive relation are V1 and V2. E1 and E2 are the residuals in the equations. This completely defines the Teel et al. model. However, to implement our method, the latent variables F1 and F2 are added as phantom variables, using the fixed 1.0 and -1.0 paths shown. With these phantom factors, the model setup is complete and can be run by EQS. However, the left part of Figure 2 also shows the covariance term $\text{Cov}(V1, F1) = C_{V1, F1}$ symbolized by the two-way curved dashed arrow connecting V1 and

F1. This represents the unstandardized association of V1 and F1. This covariance is not a parameter of the model, and hence technically it does not belong in the path diagram. It is, however, a consequence of the model and its parameters, and so we draw it to illustrate graphically what the model setup accomplishes. After using the variances of V1 and F1 to standardize the covariance, the desired correlation $\rho_{V1,F1}$ is obtained. Its square $\rho_{V1,F1}^2$ is the variance accounted for. Similarly, in the right part of the figure, we also use a phantom factor F2 to determine the covariance $C_{V2,F2}$, whose standardization gives the correlation between the optimal linear combination F2 and V2, and whose square $\rho_{V2,F2}^2$ gives the proportion of variance explained in V2. In order to carry through these steps, the first EQS input file in the Appendix is used for fitting this model to Teel et al.'s correlation matrix. Their matrix is shown under /MATRIX in the input file. (Since no raw data was provided by the authors, which is required for evaluation of the requisite multinormality assumption, we simply assume that maximum-likelihood applied to a correlation matrix is an appropriate method for their data).

Output from the program shows that the model fits the data perfectly, so we presume that Teel et al. generated their data according to their models. We obtain the estimate of $\text{Corr}(V1, F1)$ as .763. Hence our estimate of the proportion of explained variance in V1 is given as $\rho_{V1,F1}^2 = .763^2 = .582$. That is, with the Teel et al. model in Figure 1, and their data, 58.2% of the variance in V1 is explained by the estimated optimal combination of its predictors V2 and V3. Similarly, $\rho_{V2,F2}^2 = .759^2 = .574$. That is, 57.4% of the variance in the V2 is explained by the estimated optimal compound of its predictors V1 and V4.

The correlated disturbance example from Teel et al. is diagrammed in Figure 3 without Teel et al.'s dummy latent factors and errors, but with the phantom factors F1 and F2 to represent the predicted dependent variables under the model. This model is fitted to their correlation matrix using the second EQS input file in the Appendix.

Insert Figure 3 about here

Using F1, the proportion of explained variance in V1 is found to be $P^\#(V1) = .5^\# = .25$, i.e., 25% of the variance in V1 is explained in terms of variance in its predictor V3. The estimate of $\text{Corr}(V2, F2)$ is similarly found to be .673, and hence the estimate of explained variance in V2 is $P^\#(V2) = .673^\# = .453$. That is, with this model 45.3% of the variance in V2 is explained by the estimated optimal combination of its predictors V1 and V4.

Discussion

We proposed a definition of proportion of variance explained that is consistent with that implied by Teel, Bearden, and Sharma (1986), but our definition has the advantage that it can be applied to any structural model, not only latent variable models or models with reciprocal causal loops. In fact, the examples of Figure 1 actually involve no latent variables although they appear to be latent variable models. The approach is easy to implement, and it avoids questionable results such as incorrectly computed values and out of range numbers for the coefficients. In our applications to the Teel et al. data, we obtained results that are inconsistent with most of their illustrated $R^\#$ values. For example, for the model of Figure 1, implemented as in Figure 2, we obtained values of .582 and .574 for $P^\#(V1)$ and $P^\#(V2)$. In contrast, Teel et al. reported .540 and .530 for the LISREL approach; .175 and .243 for the econometric approach; and .310 and .295 for their own predicted/observed $R^\#$ approach. It is interesting to note that the EQS output on the Model Covariance Matrix gives the unstandardized variances of F1 and F2 as .31 and .295, respectively. Evidently, it is these variances that Teel et al. consider to be $R^\#$ values. Similarly, for the model of Figure 3, as noted above we obtained values of .25 and .453, which are partially inconsistent with the values reported by Teel et al. They reported .25 and .423 for the LISREL approach; .25 and .184 for the econometric approach, and .25 and .25 for their own approach. Again, the EQS output verifies that Teel et al. incorrectly gave unstandardized variances as their $R^\#$ values. In the latter example, the first value is the same for all methods, but the second differs among them.

For the reasons described above, when different results are obtained, we do not consider the alternative values to represent reasonable descriptions of variance explained in the dependent variables of their model. In contrast, we believe that our approach offers a sound definition, implementation, and interpretation. It is also easy to visualize exactly what is being estimated, as illustrated in Figures 2 and 3.

Although our definition of variance explained involves a squared correlation coefficient and hence logically must lie in the 0-1 range, we give a word of caution. Current structural equation modeling programs do not impose the constraint that the covariance matrix among the variables in a model is positive definite or at least positive semidefinite. Bentler and Jamshidian (1994) describe how such a constraint could be imposed during estimation, but their method has not been implemented in public programs to our knowledge. As a result, it remains theoretically possible that correlations outside the $[-1, +1]$ range can be estimated with such programs. Although this does not happen often in current practice, a problematic model could yield estimates of $R^2(\xi_k)$ that are outside its logical range. Typically we would expect this to occur only in the context of a model that does not fit the data and has other anomalies.

Appendix

EQS Input Files for Estimation of Proposed Measure of Explained Latent Variance

With the Teel, Bearden, & Sharma (1986) Model and Data Examples

```
/TITLE
RECIPROCAL CAUSATION RELATIONSHIP EXAMPLE. THEIR VARIABLES X1=
0 AND X2= 0# ARE DENOTED HERE V3 AND V4, RESPECTIVELY. Y1 œ ( AND
Y2 œ (# ARE DENOTED V1 AND V2. AND RESIDUALS ' AND '# ARE
DENOTED E1 AND E2, RESPECTIVELY. F1 AND F2 ARE PHANTOM FACTORS
REPRESENTING OPTIMALLY PREDICTED V1 AND V2.

/SPECIFICATIONS
CASES = 200; VARS = 4; !SAMPLE SIZE IS NOT GIVEN BY TEEL ET AL.
MATRIX=COVARIANCE;

/EQUATIONS
V1 = .5*V2 + .2*V3 + E1;
V2 = .4*V1 + .3*V4 + E2; !RECIPROCAL CAUSATION RELATIONSHIP
F1 = V1 - E1; !AUXILIARY CONSTRUCT FOR ESTIMATING P(V1)
F2 = V2 - E2; !AUXILIARY CONSTRUCT FOR ESTIMATING P(V2)

/VARIANCES
V3 = 1*; V4 = 1*; E1 = .46*; E2 = .47*;

/PRINT
COVARIANCES=YES; CORRELATIONS = YES;

/MATRIX !LOWER HALF OF TABLE 1 OF TEEL ET AL. (1986)
1
.75 1
.25 .1 1
.1875 .375 0 1

/END
```

/TITLE

CORRELATED DISTURBANCES EXAMPLE. THEIR VARIABLES $X_1 = 0$ AND $X_2 = 0$ ARE DENOTED HERE V_3 AND V_4 , RESPECTIVELY. $Y_1 \in (\cdot$ AND $Y_2 \in (\#$ ARE DENOTED V_1 AND V_2 AND RESIDUALS $' \cdot$ AND $' \#$ ARE DENOTED E_1 AND E_2 , RESPECTIVELY. F_1 AND F_2 ARE PHANTOM FACTORS REPRESENTING OPTIMALLY PREDICTED V_1 AND V_2 .

/SPECIFICATIONS

CASES = 200; VARS = 4; !SAMPLE SIZE NOT GIVEN BY TEEL ET AL.

MATRIX=COVARIANCE;

/EQUATIONS

$V_1 = .5 * V_3 + E_1$;

$V_2 = .4 * V_1 + .3 * V_4 + E_2$;

$F_1 = V_1 - E_1$; !AUXILIARY CONSTRUCT FOR ESTIMATING $P(V_1)$

$F_2 = V_2 - E_2$; !AUXILIARY CONSTRUCT FOR ESTIMATING $P(V_2)$

/VARIANCES

$V_3 = 1$ *; $V_4 = 1$ *; $E_1 = .750$ *; $E_2 = .577$ *;

/COVARIANCES

$E_1, E_2 = .216$ *;

/PRINT

COVARIANCES=YES; CORRELATIONS = YES;

/MATRIX !UPPER HALF OF TABLE 1 OF TEEL ET AL. (1986)

1				
.6163	1			
.5	.2	1		
0	.3	.0	1	

/END

FIGURE 1

MODEL WITH RECIPROCAL CAUSATION

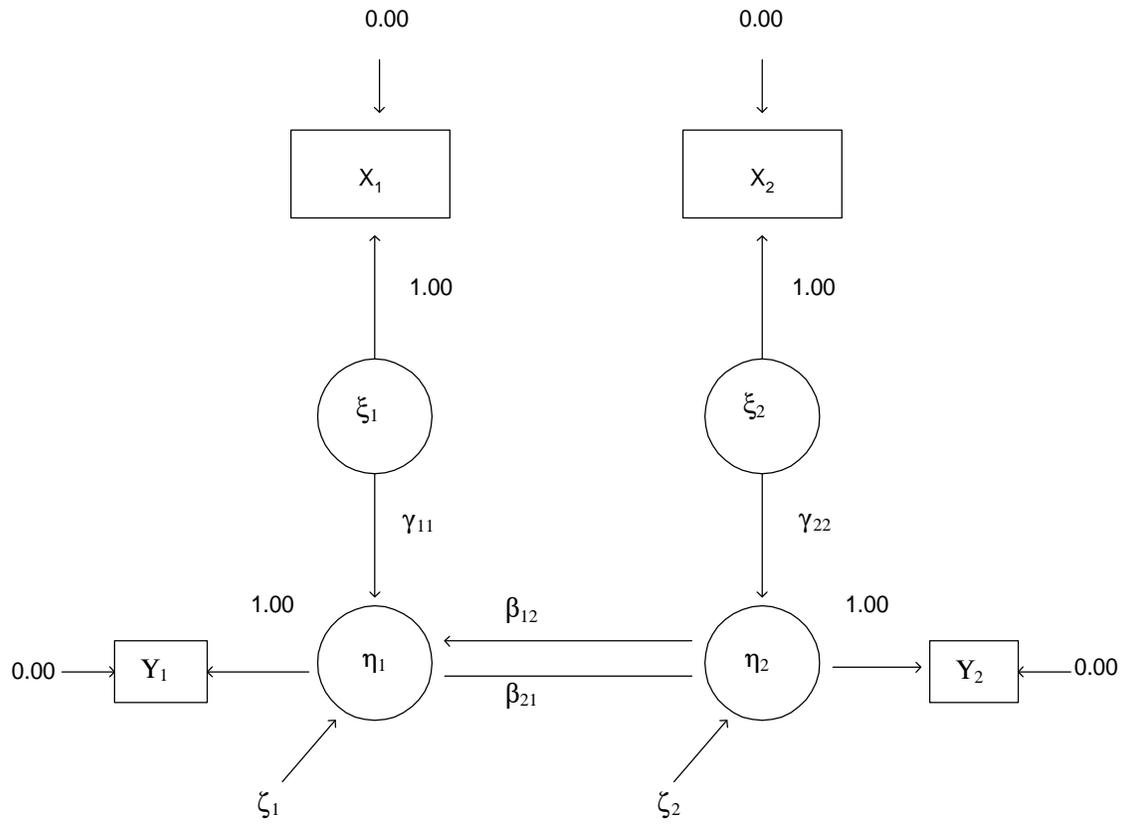


FIGURE 2

MODEL OF FIGURE 1 AS IMPLEMENTED
IN EQS, WITH AUXILIARY FACTORS F1 AND F2

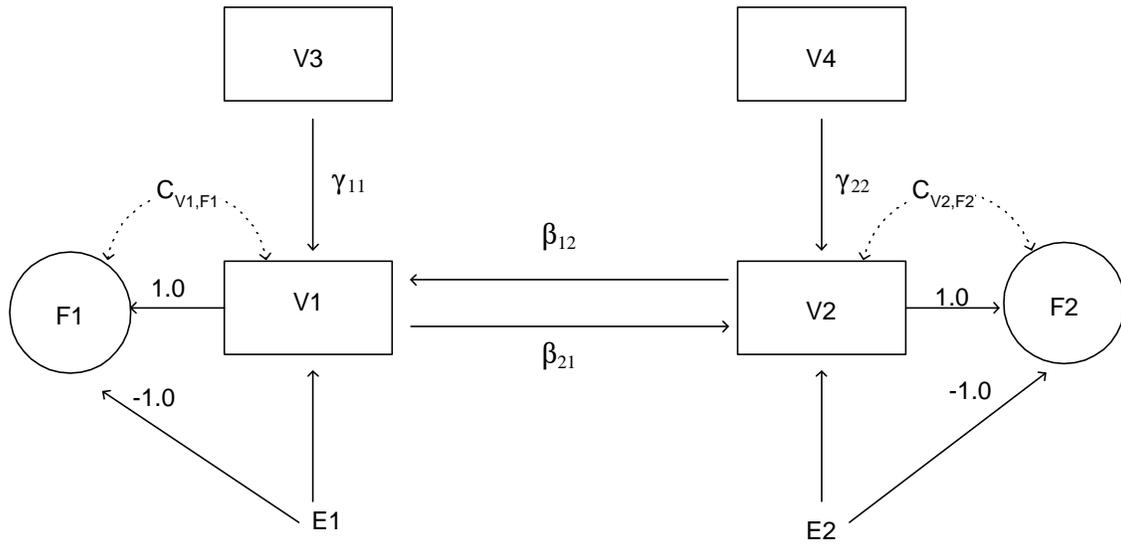
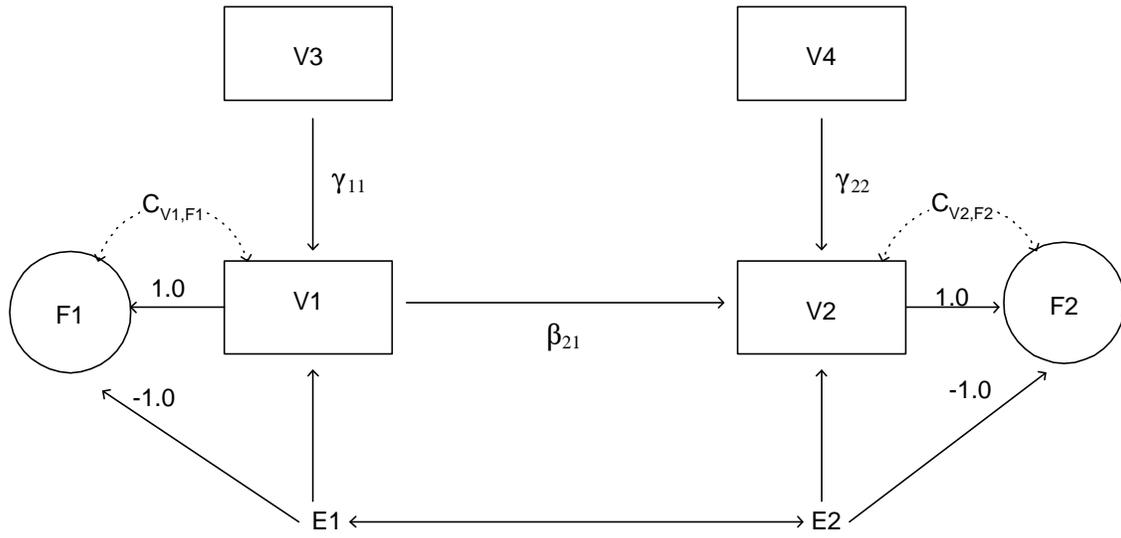


FIGURE 3

MODEL WITH ONE-WAY CAUSATION AND CORRELATED ERROR,
WITH AUXILIARY FACTORS F1 AND F2



References

- Bentler, Peter M. (1995), *EQS Structural Equation Modeling Program Manual*. Encino, CA: Multivariate Software.
- Bentler, Peter M. and Mortaza Jamshidian (1994), "Gramian matrices in covariance structure models," *Applied Psychological Measurement*, 18 (March), 79-94.
- Fox, John (1997), *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage.
- Hayduk, Leslie A. (1996), *LISREL Issues, Debates, and Strategies*. Baltimore, MD: Johns Hopkins University.
- Jain, Dipak (1994), "Regression analysis for marketing decisions," in *Principles of Marketing Research*, Richard P. Bagozzi ed. Cambridge: Blackwell Publishers, 162-194.
- Jennrich, Robert I. (1995), *An Introduction to Computational Statistics: Regression Analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Jöreskog, Karl G. and Dag Sörbom (1993), *LISREL 8 User's Guide*. Chicago, IL: Scientific Software.
- Raykov, Tenko (1997), "Estimation of composite reliability for congeneric measures," *Applied Psychological Measurement*, 21 (June), 173-84.
- Ryan, Thomas (1997), *Modern Regression Methods*. New York: Wiley.
- Schaubroeck, John (1990), "Investigating reciprocal causation in organizational behavior research," *Journal of Organizational Behavior*, 11 (January), 17-28.
- Sharma, Subhash (1996), *Applied Multivariate Techniques*. New York: Wiley.
- Shimp, Terence A. and Alican Kavas (1984), "The theory of reasoned action applied to coupon usage," *Journal of Consumer Research*, 11 (December), 795-809.
- Teel, Jesse E., William O. Bearden and Subhash Sharma (1986), "Interpreting LISREL estimates of explained variance in nonrecursive structural equation models," *Journal of Marketing Research*, 23 (May), 164-8.