SOCIAL SCIENCE RESEARCH 17, 384–399 (1988)

Estimation and Hypothesis Testing for Seemingly Unrelated Regressions: A Sociological Application

DIANE H. FELMLEE

Indiana University

AND

LOWELL L. HARGENS

University of Illinois, Urbana

Sociologists frequently use ordinary least squares (OLS) to estimate a series of regression equations from data on the same observational entities. Such "seemingly unrelated regressions" are linked by correlations among the disturbances. In this paper we review three techniques for estimating "seemingly related regressions"—OLS, Zellner's generalized least-squares method, and maximum likelihood estimation—and present an illustrative sociological example employing each technique. We discuss the conditions under which the non-OLS estimation procedures offer advantages for efficiency of estimation and hypothesis testing. © 1988 Academic Press, Inc.

Increased use of structural equation models in the social sciences has led to an emphasis on statistical procedures which simultaneously estimate entire systems of equations. A structural equation model consists of a number of simultaneous equations, all of which are necessary for determining at least one endogenous variable in the system (Kmenta, 1971, p. 532). In this paper we describe another kind of system of equations that can be estimated by simultaneous- rather than single-equation methods. These systems have been named "seemingly unrelated regressions" (SUR) because they consist of equations whose *only* links are correlations among disturbance terms. While methods for estimating SUR systems were developed over two decades ago by Zellner (1962, 1963) and have been utilized by some economists, to date sociologists have made little use of them. Instead of treating their equations as a SUR system, sociologists

The authors express their appreciation to Lawrence Cohen, Laurits Christensen, and James Lincoln for their comments on an earlier version of this paper. Reprint requests should be sent to Diane H. Felmlee, Department of Sociology, Indiana University, Ballantine Hall, Bloomington, IN 47405.

0049-089X/88 $3.00

usually estimate a series of separate regression equations from data on the same units with ordinary least squares (see, for example, Rubinson, 1976; Blau and Blau, 1982; Bursik and Webb, 1982; Crain and Mahard, 1982; Ragin *et al.*, 1982). This is unfortunate because system estimation methods for SUR equations have great potential utility for sociological research.[1]

Here we discuss two potential advantages of system SUR estimation procedures over single-equation approaches. First, such procedures increase the efficiency of coefficient estimates, which in turn increase the likelihood of rejecting the null hypothesis of no effect for explanatory variables. Second, the application of these methods enables the use of cross-equation hypothesis tests which are invalid when correlated equations are estimated separately.

DYNAMIC AND STATIC APPLICATIONS

Correlations among disturbances of equations in SUR systems result from the fact that these equations are estimated from data on the same observational entities. Such correlations can arise in research using either longitudinal or cross-sectional data. For example, instances of SUR systems occur frequently in economic research where comparative time-series analyses are carried out for characteristics of two or more units (e.g., firms, industries) observed at the same time points. Each unit has its own regression equation describing the causal processes that govern it, but because all units under study are subject to the same general economic conditions at each time point, each unit's regression equation is likely to have disturbances that are correlated with those of the other units' equations. In such cases, "contemporaneous covariances" link the disturbances of the several regression equations, and the equations as a set constitute a simultaneous equation system (Theil, 1971, pp. 297–298). Again, this situation is likely to occur whenever two or more regression equations are estimated from time-series data for the same time points.

Certain cross-sectional designs also involve SUR systems of equations. Whenever a researcher estimates a series of regression equations from data on the same observational units, the disturbances of the various equations are likely to be correlated. For example, if one estimates regression equations for the mean educational attainment of blacks and whites from SMSA-level data on the same SMSAs, the disturbances for the two equations are likely to be correlated. Similarly, estimations of two or more single-equation models from data on several levels of the

[1] While the basic issues regarding SUR systems are presented in many recent econometric texts, the major statistics books from other branches of the social sciences have largely ignored this topic. Thus the econometric treatment of SUR is neither familiar, nor accessible, to many social scientists. Our purpose here is to familiarize our noneconomist audience with this topic and to review the recent relevant econometric literature.

same organization are likely to involve correlated disturbances among the estimated equations pertaining to the various levels. These correlations arise because all levels within a given organization are influenced by the same organization-specific cultural and economic conditions. In contrast to comparative time-series analyses, cross-sectional SUR systems have seldom been estimated by efficient statistical methods.

Below we discuss three standard methods for analyzing SUR systems and compare estimation efficiency and hypothesis testing procedures among these methods. The three estimation procedures include ordinary least-squares (OLS) estimates of the separate regression equations, Zellner's generalized least-squares (GLS) method, and maximum likelihood (ML) estimation. We then illustrate these methods and their relative advantages with an empirical analysis of sociological data gathered from a set of business organizations.

METHODS OF ANALYZING SUR SYSTEMS

Estimation of SUR Systems and Ordinary Least Squares

SUR simultaneous equation systems can be represented by the general specifications (see Kmenta, 1971, pp. 517–518)

$$\begin{aligned} Y_1 &= \beta_{11} + \beta_{12}X_{12} + \beta_{13}X_{13} + \cdots + \beta_{1K}X_{1K_1} + \epsilon_1 \\ Y_2 &= \beta_{21} + \beta_{22}X_{22} + \beta_{23}X_{23} + \cdots + \beta_{2K}X_{2K_2} + \epsilon_2 \\ &\;\;\vdots \\ Y_M &= \beta_{M1} + \beta_{M2}X_{M2} + \beta_{M3}X_{M3} + \cdots + \beta_{MK}X_{MK_M} + \epsilon_M \end{aligned} \tag{1a}$$

or:

$$\mathbf{Y}_m = \mathbf{X}_m\boldsymbol{\beta}_m + \boldsymbol{\epsilon}_m \qquad (m = 1, 2 \ldots, M), \tag{1b}$$

where $\mathbf{Y}_m$ is an $(N \times 1)$ vector of observations on the mth dependent variable, $\mathbf{X}_m$ is an $(N \times K_m)$ matrix of observations on $K_m - 1$ nonstochastic independent variables (with $X_{m1} = 1$ for all m), $\boldsymbol{\beta}_m$ is a $(K_m \times 1)$ vector of regression coefficients, and $\boldsymbol{\epsilon}_m$ is an $(N \times 1)$ vector of disturbances. Note that the number of independent variables in each equation need not be the same (e.g., K_1 need not equal K_2). We make the standard assumptions that the $\boldsymbol{\epsilon}_m$ have zero means and are homoscedastic. Thus,

$$E(\epsilon_m) = 0 \tag{2}$$

$$E(\boldsymbol{\epsilon}_m\boldsymbol{\epsilon}_m') = \boldsymbol{\sigma}_{mm}\mathbf{I}_N, \tag{3}$$

where $\mathbf{I}_N$ is an $(N \times N)$ identity matrix. In addition to these specifications, we must allow the disturbances of different equations to be correlated,

$$E(\boldsymbol{\epsilon}_m\boldsymbol{\epsilon}_p') = \boldsymbol{\sigma}_{mp}\mathbf{I}_N \qquad (m, p = 1, 2 \ldots, M; m \neq p), \tag{4}$$

where the covariance of the disturbances of the m^{th} and p^{th} equations, σ_{mp}, is assumed to be constant over all observations.

Under these conditions, estimation of each of the M equations separately by ordinary least squares results in estimates of the $\boldsymbol{\beta}_m$ that are unbiased and consistent (Kmenta, 1971, p. 518) but possibly not efficient (Dhrymes, 1970, pp. 1954–1958; Kmenta, 1971, p. 519). The OLS estimators may not be efficient because they do not take into account the covariances of the disturbances across equations.

Stronger strictures against the use of OLS are encountered when one wishes to carry out tests of hypotheses involving restrictions across equations (for example, that $\beta_{m2} = \beta_{p2}$). In such equations observations are pooled and equations reestimated jointly, but the usual assumption of uncorrelated disturbances still must be met for tests based on OLS to be appropriate (Specht and Warren, 1976, pp. 553–554). Since the disturbances across equations are likely to be present in SUR systems, the joint disturbance matrix for such systems usually has nonzero off-diagonal elements and the assumptions of OLS-based tests are rarely met.

In sum, OLS estimates for SUR systems are unbiased and consistent but are are not necessarily efficient. Furthermore, OLS-based tests of cross-equation restrictions are usually inappropriate for SUR systems. This last condition is especially troublesome because the desire to carry out such tests is often a primary motivation for analyzing SUR systems.

Estimation of SUR Systems by Generalized Least Squares

Zellner (1962) shows that a two-stage procedure employing GLS estimation can be used to obtain estimates for SUR systems. Social scientists are familiar with GLS primarily as an alternative to OLS for single-equation models where the OLS assumptions of homoscedastic and non-autocorrelated disturbances are not met (Hanushek and Jackson, 1977, pp. 135–142). Zellner's application of the GLS principle to SUR systems of equations requires the assumptions of homoscedastic and nonautocorrelated disturbances for any one equation, as specified in the model in the previous section, but it allows correlations between the disturbances in different equations.

GLS estimates for regression coefficients take the general form

$$\beta^* = (\mathbf{X}'\,\boldsymbol{\Omega}^{-1}\,\mathbf{X})^{-1}(\mathbf{X}'\,\boldsymbol{\Omega}^{-1}\,\mathbf{Y}), \tag{5}$$

where $\boldsymbol{\Omega}$ is the known variance–covariance matrix of disturbances. In the case of SUR systems, $\mathbf{X}$ is a block-diagonal supermatrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & 0 & \cdot & \cdot & 0 \\ 0 & \mathbf{X}_2 & \cdot & \cdot & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdot & \cdot & \mathbf{X}_M \end{bmatrix}$$

containing N observations of the K_m independent variables for each of the M equations, Y is an $(MN \times 1)$ vector of observations of a dependent variable, and β^* is a $(\Sigma_{m=1}^{M} K_m \times 1)$ vector of coefficients. $\mathbf{\Omega}$ is a symmetric $(MN \times MN)$ matrix containing MN variances of disturbances and three types of covariances among disturbances (Theil, 1971, pp. 297–298). The first type consists of covariances of the disturbances of the N different observations for each equation and are by (3) equal to zero. The second consists of covariances of the disturbances of the N different observations across different equations and are also assumed to be equal to zero. The final type consists of covariances of the disturbances for the same observations for different equations and are by (4) allowed to be nonzero. Specifying Σ as an $M \times M$ matrix of variances of disturbances and same-unit different-equation covariances (where $\sigma_{ij} = \sigma_{ji}$),

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdot & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \cdot & \sigma_{2M} \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_{M1} & \sigma_{M2} & \cdot & \sigma_{MM} \end{bmatrix}.$$

$\mathbf{\Omega}$ equals the Kronecker product $\Sigma \times \mathbf{I}$, $\mathbf{I}$ being of order $N \times N$. Thus,

$$\mathbf{\Omega} = \begin{bmatrix}
\sigma_{11} & 0 & \dots & 0 & \sigma_{12} & 0 & \dots & 0 & & \sigma_{1M} & 0 & \dots & 0 \\
0 & \sigma_{11} & \dots & 0 & 0 & \sigma_{12} & \dots & 0 & \dots & 0 & \sigma_{1M} & \dots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\
0 & 0 & \dots & \sigma_{11} & 0 & 0 & & \sigma_{12} & & 0 & 0 & \dots & \sigma_{1M} \\
\sigma_{21} & 0 & \dots & 0 & \sigma_{22} & 0 & \dots & 0 & & \sigma_{2M} & 0 & \dots & 0 \\
0 & \sigma_{21} & \dots & 0 & 0 & \sigma_{22} & \dots & 0 & \dots & 0 & \sigma_{2M} & \dots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\
0 & 0 & \dots & \sigma_{21} & 0 & 0 & \dots & \sigma_{22} & & 0 & 0 & \dots & \sigma_{2M} \\
 & & \vdots & & & & \vdots & & & & & \vdots & \\
\sigma_{M1} & 0 & \dots & 0 & \sigma_{M2} & 0 & \dots & 0 & & \sigma_{MM} & 0 & \dots & 0 \\
0 & \sigma_{M1} & \dots & 0 & 0 & \sigma_{M2} & \dots & 0 & \dots & 0 & \sigma_{MM} & \dots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\
0 & 0 & \dots & \sigma_{M1} & 0 & 0 & \dots & \sigma_{M2} & & 0 & 0 & & \sigma_{MM}
\end{bmatrix}$$

Given $\mathbf{\Omega}$ the GLS estimates from (5) are best linear unbiased estimates (Zellner, 1962). Unfortunately, $\mathbf{\Omega}$ is rarely known and must itself be estimated. When consistent estimators of the elements in Σ are available, however, they can be used to estimate $\boldsymbol{\Omega}$ which can be substituted in the GLS formula in (5)

$$\dot{\boldsymbol{\beta}} = (\mathbf{X}'\,\dot{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\,\dot{\boldsymbol{\Omega}}^{-1}\mathbf{Y}). \tag{6}$$

Zellner (1962) shows that estimators based on this substitution are asymptotically equivalent to the GLS estimators in (5) and that they are asymptotically efficient. Zellner also demonstrates that consistent estimators of the elements in Σ can be obtained by estimating each of the SUR equations in (1a) separately by OLS, obtaining the residuals for each of the n observations in each equation, $e_{mn} = y_{mn} - \mathbf{X}_{mn}\,\tilde{\boldsymbol{\beta}}_m$ (where $\tilde{\boldsymbol{\beta}}_m$ is the OLS estimator), and computing s_{mp} as[2]

$$s_{mp} = \frac{1}{N}\sum_{n=1}^{N} e_{mn}e_{pn} \qquad (m,\, p = 1, 2, \ldots, M). \tag{7}$$

Kakwani (1967) shows that Zellner's estimator is unbiased if its expectation exists and the distribution of the disturbances is symmetrical.

We noted earlier that using OLS-based statistical tests of cross-equation restrictions, such as those described in Specht and Warren (1976), is inappropriate when the covariance matrix of disturbances contains nonzero off-diagonal elements. A more general procedure for testing cross-equation restrictions in the presence of such nonzero elements is available (Zellner, 1962; Theil, 1971, pp. 312–317). Let the cross-equation restrictions being tested be represented as

$$\mathbf{r} = \mathbf{R}\boldsymbol{\beta} \tag{8}$$

where $\mathbf{r}$ is a q by 1 vector, $\mathbf{R}$ is a q by $\Sigma_{m=1}^{M} K_m$ matrix of rank M, and $\boldsymbol{\beta}$ is a $\Sigma_{m=1}^{M} K_m$ by 1 vector of parameters (for examples see Theil, 1971, p. 313, and also the example given below). Then assuming multivariate normality of the disturbances and the validity of (8), the test statistic

$$\frac{NM - \Sigma_{m=1}^{M} K_m}{q} \times \frac{(\mathbf{r} - \mathbf{R}\boldsymbol{\beta}^*)'\,(\mathbf{R}[\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}]^{-1}\mathbf{R}')^{-1}(\mathbf{r} - \mathbf{R}\boldsymbol{\beta}^*)}{(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*)'\,\boldsymbol{\Omega}\,(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}^*)} \tag{9}$$

is distributed as F with q and $NM - \Sigma_{m=1}^{M} K_m$ degrees of freedom. The $\boldsymbol{\beta}^*$ vector and the $\boldsymbol{\Omega}$ matrix in this expression come from formula (5),

[2] Rather than using N as the divisor here, $N - K_m$ is sometimes used. Although this produces no difficulties when all equations have the same number of independent variables, when equations have different numbers of independent variables, the use of

$$s_{mp} = \frac{1}{N - K_m}\sum_{n=1}^{N} \mathbf{e}_{mn}\mathbf{e}_{pn}$$

results in $s_{mp} \neq s_{pm}$. Thus, a third expression for s_{mp} is sometimes used:

$$s_{mp} = \frac{1}{\sqrt{(N - K_m)(N - K_p)}}\sum_{n=1}^{N} e_{mn}e_{pn}.$$

indicating that $\boldsymbol{\Omega}$ must be known for the test statistic to have an F distribution. When one only has a consistent estimator of $\boldsymbol{\Omega}$ and substitutes $\dot{\boldsymbol{\beta}}$ and $\dot{\boldsymbol{\Omega}}$ from Eq. (6) into the above expression, the test statistic is asymptotically distributed as F (see also Theil, 1971, pp. 402–403).

Although Zellner's two-stage GLS method for estimating SUR systems and calculation of the test statistic given above are quite laborious when carried out by hand, they are options in widely available computer programs.[3]

Estimation of SUR SYSTEMs by Maximum Likelihood Methods

In his original paper on SUR systems, Zellner (1962) pointed out that the two-stage GLS procedure he discussed could be used as a basis for an iterative procedure for estimating SUR systems. After obtaining the GLS $\dot{\boldsymbol{\beta}}$ vector, new residuals can be computed using $e_{mn} = y_{mn} - \mathbf{X}_{mn}\dot{\boldsymbol{\beta}}_{mn}$, and new s_{mp} values can be obtained using Eq. (7). These new s_{mp} values can then be used in a second GLS estimation of the SUR system, and the whole process can be repeated until the coefficient estimates converge. This iterative procedure has been shown to yield maximum likelihood estimates for SUR systems under the assumption that the disturbances are drawn from a multivariate normal population (Kmenta and Gilbert, 1968; Oberhofer and Kmenta, 1974). Most computer programs that perform two-stage Zellner estimation of SUR systems also include an option for carrying out the iterative estimation.

One can also obtain maximum likelihood estimates directly (Kmenta and Gilbert, 1970). If we assume multivariate normality of disturbances in addition to the standard assumptions described previously, we can construct the logarithm of the joint likelihood function as

$$L = -\frac{MN}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).$$

Maximizing L with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$ yields the maximum likelihood estimates

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\,\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\,\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Y}), \tag{10}$$

$$\hat{\boldsymbol{\Omega}} = [\mathbf{e}_1\mathbf{e}_2 \cdots \mathbf{e}_M]'[\mathbf{e}_1\mathbf{e}_2 \cdots \mathbf{e}_M], \tag{11}$$

where

$$\mathbf{e}_m = \mathbf{Y}_m - \mathbf{X}_m\hat{\boldsymbol{\beta}}_m, \qquad m = 1, 2, \ldots, M.$$

Since (1b) is a simultaneous-equation system, one can use a computer program capable of estimating parameters by maximum likelihood pro-

[3] Such computer programs are SAS (SAS Institute, 1979), SHAZAM (White, 1978), and Time Series Processor (TSP) (Caves and Tretheway, 1978). The iterative Zellner ML procedure discussed in the next section is also available in these programs.

cedures for structural equation models, such as LISREL (Joreskog and Sorbom, 1978), to solve Eqs. (10) and (11). The maximum likelihood estimator, $\hat{\boldsymbol{\beta}}$, is consistent, asymptotically efficient, and asymptotically normally distributed—the same asymptotic properties as those that the Zellner two-stage GLS estimator has been shown to have (Kmenta and Gilbert, 1968; Oberhofer and Kmenta, 1974).

Once maximum likelihood estimators have been obtained, cross-equation hypotheses can be tested. Such tests can use the F statistic in Eq. (9). An alternative procedure can be used to test cross-equation equality restrictions if analysis is done with ML using LISREL (Joreskog and Sorbom, 1978). Under the assumptions that observed variables have a multivariate normal distribution, LISREL generates χ^2 statistics for estimated models and these can be calculated for two models, one estimated without the linear constraints imposed by the null hypothesis and one estimated with these constraints. The difference between these χ^2 is asymptotically distributed as χ^2 with the degree of freedom equal to the difference in the number of estimated parameters.

COMPARISON OF THE THREE METHODS

The Zellner two-stage GLS and ML estimators of SUR systems are asymptotically efficient, and therefore when dealing with large samples we would expect these estimators to have smaller standard errors than those of the OLS estimators. There are two major cases in which no gain in efficiency is achieved by using the GLS or ML estimators. The first consists of the case in which all covariances among the disturbances for the various equations equal zero. In this case the "seemingly unrelated" equations are in fact unrelated, and OLS estimators for the separate equations are best linear unbiased estimates (Hanushek and Jackson, 1977, p. 229). The second case occurs when each of the seemingly unrelated regressions involves exactly the same independent variables, i.e., $\mathbf{X}_m = \mathbf{X}_p$ $(m, p = 1, 2, \ldots, M)$ (Zellner, 1962, 351).[4] The gains in efficiency provided by the two-stage GLS and ML estimators in comparison to OLS estimators are greater to the extent that neither condition is met, and are greatest when the independent variables in each equation are orthogonal to the independent variables in the other equations and when the disturbances in the different equations are highly correlated.

The two-stage GLS and ML estimators are asymptotically more efficient

[4] Thus, exploratory analyses which regress a number of dependent variables on the same set of independent variables will not benefit from the use of SUR estimation. This should not be confused, however, with the situation where independent variables have the same name, but different values, for different equations—a typical occurrence in SUR estimation for time-series data (see Zellner, 1962), and one that can also occur for cross-sectional data as well.

than OLS, but the small sample properties of the first two procedures are generally unknown. Zellner (1963) proved that when the independent variables in each equation are orthogonal to those in the other equations, the efficiency of his two-stage GLS estimators relative to OLS estimators is directly proportional to sample size and the magnitude of the correlations among the disturbances. For example, with a sample size of 11 and a correlation between disturbances of +.5 in a two-equation system, the ratio of the variance of the Zellner estimator to those of OLS equals .86, and when the sample size under these conditions increases to 35 this ratio equals .80. Increases in the correlation between the disturbances under these conditions result in very substantial gains in efficiency: with a sample size of 11 and a correlation between disturbances of +.9, the value of the ratio equals .22 (Zellner, 1963, p. 983).

General small sample properties of either the two-stage GLS or ML estimators when independent variables are not orthogonal across equations have not been determined analytically, although progress on certain aspects of the problem has been made (Phillips, 1977; Bruesch, 1980). Kmenta and Gilbert (1968) carried out Monte Carlo experiments on several small sample SUR systems and reported that in all instances, save the two conditions noted above, the Zellner two-stage and ML estimators had smaller standard errors than those for the OLS estimator. They also noted that the ML estimators tended to have smaller standard errors than the two-stage estimators when correlations among independent variables were low and correlations among disturbances were high, but that the latter tended to yield smaller estimates of standard errors under other conditions. Kmenta and Gilbert point out that it is somewhat hazardous to generalize from their experiments, however, because they are based on a restricted range of models and because differences between estimators are themselves subject to random sampling variability. Thus, although OLS estimators of SUR systems for small samples are probably less efficient than either the two-stage GLS estimators or ML estimators, it is still premature to judge the relative efficiency of the latter two estimators. Because of this, and because both estimators have the same asymptotic properties, many recommend the two-stage GLS estimators because of their greater computational simplicity. We suspect, however, that those with access to electronic computers and software systems that can obtain the two-stage GLS and ML estimators will usually find little difference in terms of the time and expense that they entail.

AN EXAMPLE

To illustrate the differences in results that can be obtained from the three estimators discussed above, we present an illustrative analysis involving four structural characteristics of a sample of 26 Japanese business subsidiaries in Southern California.[5] The structural characteristics are (i) specialization, the extent to which organizational functions are assigned as specialized duties to individuals; (ii) centralization, the level in the official hierarchy of authority given final responsibility for decisions; (iii) formalization, the number of specific role-defining documents in the organization and the degree to which these are distributed among employers; and (iv) ranks, the vertical span of the organization as measured by the number of job positions in the longest line between the lowest level employees and the top-ranking executive. The instruments used to measure these structural properties are based on the Aston scales (see Lincoln *et al.*, 1978).

We began our analysis by examining the relationships of each of the four dependent variables with all of the independent variables included in the Lincoln *et al.* study.[6] Next, we dropped those relationships that fell far below conventional levels of statistical significance and then estimated the resulting SUR system given in[7]

$$\begin{aligned} Y_1 &= \beta_{10} + \beta_{11}X_{11} + b\beta_{12}X_{12} + \beta_{13}X_{13} + \epsilon_1 \\ Y_2 &= \beta_{20} + \beta_{21}X_{21} + \epsilon_2 \\ Y_3 &= \beta_{30} + \beta_{31}X_{31} + \epsilon_3 \\ Y_4 &= \beta_{40} + \beta_{41}X_{41} + \beta_{42}X_{42} + \beta_{43}X_{43} + \beta_{44}X_{44} + \epsilon_4, \end{aligned} \tag{12}$$

where Y_1 is SPEC, specialization; Y_2 is CENT, centralization; Y_3 is FORM, formalization; Y_4 is RANK, ranks; X_{11} is SIZE, the natural logarithm of the total number of employees in the organization; X_{12} is JAMER, the proportion of the organization's employees who are American citizens of Japanese descent; X_{13} is MANUF, which measures whether the function of the organization was manufacturing or not; X_{21} is JAMER; X_{31} is AUTO, the automaticity range index developed by the Aston group to measure technological complexity; X_{41} is SIZE; X_{42} is AUTO; X_{43} is JAMER; and X_{44} is DIST, which measures whether the function of the

[5] These data were collected as part of a study of cultural effects on organizational structure. We chose to study the subsidiaries in order to explore the small-sample properties of the three estimation techniques for SUR systems; subsidiaries are the largest subcategory of the organizations in the study. We thank James R. Lincoln, who kindly made these data available for our analysis.

[6] Since we studied only the subsidiaries, the dummy variables for organizational status were not included. All other variables in the original study were examined.

[7] Ordinarily, the specification of independent variables for each dependent variable in an SUR system should be based on empirical knowledge or theoretical expectations rather than the empirical procedure we employed. Since our purpose is to illustrate differences in results produced by the three estimation methods, this deviation from ideal practice is less objectionable than if we were attempting to draw substantive conclusions about the dependent variable. We also note that if the same independent variables had qualified for inclusion in each of the four reestimated equations, there would have been no reason to obtain GLS or ML estimates of the equations because this is one of the instances in which OLS are best linear unbiased estimates.

organization is distribution or not.[8] Since observations of all variables for each of the four equations are obtained from each of the 26 organizations, one would expect organization-specific characteristics not incorporated in the above equations to produce correlations among the disturbances for the equations. As a result, one would expect OLS estimates of the parameters of the four-equation system to have larger standard errors than those obtained from either the two-stage GLS procedure or the maximum likelihood method. Given the small size of our sample of organizations, however, this expectation, based on the asymptotic properties of the last two methods, may not be realized.

Panel A of Table 1 presents the standard errors associated with each of the nine coefficients in (12) obtained by estimating the SUR system with each of the three methods discussed above. In addition, panel A includes the ratios of the variances of the GLS and ML estimators to those of the OLS estimators. Panel B presents the nine coefficients estimated by the three methods and indicates those that are sufficiently large to enable one to reject the null hypothesis that $\beta = 0$ with varying probabilities of α-error and two-tailed tests.

The results in panel A show that the estimated standard errors for the two-stage GLS and ML methods are uniformly smaller than those for OLS. Thus, our results are consistent with the claim that both estimators are more efficient than OLS estimators even in very small samples. The estimated standard errors yielded by maximum likelihood tend to be slightly smaller than those yielded by the two-stage GLS method with the exception of those for β_{12} and β_{43}. Similarly, the ratios of the variances in panel A show that the ML and GLS estimators have smaller variances than the OLS estimators, especially for the equations with more than one exogenous variable, Eqs. (1) and (4).[9] For some coefficients the

[8] Since the data for our analysis are cross-sectional, we examined scattergrams of the residuals for the OLS estimates of the four equations in order to determine if they suggested violations of the assumption of homoscedasticity (Draper and Smith, 1966, pp. 86–92). Only the first equation showed any systematic variation in the dispersion of residuals with the predicted values of the dependent variable, and in that instance it consisted of smaller residuals for the three cases with the lowest predicted values. Reestimation of the first equation with weighted least squares, either separately or as part of an SUR system, produced results which differed little from those reported below and which yield the same statistical decisions.

[9] The explanatory variables in Eqs. (2) and (3) are subsets of the explanatory variables in the other equations, and the variance ratios for the coefficients in these two equations are the largest. Revankar (1976) shows that in a two-equation system in which the explanatory variables in one equation, equation j, represent a proper subset of those in the other equation, equation k, the two-stage GLS estimators of equation j (with restricted or unrestricted residuals) are identical to the OLS estimators. It is possible that a generalization of his argument to a four-equation system accounts for the small drop in estimated variance in Eqs. (2) and (3) when the non- OLS estimators are used.

TABLE 1
Standard Errors, Variance Ratios, and Parameter Estimates for OLS, Two-Stage GLS, and ML Estimators of Structural Characteristics of 26 Organizations

	Estimation procedure			Ratios of GLS and ML variances to those for OLS	
Parameters	OLS	Two-stage GLS	ML	GLS/OLS	ML/OLS
		A. Standard errors			
β_{11}	.428	.330	.322	.594	.566
β_{12}	3.562	3.067	3.075	.741	.745
β_{13}	.886	.694	.681	.614	.591
β_{21}	11.847	11.378	11.377	.922	.922
β_{31}	.607	.558	.550	.845	.821
β_{41}	.191	.164	.162	.737	.719
β_{42}	.258	.221	.219	.734	.721
β_{43}	1.515	1.353	1.366	.798	.813
β_{44}	.428	.366	.360	.731	.707
	B. Parameter estimates and significance levels				
β_{11}	2.366***	2.633***	2.664***		
β_{12}	−6.018	−8.089**	−8.351**		
β_{13}	− .998	−1.274*	−1.306*		
β_{21}	21.836*	21.371*	21.330*		
β_{31}	2.110***	2.127***	2.112***		
β_{41}	.960***	.893***	.874***		
β_{42}	.666**	.761**	.792***		
β_{43}	−3.571*	−3.389**	−3.312**		
β_{44}	−1.121**	−1.314***	−1.382***		

* $.01 < P \leq .05$.
** $.001 < P \leq .01$.
*** $P \leq .001$.

change in variance is quite substantial. In the case of β_{11}, for instance, using ML rather than OLS results in a decrease in variance of over 43%.

We also note that in this example the estimates of the correlations among disturbances in different equations are fairly small, ranging from −.029 to −.340, while most of the correlations among the independent variables are of moderate sizes. These conditions were not among those which favored maximum likelihood estimators over Zellner GLS estimators in Kmenta and Gilberts' Monte Carlo experiments, but in our analysis the former tend to outperform the latter. We take such results as reinforcing the caution expressed above concerning premature generalization about the relative efficiencies of the GLS and ML estimators.

In general, however, the Zellner two-stage GLS and ML methods appear to be superior to OLS in terms of efficiency of estimation in this empirical example. This is true in spite of the considerably less than ideal conditions favoring gains in efficiency over OLS estimation—small sample size, nonorthogonal regressions, and low correlations among disturbances.

The results in panel B of Table 1 show that coefficients for the same variables vary considerably in their magnitude and direction for different equations. In order to test whether such variation is statistically significant one could employ the test statistic discussed earlier. In the present example we might want to test the null hypothesis that the coefficients for an independent variable that appears in more than one equation are equal across equations. For instance, we could test the hypothesis $\beta_{12} = \beta_{21} = \beta_{43}$, the hypothesis that the coefficients for JAMER in Eqs. (1), (2), and (4) are equal.[10] This hypothesis can be expressed in terms of formula (9) as

$$
\underset{\mathbf{r}}{\begin{bmatrix} 0 \\ 0 \end{bmatrix}} = \underset{\mathbf{R}}{\begin{bmatrix} 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix}} \times \underset{\boldsymbol{\beta}}{\begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{21} \\ \beta_{31} \\ \beta_{41} \\ \beta_{42} \\ \beta_{43} \\ \beta_{44} \end{bmatrix}}
$$

The alternative hypothesis is that at least one of these two restrictions is not true.

These procedures can obviously be used to carry out a large number of tests about the equivalence of one or more coefficients across different equations. Given that (9) is only asymptotically distributed as F, however, use of these tests with small samples yields results that should be interpreted cautiously.

CONCLUDING REMARKS

Seemingly unrelated regression methods are applicable whenever researchers are estimating separate single-equation models on the basis of data from common sources (time points, organizations, regions, individuals,

[10] For these data, there is no clear substantive reason for testing this hypothesis. We discuss it here to illustrate how we would go about testing hypotheses about the equality of coefficients in different equations.

etc.). In such cases, it is reasonable to think that the disturbances of the various equations are correlated because observations taken from the same source are likely to be simultaneously affected by variables not included in the equations. Estimation procedures which take the correlations between the disturbances into account are asymptotically more efficient than OLS. We believe that sociologists interested in topics involving areal or organizational units, or time-series analyses, often encounter situations in which the use of SUR methods is appropriate. Concern over the efficiency of estimation techniques in such analyses is often heightened because there is a small number of cases (often less than 100) and they involve the evaluation of competing theories about a phenomenon in terms of the statistical significance of one or more variables in each equation. Under these conditions, use of an estimation technique which is likely to be inefficient on a priori grounds is unjustifiable.

When concern over the efficiency of estimates is paramount, researchers might consider creating a SUR system to take advantage of the greater efficiency of SUR methods. For example, a study of determinants of organizational centralization in a sample of organizations would usually be based on OLS estimation of coefficients for the effects of independent variables on centralization alone; but more efficient estimates may well be obtained by examining other dependent variables as well and estimating a set of SUR equations by either the Zellner two-stage or ML methods. Such a strategy is similar to those, such as the use of instrumental variables, that use additional information to enable researchers to estimate otherwise unidentifiable coefficients. Of course, if one knew the sources of the correlations between disturbances in a SUR system and had measures of them, it would be possible to make the correlations between disturbances equal to zero by including the measures in the equations. In such instances OLS estimation of the equations is most efficient. Since researchers rarely have either the required knowledge about the sources of these correlations or measures of them, the use of the SUR techniques discussed in this paper should be the rule rather than the exception.[11] We do not wish to imply, however, that SUR techniques absolve the researcher from concern with issues of proper model specification. Like other methods which estimate a set of equations simultaneously, SUR will carry specification errors in any one equation throughout the estimation of the entire system (Wonnacott and Wonnacott, 1970, pp. 397–400; see also Burt, 1976). Thus, the possible gains in efficiency made possible through the use of SUR methods must always be assessed against the possibility of errors due to model misspecification.

[11] Doreian (1980) discusses the use of models that incorporate correlated disturbances, as opposed to those which employ measures of the sources of the correlated disturbances, for the analysis of spatial data. He argues that when both approaches are feasible, the choice between them should be made on substantive grounds.

As we noted earlier, both the Zellner two-stage GLS and the maximum likelihood estimators have the same asymptotic properties and as yet unknown small sample properties. Thus, there are no established statistical grounds for choosing one rather than the other, and the choice can therefore be made on grounds of relative convenience and expense. In our analysis we found no significant difference between the two methods even on these grounds.

Finally, the techniques presented in this paper have been extended to include nonlinear SUR models (Gallant, 1975; Faurot and Fon, 1978), and SUR models with AR(1) autoregression processes (Parks, 1967; Kmenta and Gilbert, 1970).[12] These developments further increase the applicability of the non-OLS approaches to SUR systems reviewed here and make it possible to estimate these systems of equations under the conditions, such as heteroscedasticity and autoregression, that first motivated the development of alternatives to single-equation OLS estimation.

REFERENCES

Blau, J. R., and Blau, P. M. (1982). "The cost of inequality: metropolitan structure and violent crime," American Sociological Review 47, 114–129.

Breusch, T. (1980). "Useful invariance results for generalized regression models," Journal of Econometrics 13, 327–340.

Bursik, R. J., and Webb, J. (1982). "Community change and patterns of delinquency," American Journal of Sociology 88, 24–42.

Burt, R. S. (1976). "Interpretational confounding of unobserved variables in structural equation models," Sociological Methods and Research 5, 3–49.

Caves, D. W., and Tretheway, M. W. (1978). "Time series processor," University of Wisconsin, Madison, unpublished manuscript.

Crain, R. L., and Mahard, R. E. (1982). "The consequences of controversy accompanying institutional change: The case of school desegregation," American Sociological Review 47, 697–708.

Dhrymes, P. J. (1974). Econometrics, Springer-Verlag, New York.

Dielman, T. E. (1983). "Pooled cross-sectional and time series data: A survey of current statistical methodology," The American Statistician 37, 111–122.

Doreian, P. (1980). "Linear models with spatially distributed data: Spatial disturbances or spatial effects?" Sociological Methods and Research 9, 29–60.

Draper, N. R., and Smith, H. (1966). Applied Regression Analysis, Wiley, New York.

Faurot, D. J., and Fon, V. (1978). "A computer program for seemingly unrelated nonlinear regressions," Econometrica 46, 479.

Gallant, A. R. (1975). "Seemingly unrelated nonlinear regressions," Journal of Econometrics 3, 35–50.

Hanushek, E. A., and Jackson, J. E. (1977). Statistical Methods for Social Scientists, Academic Press, New York.

Joreskog, K. G., and Sorbom, D. (1978). "LISREL IV—A General Computer Program

[12] See Srivastava and Dwivedi (1979) and Dielman (1983) for surveys of the literature on estimation of seemingly unrelated regression equations.

for Estimation of a Linear Structural Equation System by Maximum Likelihood Methods," National Education Resources, Chicago.

Kakwani, C. E. (1967). "The unbiasedness of Zellner's seemingly unrelated regression equations estimator." Journal of the American Statistical Association **62,** 141–142.

Kmenta, J. (1971). Elements of Econometrics, MacMillan, New York.

Kmenta, J., and Gilbert, R. F. (1968). "Small sample properties of alternative estimators of seemingly unrelated regressions," Journal of the American Statistical Association **63,** 1180–2000.

Kmenta, J., and Gilbert, R. F. (1970). "Estimation of seemingly unrelated regressions with autoregressive disturbances," Journal of the American Statistical Association **65,** 186–197.

Lincoln, J. R., Olson, J., and Hanada, M. (1978). "Cultural effects on organizational structure: The case of Japanese firms in the United States," American Sociological Review **43,** 829–847.

Oberhofer, W., and Kmenta, J. (1974). "A general procedure for obtaining maximum likelihood estimates in generalized regression models," Econometrica **42,** 579–590.

Parks, R. W. (1967). "Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated," Journal of the American Statistical Association **62,** 500–509.

Phillips, P. C. B. (1977). "An approximation to the finite sample distribution of Zellner's seemingly unrelated regression estimator," Journal of Econometrics **6,** 147–164.

Ragin, C. C., Coverman, S., and Hayward, M. (1982). "Major labor disputes in Britain, 1902–1930: The relationship between resource expenditure and outcome," American Sociological Review **47,** 238–352.

Revankar, N. S. (1976). "Use of restricted residuals in SUR systems: Some finite sample results," Journal of the American Statistical Association **71,** 183–188.

Rubinson, R. (1976). "The world-economy and the distribution of income within states: A cross-national study." American Sociological Review **4,** 638–659.

SAS Institute Inc. (1979). SAS Users Guide, SAS Institute, Inc., Cary, NC.

Specht, D. A., and Warren, R. D. (1976). "Comparing causal models." Sociological Methodology (D. R. Heise, Ed.), pp. 46–82, Jossey-Bass, San Francisco, 1976.

Srivastava, V. K., and Dwivedi, T. D. (1979). "Estimation of seemingly unrelated regression equations: A brief survey," Journal of Econometrics **10,** 15–32.

Theil, H. (1971). Principles of Econometrics, Wiley, New York.

White, K. J. (1978). "A general computer program for econometric methods-SHAZAM," Econometrica **46,** 239–240.

Wonnacott, R. J., and Wonnacott, T. H. (1970). Econometrics, Wiley, New York.

Zellner, A. (1962). "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias," Journal of the American Statistical Association **57,** 348–368.

Zellner, A. (1963). "Estimation of seemingly unrelated regressions: Some exact finite sample results," Journal of the American Statistical Association **58,** 977–982.