CLAES FORNELL and FRED L. BOOKSTEIN*

In marketing applications of structural equation models with unobservable variables, researchers have relied almost exclusively on LISREL for parameter estimation. Apparently they have been little concerned about the frequent inability of marketing data to meet the requirements for maximum likelihood estimation or the common occurrence of improper solutions in LISREL modeling. The authors demonstrate that partial least squares (PLS) can be used to overcome these two problems. PLS is somewhat less well-grounded than LISREL in traditional statistical and psychometric theory. The authors show, however, that under certain model specifications the two methods produce the same results. In more general cases, the methods provide results which diverge in certain systematic ways. These differences are analyzed and explained in terms of the underlying objectives of each method.

Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory

Though they were introduced to marketing only recently, structural equation models with unobservable variables are beginning to change the conventions of marketing research methodology (Bagozzi 1980; Fornell 1982). For social science in general, the new structural equations approach is closely identified with the maximum likelihood covariance structure analysis generalized by Jöreskog (1970, 1973, 1979) and the associated computer program LISREL (Jöreskog and Sörbom 1978, 1981). For marketing, in particular, LISREL has been used for parameter estimation in nearly every application of structural modeling. As powerful as this method is, one may not realistically assume that all problems amenable to use of structural equation models are also suited to LISREL. There are other protocols of structural estimation which impose different assumptions about data, theory, and the ties between unobservable variables and indicators. Marketing data often do not satisfy the requirements of multinormality and interval scaling, or attain the sample size required by maximum likelihood estimation. More fundamentally, two serious problems often interfere with meaningful covariance structure analysis: improper solutions (i.e., solutions outside the admissible parameter space) and factor indeterminacy. Wold’s (1963, 1965, 1975, 1980a, b, 1982) method of partial least squares (PLS) avoids many of the restrictive assumptions underlying maximum likelihood (ML) techniques and ensures against improper solutions and factor indeterminacy. As an illustration of situations in which LISREL is not well suited and PLS is a feasible alternative, we estimate three models for a single data set. The first model compares traditional LISREL ML estimates with PLS estimates for a case in which LISREL produces improper solutions. Contrary to common belief, improper estimates do not necessarily stem from sample variance, lack of fit, or ML estimation, but may be due to the path-analytic fitting objective behind LISREL. We show that the removal of factor indeterminacy via PLS provides an effective cure. In a second model, wherein all factors are explicitly defined, LISREL avoids improper solutions by giving estimates identical to those of PLS. The third model extends the second in a direction consistent with the consumer behavior theory it embodies. Again, the LISREL results are outside the admissible parameter space.

*Claes Fornell is Associate Professor of Marketing, Graduate School of Business Administration, and Fred L. Bookstein is Associate Professor of Radiology, Medical School, University of Michigan.

The authors express their gratitude to Herman Wold for his detailed comments on a previous draft of this article.

1Previously known as NIPALS (nonlinear iterative partial least squares) or NILES (nonlinear iterative least squares).
PARTIAL LEAST SQUARES IN STRUCTURAL MODELING

In partial least squares, the set of model parameters is divided into subsets estimated by use of ordinary multiple regressions that involve the values of parameters in other subsets. An iterative method provides successive approximations for the estimates, subset by subset, of loadings and structural parameters. Extending his theory of fixed-point estimation, Wold (1965) developed this method for structural models with unobservable variables (1974, 1975, 1980a, b). As is the case with LISREL, many of the early elaborations and applications originate from Sweden (Ågren 1972; Areskoug, Wold, and Lytkens 1975; Bergström 1972, Bodin 1974; Lytkens 1966, 1973; Noonan and Wold 1977). In the United States, Hui (1978) extended the model to nonrecursive systems, and Bookstein (1980, 1981) provided a geometrical restatement of its protocols.

Recent discussions of both PLS and LISREL are available from Joreskog and Wold (1981). PLS has been applied in a variety of disciplines, including economics (Apel 1977), political science (Meissner and Uhle-Fassing 1981), psychology of education (Noonan 1980; Noonan and Wold 1980), chemistry (Kowalski, Gergeriach, and Wold 1981), and marketing (Jagpal 1981).

Model Structure

The systematic part of the predictor relation in PLS is the conditional expectation of the predictands for given values of the predictors. The structural relations ("inner relations") are thus specified as stochastic. We write this as

\[ E(\eta_1, \ldots, \eta_k) = \beta^* \eta + \Gamma \xi \]

where \( \eta = (\eta_1, \eta_2, \ldots, \eta_k) \) and \( \xi = (\xi_1, \xi_2, \ldots, \xi_p) \) are vectors of unobserved criterion and explanatory variables, respectively, \( \beta^* (m \times m) \) is a matrix of coefficient parameters (with zeros in the diagonal) for \( \eta \), and \( \Gamma (m \times n) \) is a matrix of coefficient parameters for \( \xi \).

The measurement equations ("outer relations") are

\[ y = \Lambda \eta + \epsilon \]
\[ x = \Lambda \xi + \delta \]

where \( y' = (y_1, y_2, \ldots, y_p) \) and \( x' = (x_1, x_2, \ldots, x_q) \) are the observed criterion and explanatory variables, respectively, \( \Lambda, (p \times m) \) and \( \Lambda, (q \times n) \) the corresponding regression matrices, and \( \epsilon \) and \( \delta \) are residual vectors.

Relationship Between Unobserved and Measured Variables

In PLS, the unobservable variables are estimated as exact linear combinations of their empirical indicators

\[ \eta = \pi_\eta y \]
\[ \xi = \pi_\xi x \]

where \( \pi_\eta (p \times m) \) and \( \pi_\xi (p \times m) \) are regression matrices.

As is evident from equations 2, 3 and 4, 5, the unobserved constructs can be viewed either as underlying factors or as indices produced by the observable variables. That is, the observed indicators can be treated as reflective or formative. Reflective indicators are typical of classical test theory and factor analysis models; they are invoked in an attempt to account for observed variances or covariances. Formative indicators, in contrast, are not designed to account for observed variables; they are used to minimize residuals in the structural relationship.

The choice between reflective and formative modes, which substantially affects estimation procedures, has hitherto received only sparse attention in the literature. Figure 1 exemplifies the choices to be made. Deciding how unobservables and data should be related involves three major considerations: study objective, theory, and empirical contingencies.

If the study is intended to account for observed variances, reflective indicators (Figure 1, Mode A) are most suitable. If the objective is explanation of abstract or

---

Figure 1

THREE DIFFERENT MODES OF RELATING UNOBSERVABLES TO EMPIRICAL INDICATORS

Mode A: Reflective indicators

Mode B: Formative indicators

Mode C: Formative indicators for the exogenous construct, reflective indicators for the endogenous construct

---

2See also Bagozzi and Fornell (1982).
"unobserved" variance, formative indicators (Figure 1, Mode B) would give greater explanatory power. Both formative and reflective indicators can also be used within a single model. For instance, if one intends to explain variance in the observed criterion variables by way of the unobservables, the indicators of the endogenous construct should be reflective and those of the exogenous construct should be formative; the result is a mixed-mode estimation (Figure 1, Mode C). Modes A and B represent two separate principles—mode A minimizes the trace of the residual variances in the "outer" (measurement) equations and mode B minimizes the trace of the residual variances in the "inner" (structural) equation, both subject to certain systematic constraints (discussed subsequently). Mode C is a compromise between the two principles.

Indicator mode is also shaped by an aspect of the substantive theory behind the model: the way in which the unobservable construct is conceptualized. Constructs such as "personality" or "attitude" are typically viewed as underlying factors that give rise to something that is observed. Their indicators tend to be realized, then, as reflective. In contrast, when constructs are conceived as explanatory combinations of indicators (such as "population change" or "marketing mix") which are determined by a combination of variables, their indicators should be formative.

Finally, the choice of indicator mode involves an empirical element. In the formative mode, sample size and indicator multicollinearity affect the stability of indicator coefficients, which in this mode are based on multiple regressions. In the reflective mode, indicator coefficients are based on simple regressions and thus are not affected by multicollinearity. If the indicators are highly collinear but one nonetheless desires optimization of explained structural model variance, one might estimate model B but use loadings, rather than regression weights, for interpretation. This approach is illustrated in our subsequent analyses.

Should the considerations involving study objective, theory or conceptualization, and empirical contingencies be contradictory, the selection of indicator mode may be difficult. For example, one may wish to minimize residual variance in the structural portion of the model (which suggests use of formative indicators), even though the constructs are conceptualized as giving rise to the observations (which suggests use of reflective indicators). In such cases the analyst might estimate twice, once in each mode. If the results correspond, there is no problem. If they differ, a compromise might be worked out using the factor structures of the blocks separately (as suggested by Bookstein 1981); otherwise, a decision as to the overriding concern must be made.

[1] "Population change" is presumed to be determined by natality, mortality, and migration. This example is from Hauser's (1973) criticism of sociologists' overreliance on reflective indication.

Estimation

The PLS model is estimated by (1) the loadings \( (\Lambda_x, \Lambda_y) \) or weights \( (\pi_x, \pi_y) \) which describe how the observations relate to the unobservables, and (2) the structural relations \( (\beta, \Gamma) \), whereby values of unobservables influence values of the other unobservables in the system. Instead of optimizing a global scalar function, PLS estimates by way of a nonlinear operator for which the vector of all estimated item loadings \( (\Lambda_x, \Lambda_y) \) is a fixed point. Since its introduction by Wold in 1963, the theory of fixed point (FP) estimation has been discussed by Lytkens (1968, 1973) and Areskoug (1981), and in a collection of papers edited by Wold (1981). Several developments using some form of FP can be found in the recent psychometric literature (Carroll, Pruzansky, and Kruskal 1980; deLeeuw, Young, and Takane 1976; Kroonenberg and deLeeuw 1980; Kruskal 1980; Young et al. 1976; Perreault and Young 1980; Sands and Young 1980).

FP differs from ML models such as LISREL in its basic principles and assumptions. In ML estimation, the probability of the observed data given the hypothesized model is maximized. Wold's PLS estimation, which is a least squares approach, minimizes residual variances under an FP constraint. ML estimators assume a parametric model, a family of joint distributions for all observables; PLS operates as a series of interdependent OLS regressions, presuming no distributional form at all. FP estimation, then, bears no resemblance to the search for zeros of certain derivatives which characterizes the estimation of ML models.

The distinction between optimizing and fixed-point methods may be compared with the two main models for solving multiple regressions. If we state as our goal construction of a linear form \( e + b_{x1} x_1 + b_{x2} x_2 \) which estimates \( y \) with least error variance, we face a problem in direct minimization. But we may instead invoke the vocabulary of path analysis, referring to the total effect \( b_{xy} \) of variable \( x \) on variable \( y \) and attempting to partition this into a direct effect \( b_{x1} x_1 \) and an indirect effect \( b_{x2} x_2 \) mediated by \( x_1 ' \) s effect on \( x_2 \). The direct and indirect effects taken together must constitute the total; we must have

\[
b_{x1} x_1 + b_{x2} x_2 = b_{xy}
\]

and similarly for \( x_2 \). The result is a system of two equations with two unknowns which are, of course, identical to the normal equations of the usual approach. In this second version all explicit minimization is relegated to the bivariate analyses, the coefficients \( b_{x1}, b_{x2}, b_{x3} \) being solutions of an easier optimum problem, the simple regression. Multiple regression may be thought of, then, as a revision by joint constraints of simple regression coefficients independently arrived at.

A similar distinction separates PLS from LISREL in the structural analysis of systems of latent variables. LISREL poses and solves the global optimization prob-
lem (maximization of likelihood) explicitly. PLS limits its explicit optimization computations to the now-familiar case of ordinary multiple regression. The separate simple analyses are jointly adjusted by nonlinear algebraic constraints according to the specific model specification. For example, consider the specification of Figure 1, Mode B. We estimate this model by giving arbitrary starting values to the weights \( \pi_{ni} \) and \( \pi_{n2i} \), normalizing \( \eta \) to unit variance and regressing it on \( x_1 \) and \( x_2 \). Let \( \hat{\xi} \) be the predicted value of \( \eta \) in this regression. \( \hat{\eta} \) is estimated by a corresponding procedure as the predicted value of \( \hat{\xi} \). That is, \( \pi_{n1i} \) and \( \pi_{n2i} \) are given arbitrary starting values and \( \xi \) is normalized to unit variance and regressed on \( y_1 \) and \( y_2 \). \( \eta \) and \( \xi \) are replaced by \( \hat{\eta} \) and \( \hat{\xi} \), each normalized to unit variance, and the procedure is repeated \( k \) times until \( \hat{\eta}_k \) equals \( \hat{\eta}_{k-1} \) and \( \hat{\xi}_k \) equals \( \hat{\xi}_{k-1} \). At this point, the estimated \( \hat{\eta}_k \) is regressed on the estimated \( \hat{\xi}_k \) to obtain the structural relationship between the two. The regression of \( \hat{\eta}_k \) on \( y_1 \) and \( y_2 \) gives the estimated criterion weights \( \hat{\pi}_{n1} \) and \( \hat{\pi}_{n2} \) and the regression of \( \hat{\xi}_k \) on \( x_1 \) and \( x_2 \) gives the estimated predictor weights \( \hat{\pi}_{n1i} \) and \( \hat{\pi}_{n2i} \). To estimate a model with reflective indicators (e.g., Figure 1, Mode A), we follow a similar procedure with the exception that the observed variables (i.e., \( x \)'s and \( y \)'s) are predictands and the unobserved (i.e., \( \eta \) and \( \xi \)) are predictors.

**Assumptions**

From equation 1 it follows that \( E(\eta'e') = E(\xi'e') = 0 \), where \( \xi = \eta - E(\eta) \) is a vector of residuals. From equations 2 and 3 we have \( E(e) = E(\delta) = E(ye') = E(x\delta') = 0 \). We standardize such that \( E(\eta') = E(\xi') = 0 \) and \( \text{Var}(\eta) = \text{Var}(\xi) = \text{Var}(x_i) = \text{Var}(y_i) = 1 \), all \( i,j,k,r \), and \( E(x) = E(y) = 0 \). The standardization of the observed variables is not an essential assumption. It is made here because we are working with variables of different scales. For predictive purposes, location parameters can be estimated and the preceding standardization dropped.

The residual covariance structure is not restricted in PLS. We define \( E(ee') = \theta_e, E(\delta\delta') = \theta_\delta \), and \( E(\xi\xi') = \Psi \). In contrast to covariance structure models (such as LISREL) in which the objective is to minimize the residual covariance matrix (by reproducing the observed covariances), PLS aims to minimize the trace (sum of the diagonal elements) of \( \Psi \) and, with reflective specification, also \( \theta_{\delta} \) and \( \theta_{\eta} \). Because the off-diagonal elements are not among the unknowns of the model and because the unobservables are explicitly estimated, there are no identification problems for recursive PLS models.

The fixed-point estimation addresses the problem of unknown unobservables by substituting the proxy estimates in an iterative manner (as described before for a simple model).

Because PLS estimation involves no assumptions about the population or scale of measurement, there are no distributional requirements. Residual variances are minimized to enhance optimal predictive power. In contrast, residual covariances in LISREL are minimized for optimal parameter accuracy.

**ANALYSES**

We estimate LISREL and PLS models from a small data set, with collinear indicators, in the context of a study of consumer exit and voice caused by dissatisfaction. The theoretical models and the data set are deliberately chosen to illustrate a case to which LISREL (or other covariance structure models) is not well suited.

The analysis was suggested by Hirschman's (1970) exit-voice theory of consumer reaction to dissatisfaction. This theory states that individuals react to discrepancies between desired and actual states of affairs by the market mechanism of exit-voice or by expressing their dissatisfaction through voice. The exiting consumer makes use of the market by switching brands, terminating usage, or by shifting patronage—all economic actions. In contrast, voice is a political action—a verbal protest directed at the seller, and, if remedy is not obtained, possibly via third parties. Hirschman posits that when the exit option is blocked or when cross-elasticities are low, voice will increase. By this reasoning, exit should dominate in highly competitive markets, whereas the more a market resembles a monopoly, the more voice would be expected. Because seller concentration is a measure of monopoly power, we hypothesize that concentration is related negatively to exit and positively to voice.

**Data**

Both census data and nationwide sample data were used. Data on consumer exit and voice were obtained from Best and Andreasen (1977). From telephone interviews in 24 cities during February and March 1975, they generated a total of 2419 responses (a response rate of 80.3%). Data on concentration were gathered from the 1972 Census of Manufacturers' four-digit concentration ratios. These ratios were adjusted with respect to breadth of product category and geographic scope.

In total, seven variables were used: four measures of concentration, two measures of voice (aided and unaided recall), and one measure of exit. The voice and exit measures were expressed as the proportion of respond-
Table 1

| Exit (Y1) | 1.000 |
| Voice 1 (Y2) | -0.207 | 1.000 |
| Voice 2 (Y3) | -0.128 | 0.691 | 1.000 |
| CR 4 (X1) | 0.121 | 0.306 | -0.178 | 1.000 |
| CR 8 (X2) | -0.001 | 0.268 | 0.144 | -0.914 | 0.972 | 1.000 |
| CR 20 (X3) | -0.114 | 0.294 | 0.148 | 0.826 | 0.893 | -0.968 | 1.000 |

ents who recalled having taken action in each of those categories. From a total of 34 product and service categories in the Best and Andreasen study, we retained 24 for analysis. Though the data are based on a very large number of observations on consumers and firms, the unit of analysis makes the sample size small: the input data are summarized in a matrix of 24 cases by 7 variables. The deleted categories either had no corresponding SIC code for concentration ratios or were too general to be meaningful. Table 1 is the resulting correlation matrix.

Model 1

In LISREL applications, the most common way of relating unobservables to data is by means of reflective indicators. In this mode the model attempts to explain the observed correlations. Reflective indicators in a PLS model imply that the primary objective is to explain the variances of the observed variables. As a starting point for comparative analysis we estimated the first model using reflective indicators (see Figure 2).

In equation form, model 1 is set forth as follows.

\[
\begin{align*}
\eta_1 &= \gamma_1 \xi + \epsilon_1 \\
\eta_2 &= \lambda_2 \gamma_1 + \epsilon_2 \\
\eta_3 &= \lambda_3 \gamma_1 + \epsilon_3
\end{align*}
\]

Figure 2

MODEL 1

The results in Table 2 illustrate a problem most LISREL users probably have encountered more than once (cf. Areskoug 1982; Bentler 1976; Driel 1978; Jöreskog 1979): some variance estimates are negative (in this case, the diagonals of both \( \theta_1 \) and \( \theta_2 \) contain elements with improper signs) and the corresponding standardized loadings (which in this case are correlation coefficients) are greater than one. These are unacceptable results.\(^6\)

In Table 2, the estimates via PLS for Model 1 are also reported. PLS does not produce improper estimates, as all residual variances are actual regression residuals; they are not inferred from the data. The PLS results are thus interpretable; they suggest a significant relationship between concentration and voice along the lines suggested by Hirschman, but show no relationship between concentration and exit. The model is satisfactory insofar as the measurement residuals are small and the loadings significant (with the exception of \( \lambda_3 \)). Overall, the PLS estimates provide limited support for the hypothesis that exit-voice is affected by concentration. The LISREL estimates suggest several possibilities: (1) the theory is wrong, (2) the data are inaccurate, (3) the sample size is too small, or (4) covariance structure analysis is not appropriate for the analysis task.

Model 2

Whatever the cause of the improper solutions, the problem can be circumvented by attending to variances instead of correlations, that is, by working with components rather than factors. Components, which are exact linear combinations of their indicators, "maximize variance" whereas factors "explain covariance." To explore this amelioration we take advantage of the cir-

\(^6\)A common practice for circumventing the problem is to fix the negative variance at zero and reestimate the model, apparently on the grounds that the offending estimate is typically low and insignificant. However, this approach has both theoretical and practical flaws. The model to which it leads is based on neither the principal components nor the common-factor model (Bentler 1976). Also, forcing one offending variance to zero will possibly cause the problem to reappear in other variance estimates. Another approach is to use "various tricks to force the program to stay within the admissible parameter space" (Jöreskog 1981). However, this is not always possible, typically leads to a reduction in goodness of fit, and has questionable theoretical support. As shown by Fornell and Bookstein (1982), inadmissible solutions or so-called "Heywood cases" are not necessarily artifacts of maximum likelihood estimation. It is easy to show that the exact algebraic solutions can all be inadmissible; their pooling thus imputes a solution which is still improper, regardless of estimation method.
Table 2
MODEL 1: PLS AND LISREL ESTIMATES WITH REFLECTIVE INDICATORS

<table>
<thead>
<tr>
<th>PLS estimate</th>
<th>Critical ratio&lt;sup&gt;a&lt;/sup&gt;</th>
<th>LISREL estimate&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Critical ratio&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ&lt;sub&gt;1&lt;/sub&gt;</td>
<td>.96</td>
<td>4.8</td>
<td>.85</td>
</tr>
<tr>
<td>λ&lt;sub&gt;2&lt;/sub&gt;</td>
<td>.99</td>
<td>7.6</td>
<td>.96</td>
</tr>
<tr>
<td>λ&lt;sub&gt;3&lt;/sub&gt;</td>
<td>.99</td>
<td>6.8</td>
<td>1.01</td>
</tr>
<tr>
<td>λ&lt;sub&gt;4&lt;/sub&gt;</td>
<td>.95</td>
<td>3.1</td>
<td>.97</td>
</tr>
<tr>
<td>γ&lt;sub&gt;1&lt;/sub&gt;</td>
<td>.02</td>
<td>1.4</td>
<td>.05</td>
</tr>
<tr>
<td>γ&lt;sub&gt;2&lt;/sub&gt;</td>
<td>.27</td>
<td>2.3</td>
<td>.12</td>
</tr>
<tr>
<td>β&lt;sub&gt;1&lt;/sub&gt;</td>
<td>1.00&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td>1.00&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>β&lt;sub&gt;2&lt;/sub&gt;</td>
<td>.96</td>
<td>3.4</td>
<td>1.49</td>
</tr>
<tr>
<td>β&lt;sub&gt;3&lt;/sub&gt;</td>
<td>.87</td>
<td>1.5</td>
<td>.47</td>
</tr>
<tr>
<td>β&lt;sub&gt;4&lt;/sub&gt;</td>
<td>.99</td>
<td>not estimated</td>
<td>.99</td>
</tr>
<tr>
<td>η&lt;sub&gt;1&lt;/sub&gt;</td>
<td>.93</td>
<td>not estimated</td>
<td>.97</td>
</tr>
</tbody>
</table>

<sup>a</sup>Jackknife estimate divided by jackknifed standard error (r-statistic).
<sup>b</sup>Standardized estimate.
<sup>c</sup>Maximum likelihood LISREL estimate divided by estimated standard error.
<sup>d</sup>Fixed parameter.

cumstance that certain component structures can be estimated by both PLS and LISREL. The MIMIC model (Bagozzi, Fornell, and Larcker 1981; Jöreskog and Goldberger 1975; Stapelton 1978) is one such case.

Let this LISREL model be:

\[
\Phi_\gamma = \begin{bmatrix} 0 & .08 \\ -0.14 & 0.24 \end{bmatrix} \quad \Phi_\xi = \begin{bmatrix} 0^d & -1.21 \\ 0^d & 0^d \end{bmatrix} \]

Let this PLS version of the model be:

\[
\gamma_1 = \eta_1 + \epsilon_1 \\
\gamma_2 = \lambda_2 \eta_1 + \epsilon_2 \\
\gamma_3 = \lambda_3 \eta_1 + \epsilon_3 \\
\gamma_4 = \lambda_4 \eta_1 + \epsilon_4 \\
\]

\[
\Phi_\eta = \begin{bmatrix} 0^d & -1.21 \\ 0^d & 0^d \end{bmatrix} \quad \Phi_\chi = \begin{bmatrix} 0^d & .08 \\ 0^d & 0^d & -0.02 \end{bmatrix} \]

with \( \Phi_\xi \) symmetric. The LISREL model is illustrated in Figure 3 and its estimates are reported in Table 3. For purposes of interpretation we make reference to the loadings for the \( x \)-block, computed as

\[
\Lambda_x = R_x \Gamma_x \\
\]
to yield \( \lambda_{x1} = .37, \lambda_{x2} = .22, \lambda_{x3} = .21, \lambda_{x4} = .09 \).

The PLS version of the model is:
The results, according to a two-construct mode B estimation (see Figure 4), also appear in Table 3. Clearly, the PLS and LISREL solutions have identical \( x \)-weights (\( \gamma \)'s in LISREL, \( \gamma \)'s in PLS). Further, the loadings for the \( y \)-side (\( \lambda \)'s in LISREL) are equal to the loadings of the PLS results up to a factor of \( 1/\gamma_{\text{PLS}} \).

Thus, if LISREL is specified in a MIMIC version with \( \Theta \), symmetric, it will always produce results identical to those from a two-construct PLS model with formative indicators. The formative specification with \( \xi = 0 \)—that unobservables be exact linear combinations of their indicators—is not as restrictive as it may appear. When the errors of the \( y \)-variables are correlated, the error term \( \xi \) of the unobservable is distributed instead throughout the elements of \( \Theta \) (see Hauser and Goldberger 1971).

The PLS estimate of the structural parameter \( \gamma_{\text{PLS}} \) in Model 2 is larger than either of the two estimates \( \gamma_1, \gamma_2 \) in Model 1. This is because the objective of a formative-mode model is to minimize the trace of the residual matrix \( \Psi \) (the variance-covariance matrix of \( \zeta \)), so that the measurement portion of Model 2 absorbs the largest possible part of the total residual, subject to the constraint of being a fixed point. The formative formulation,
then, imputes a stronger relationship between concentration and the construct combining exit with the two voice measures.

Recapitulation of Models 1 and 2

We pause here to review certain important distinctions we have established. Model 1 presumes indicators which are reflective of the constructs. The LISREL estimation yielded negative variances and standardized loadings (correlations) greater than one. The PLS estimates, which by construction cannot yield such improprieties, showed small measurement variances but also low estimated structural parameters.

Model 2 involves formative indicators. Estimates in this form focus on the variance in the structural portion of the model so that more of the net failure of fit is partitioned into measurement residuals. As is evident from Table 3, Model 2 thus exhibits a high structural parameter estimate at the expense of large measurement residuals and covariances. We emphasize that the choice between formative and reflective indicators is not merely a matter of empirical statistical fact. Choice of indicator mode brings conceptual, theoretical, and empirical observations to bear together on the objectives of the study; the partitioning of error variance can be manipulated only insofar as it depends on this choice. In particular, a LISREL MIMIC model specified without a disturbance term and with correlated measurement residuals is equivalent to a formative PLS model with two constructs.

Model 3

Though Model 2 produced identical results from LISREL and PLS, it is not an ideal formulation because no distinction is made, at the abstract level, between exit and voice. Our third model makes this distinction. Let us assume that the objective is to explain the observed y-variables (i.e., in LISREL, their correlations; in PLS, their variances). As in the MIMIC model, assume also that concentration is formed by its indicators without any surplus variance. We arrive at the model depicted in Figure 5 with both formative and reflective indicators. The LISREL equations are

\[
\begin{align*}
(15) & 
\begin{bmatrix}
1 & 0 \\
\beta & 1
\end{bmatrix}
\begin{bmatrix}
\eta_1 \\
\eta_2
\end{bmatrix} =
\begin{bmatrix}
\gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 \\
0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{bmatrix} +
\begin{bmatrix}
0 \\
0
\end{bmatrix}
\end{align*}
\]

\[
(16) 
\begin{align*}
\eta_1 &= \lambda_{\eta_1} \xi + \epsilon_1 \\
\eta_2 &= 0 \lambda_{\eta_2} \xi + \epsilon_2 \\
\eta_3 &= 0 \lambda_{\eta_3} \xi + \epsilon_3
\end{align*}
\]

where \(\eta_1\) is concentration, \(\eta_2\) is voice, and \(y_1\) is exit.

The PLS model is

\[
(17) 
\begin{align*}
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
\eta_1 \\
\eta_2
\end{bmatrix} =
\begin{bmatrix}
\gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 \\
0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{bmatrix} +
\begin{bmatrix}
\xi \\
\epsilon_1
\end{bmatrix}
\end{align*}
\]

\[
(18) 
\xi = \eta_{\xi_1} \pi_{\xi_2} \pi_{\xi_3} \pi_{\xi_4}
\]

\[
(19) 
\begin{align*}
y_1 &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \end{bmatrix} + \epsilon_1 \\
y_2 &= \begin{bmatrix} 0 & \lambda_{\eta_2} \end{bmatrix} \begin{bmatrix} \eta_2 \end{bmatrix} + \epsilon_2 \\
y_3 &= \begin{bmatrix} 0 & \lambda_{\eta_3} \end{bmatrix} \begin{bmatrix} \eta_3 \end{bmatrix} + \epsilon_3
\end{align*}
\]

where \(\eta_1\) is exit (\(y_1\)), \(\eta_2\) is voice, and \(\xi\) is concentration.

The two sets of estimates are reported in Table 4. Again, the LISREL solution contains an inadmissible result: the variance of \(\epsilon_{\xi_3}\) is negative and the corresponding standardized loading \(\lambda_{\eta_3}\) is greater than one. Because these results are uninterpretable, let us focus our attention on the PLS estimates.

From the structural parameters \((\gamma_1, \gamma_2)\), we note a negative relationship between concentration and exit and a positive relationship between concentration and voice. However, the concentration-voice relationship is rather weak with a low critical ratio. The relationships between voice and its indicators (measured by \(\lambda_{\eta_2}\) and \(\lambda_{\eta_3}\)) also have low critical ratios. Further, the residual covariance matrix \(\Theta_{\epsilon}\) contains very large elements. Before interpreting these results, let us explore them in detail by use of the testing system of Fornell and Larcker (1981a) and the "blindfolding" relevance measures suggested by Wold (1982).

In Table 5, the average variance accounted for (AVA) is simply the mean \(R^2\) of the structural model (in this case the mean squared structural parameter). This is a statistic indicating the predictive power of the structural model without regard to the measurement model. Its significance can be determined by using Miller's (1975)
analogue to the traditional F-test in regression (see Fornell and Larcker 1981a, p. 48). Average variance extracted (AVE) is the mean-squared loading for each of the three blocks of indicators. As shown in Table 4, the indicators of concentration have acceptable critical ratios, whereas the critical ratios for the voice indicators are low. This finding may seem surprising in view of the high loadings for voice and the low loadings for concentration, but the reason is that the jackknifed standard errors for the elements in $\pi^2$ (see equation 4) are high in relation to their jackknifed estimates and vice versa for $\pi^1$.

Redundancy, which is the product of the squared structural correlation and AVE, measures the power of the exogenous construct for predicting the $y$-variables. As exit has only one indicator, and therefore no measurement residual, its AVE is one. Thus, redundancy for exit is equal to its squared correlation with concentration. The significance can be assessed by examining the jackknifed critical ratio of $\gamma_1$ (see Table 4) or by Miller's (1975) F-test. Both these tests suggest a significant relationship.

Redundancy for voice, which has two indicators, can also be assessed by Miller's F-test or by a jackknifing procedure using a Stone-Geisser test for predictive relevance (Geisser 1974; Stone 1974). Like the jackknifed standard errors, the Stone-Geisser estimate is nonparametric. An adaptation of Ball's $Q^2$ is the test criterion, $Q^2$ having the form (see Lohmoller 1981)

\[
Q^2 = 1 - \frac{E}{O}
\]

Table 4
MODEL 3: PLS AND LISREL ESTIMATES

<table>
<thead>
<tr>
<th>PLS estimate</th>
<th>Critical ratio$^a$</th>
<th>LISREL estimate$^b$</th>
<th>Critical ratio$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1$</td>
<td>-1.62 (-.03)$^a$</td>
<td>2.6</td>
<td>2.56 (.19)$^b$</td>
</tr>
<tr>
<td>$\pi_2$</td>
<td>6.35 (.05)</td>
<td>3.8</td>
<td>-8.12 (.05)</td>
</tr>
<tr>
<td>$\pi_3$</td>
<td>-10.76 (.14)</td>
<td>4.1</td>
<td>11.68 (.01)</td>
</tr>
<tr>
<td>$\pi_4$</td>
<td>6.42 (.33)</td>
<td>12.1</td>
<td>-6.32 (.50)</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-1.62 (-.03)$^a$</td>
<td>17.6</td>
<td>-10.76 (.14)</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>6.35 (.05)</td>
<td>.87</td>
<td>6.42 (.33)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>87</td>
<td>1.1</td>
<td>.52</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>96</td>
<td>1.5</td>
<td>.29</td>
</tr>
<tr>
<td>$\xi_1$</td>
<td>72</td>
<td>not estimated</td>
<td>2.40</td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>.95</td>
<td>not estimated</td>
<td>.73</td>
</tr>
<tr>
<td>$X$</td>
<td>1.1</td>
<td>10.3 (8 d.f.)</td>
<td>.99</td>
</tr>
</tbody>
</table>

$^a$Jackknife estimate divided by jackknifed standard error (t-statistic).
$^b$Standardized estimate.
$^c$Maximum likelihood LISREL estimate divided by estimated standard error.
$^d$PLS notation; see Figure 5.
$^e$LISREL notation; see Figure 5.
$^f$Loadings in parentheses.
$^g$Fixed parameter.

Table 5
MODEL 3: ADDITIONAL SUMMARY STATISTICS FOR PLS

<table>
<thead>
<tr>
<th></th>
<th>Average variance accounted for (AVA)</th>
<th>Average variance extracted (AVE)</th>
<th>$\rho_1$ (concentration)</th>
<th>$Q^2$ (voice)</th>
<th>$P_{\alpha}$ (exit)</th>
<th>Redundancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E^2$ (exit/concentration)</td>
<td>.17$^a$</td>
<td>.03$^a$</td>
<td>.84$^a$</td>
<td>1.00</td>
<td>.28$^a$</td>
<td>.04$^a$</td>
</tr>
</tbody>
</table>

$^a$Significant at .05.
$^b$Significant at .05 (Miller's F-test).
$^c$For corresponding critical ratios, see Table 4; see also footnote 8.
$^d$Predictive relevance, $Q^2 = - .61$. 

All critical ratios for indicator-construct relationships in the PLS program used here were based on the jackknifing of weights (i.e., $\pi_1$ and $\pi_2$). Alternatively, one could use the loadings (i.e., $A_1$ and $A_2$) as the basis for jackknifing. Because weights typically have larger standard errors, they often provide more conservative estimates.
where $O$ is the sum of squares of the $y$-variables for each endogenous construct with more than one indicator (i.e., in our example, $y_2$ and $y_3$) and $E$ is the error sum of squares. Following Stone (1974), we estimate $Q^2$ by omitting one case from the analysis and predicting this case from the remaining cases. As in jackknifing, the estimation is repeated over all cases to yield an overall index of $Q^2$ as a measure of redundancy without loss of degrees of freedom. If $Q^2 = 1$, the observed endogenous variables can be perfectly reconstructed by the model. The model is said to have predictive relevance if $Q^2 > 0$. As shown in Table 5, the redundancy for voice is low, $Q^2$ is not positive, and the model is thus not relevant for predicting the observed voice indicators ($y_2$, $y_3$).

In sum, Model 3 indicates a negative relationship between market concentration and consumer exit (as a response to dissatisfaction) and a positive relationship between market concentration and consumer voice (as a response to dissatisfaction) as was predicted by theory. However, as is shown by the residual covariance matrix $\Theta_b$, only a very small fraction of the variance of the four concentration ratios ($x$-variables) is included in the structural portion of the model. Further, the relationship between concentration and the observed voice variables is too weak to be useful in a predictive sense. The model thus provides limited empirical support for the concentration/exit-voice relationships. It is not surprising that the measurement residuals are very high for the concentration construct because it is well known that its indicators—concentration ratios—are only crude proxy measures of monopoly power. That is not to say that concentration ratios do not have some theoretical or empirical support. If they are crude proxies, at least some systematic variance must be due to monopoly power. If we assume that the exit-voice theory is correct, it is this variance that we retain for the structural portion of the model and it is very small.

The finding that the model cannot be used to predict observed voice (as opposed to observed exit) is not surprising. A multitude of factors contribute to consumer complaining behavior (voice). The potential for exit, in contrast, is reduced when the opportunities for brand and product switching are restricted. As a result, there is less exit when concentration is high.

**DISCUSSION**

Under the classical assumptions of independence and normally distributed residuals, ML and OLS estimates in regression analysis are identical. In structural equation modeling this is not the case. Except as applied to certain MIMIC models, PLS and LISREL have different objectives and present systematically different results. LISREL attempts to account for observed covariances, whereas PLS aims to account for variances at the observed and/or abstract level (depending on indicator mode). Other major differences between the models include assumptions about factor structure, mechanisms of statistical inference, matters of identification, and interpretation of measurement error, as well as frequency of convergence. Unlike ML techniques, PLS makes minimal demands about measurement scales, sample size, and the distribution of residuals. Small sample sizes—sometimes fewer than the number of variables (Wold 1980c)—can be sufficient for descriptive PLS analyses. Moreover, in contrast to ML, PLS estimation does not involve a statistical model and thus avoids the need for assumptions about scales of measurement. Nominal, ordinal, and interval-scaled variables are permissible in PLS in the same ways as in ordinary regression.11

A primary difference between PLS and LISREL is the structure of unobservables. LISREL specifies the residual structures, whereas PLS specifies the estimates of the unobservables explicitly. This difference bears important implications which have long been debated in the psychometric literature (see the review by Steiger 1979). The main defense of the factor model is that it allows for imperfect measurement by assigning surplus variance to the unobservables. However, such measurement error implies certain disturbing consequences. An infinite number of unobservables may bear the same pattern of correlations with observed variables and yet be only weakly or even negatively correlated with each other (Mulaik and McDonald 1978). For exploratory analyses, such indeterminacy can be very problematic. In confirmatory structural equations, indeterminacy has been thought to be less of a problem by reason of the presumed existence of “prior knowledge” which rules out conflicting explanations. Because the chi square statistic of fit in LISREL is identical for all possible unobservables satisfying the same structure of loadings, a priori knowledge is necessary. However, indeterminacy can create difficulties for confirmatory studies. In some cases several hypothesized models have accounted for the same data equally well (see Mulaik 1976). Thus, confirmatory studies are not necessarily free from the problem of having several interpretations. In this article we suggest an equally serious problem: only indeter-

---

11This is not to imply that all LISREL variables require metric data or multinormal distributions. Dichotomous variables can be handled as fixed $x$-variables or by use of several-group analysis. If we assume an underlying bivariate normal distribution, ordinal variables can be handled by estimating the polychoric correlation instead of the product moment correlation. However, because these correlation coefficients are estimated separately, there is no guarantee that the resulting matrix of correlations is positive definite (which means that ML estimation cannot be used). Further, these correlations cannot be used to correctly estimate standard errors or the chi square goodness-of-fit measure (Jöreskog and Sörbom 1981). When deviations from multinormality cannot be ignored, LISREL V provides an unweighted least squares fitting function but not the associated standard errors or the ML goodness-of-fit interpretation of the chi square.
minimize factors can have improper loadings. Because improper loadings lead to negative variances, such results do not have several interpretations, but rather have none at all.

For the discipline of marketing, which is often concerned with prediction and control, other drawbacks flow from indeterminacy. As the factor scores are indeterminate, specific case predictions are not possible without prior estimation of the scores themselves. Similarly, testing for outliers and modeling of factor scores cannot be done. PLS avoids factor indeterminacy by explicitly defining the unobservables. Factor scores for prediction or further modeling are then readily available.

PLS estimators lack the parameter precision of ML estimation in achieving optimal prediction. Given multivariate normality, LISREL estimates are efficient in large samples, and support analytic estimates of asymptotic standard errors. In exchanging much greater a priori assertion for statistical inference, LISREL is a model for theory testing and a general method for covariance structure analysis in which a theoretical model is specified in terms of covariances and tested against empirical data. PLS is not a model in the same sense; instead of factor analysis or covariance structure analysis, it belongs to the same class of models as canonical correlation, principal components, and regression analysis. It too can be tested against empirical data but via different procedures. The assessment of fit in PLS follows a logic similar to regression analysis and depends on the estimation mode. For a model with reflective indicators, Stone-Geisser's $Q^2$ of redundancy would be appropriate. For a model with formative indicators, where the objective is to minimize the residuals of the structural equation, Miller's F-test or the nonparametric jackknifed standard errors can be used to assess goodness of fit. The corresponding test in LISREL is the likelihood ratio chi square which evaluates goodness of fit with respect to the covariance matrix. In contrast to the PLS measures of fit, a good covariance fit does not support any conclusions beyond those related specifically to accounting for observed covariance. In fact, it is easy to show that the strength of observed variable relationships is negatively associated with goodness of fit (Fornell and Larcker 1981a, b). Therefore, other statistics (e.g., critical ratios) must be used for statistical evaluation of construct relationships and loadings in LISREL.

The factor model underlying LISREL allows more errors in measurement than the components model invoked by PLS. The measurement residuals in LISREL group specific variance with measurement error; no distinction is made between the two. The measurement residuals in PLS are specific if the off-diagonal terms in $\Theta_2$ and $\Theta_3$ are statistically indistinguishable from zero; otherwise they are systematic which means the construct in question is not unidimensional. Thus, LISREL draws upon the classical true-score theory in psychometrics, but ignores the problem of separating measurement error from specific variance. PLS, however, ignores classical true-score theory, but separates measurement residuals (specific and systematic) from the structural equation variance. As correlations between indicators of the same construct decrease, the estimated association between unobservables increases in LISREL. In contrast, the PLS estimate does not go "beyond the data." If the theoretical model is correct and the indicators are valid measurements of the constructs (despite low correlations) the LISREL estimate would be correct whereas the PLS estimate would be biased downward. If one had reason to doubt the accuracy of the theoretical model and/or the validity of the indicators, the LISREL estimate would be exaggerated and more credence could be given to the PLS estimate.

**SUMMARY**

For the marketing analyst, the choice between LISREL and PLS is neither arbitrary nor straightforward. Both apply to the same class of models—structural equations with unobservable variables and measurement error—but they have different structures and objectives.

- LISREL attempts to account for observed covariances, whereas PLS aims at explaining variances (of variables observed and/or unobserved).
- LISREL offers statistical precision in the context of stringent assumptions; PLS trades parameter efficiency for prediction accuracy, simplicity, and fewer assumptions.
- Both models treat measurement residuals, but in different ways. PLS separates out "irrelevant" variance from the structural portion of the model; LISREL combines specific variance and measurement error into a single estimate and adjusts for attenuation.
- LISREL requires relatively large samples for accurate estimation and relatively few variables and constructs for convergence; PLS is applicable to small samples in estimation as well as testing (via jackknifing and the Stone-Geisser test) and appears to converge quickly even for large models with many variables and constructs (Lohmöller 1982).

Only the general research setting can determine the appropriate modeling approach. It is within this context that the LISREL benefits of statistical parameter efficiency and "disattenuated" relationships should be weighed against the problems of indeterminacy and the
TWO STRUCTURAL EQUATION MODELS

necessity of substantial theoretical knowledge. Our analyses suggest that when relevant correlation submatrices (block by block) are not of the appropriate reduced rank, factor indeterminacy is more serious than generally acknowledged. The frequent occurrences of improper and uninterpretable solutions advise against the use of LISREL unless (1) its objectives are consistent with the objectives of the study and (2) its assumptions are verifiably true; if they are not, PLS is a feasible alternative.

REFERENCES


———, H. Wold, and E. Lyttkens (1975), "Six Models with Two Blocks of Observables as Indicators for One or Two Latent Variables," Research Report No. 6, Department of Statistics, University of Göteborg, Sweden.


——— and Arthur S. Goldberger (1975), "Estimation of a


