BEBR
FACULTY WORKING
PAPER NO. 89-1555

Linearized Estimation of Nonlinear Simultaneous Equation Systems

Anil K. Bera
Ray P. Byron

College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 89-1555

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

April 1989

Linearized Estimation of Nonlinear Simultaneous Equation Systems

Anil K. Bera, Associate Professor
Department of Economics

Ray P. Byron
Australian National University

LINEARIZED ESTIMATION OF NONLINEAR SIMULTANEOUS EQUATION SYSTEMS

Anil K. Bera
University of Illinois
Champaign, Illinois, U.S.A.

and

Ray P. Byron
Australian National University
Canberra, A.C.T., Australia

SUMMARY

Linearized estimation procedure of Byron and Bera for nonlinear single equation functions is generalized to nonlinear simultaneous equation systems. The procedure uses variable linearization through Taylor series expansion rather than parametric linearization as used in the Gauss-Newton and Newton-Raphson methods. It does not require recomputation of derivatives and reinversion of weighting matrices at each iteration stage. Both limited information and full information methods are discussed. Finite sample properties of the estimators are investigated through simulation, and an empirical example is provided to highlight the usefulness of the suggested procedure.

1. Introduction

The appearance of nonlinearity is the rule rather than the exception in economic models because the affects of some economic variables on other variable(s) are not simply additive. Many examples can be given in support of this statement. Unfortunately, econometricians have engaged primarily in problems arising from linear models. Although certain solutions of nonlinear problems are available, applied econometricians seem to prefer estimating linear approximation of their nonlinear models rather than applying nonlinear techniques to original models specified by economic theory. One reason for this may be that the existing procedures are not simple enough to apply in many situations. In an earlier work reported in Byron and Bera (1983), we proposed a simple estimation technique for nonlinear single equation functions. There, by taking a Taylor series expansion around a feasible point in the observation space, the nonlinear function was expressed as a linear function in some observable variables and transformed parameters (some function of the original parameters) plus a remainder term. Consistent estimates of the transformed parameters were obtained by applying iterative ordinary least squares (OLS) with a remainder correction, and from these estimates the original parameters were recovered. An advantage of this procedure was that we avoided some of the computational burden of the Gauss-Newton or Newton-Raphson methods, e.g., at each iteration stage recalculation of derivatives and reinversion of weighting matrix were not required. Moreover, a by-product of the procedure, the transformed parameters had interesting interpretations. They could be viewed as local derivatives or as local elasticities of the nonlinear function. In this paper we extend those results to the simultaneous equation case. Our identification and estimation techniques have close resemblance to the procedures for the linear system. The plan of the paper is as follows. In Section 2, we specify the model and in Section 3, an overview of the literature is given. Section 4 describes the linearized limited information estimation procedure. In Section 5, through a Monte Carlo study we compare our procedure with some of the existing ones. In Section 6, we briefly discuss the linearized full information procedure and provide an empirical example. The paper is closed in Section 7 with some concluding remarks.

2. The Model

We consider the following nonlinear simultaneous equation system (NLSES)

$$q(y_i, x_i, \theta) = u_i, \qquad i = 1,2,\ldots,N \tag{1}$$

where q is a vector valued function and its j-th component is denoted by $q_j(y_i, x_i, \theta_j)$, $j = 1,2,\ldots,m$,

y_i is an m x 1 vector of endogenous variables,

x_i is a k x 1 vector of exogenous variables,

θ_j is a p_j x 1 vector of unknown parameters,

$\theta = (\theta_1', \theta_2', \ldots, \theta_m')'$, $\sum_{j=1}^{m} p_j = p$,

and

u_i is an m x 1 vector of random disturbances, with the following assumptions:

Assumption 1. (i) $y_i \in \mathcal{Y} \subset R^m$, $x_i \in \mathcal{X} \subset R^k$, $u_i \in \mathcal{E} \subset R^m$,

and (ii) θ^* - the true value of θ, is an interior point of Θ, a compact set in R^p.

Assumption 2. q is twice continuously differentiable with respect to all its arguments.

Assumption 3. Random disturbances u_i's are identically and independently distributed with mean zero and variance covariance matrix $\Sigma = ((\sigma_{ij}))$.

3. An Overview of the Literature

3.1 Identification

It is not possible to give standard rank and order type identification conditions for general NLSES. Fisher (1966) generalized the rank and order conditions for linear system to nonlinear system where nonlinearity occurs only through variables. His identification procedure is as follows: given a system, take all possible nonlinear combinations of the equations such that no extra variables other than those are already in the system appear. Combine the original equations and the equations implied by the above operation. Then apply usual rank and order conditions to the augmented system. This procedure deals with only a special case of NLSES and, even then, sometimes it might be tedious to verify the conditions for a large system.

Gallant (1977) defined identification with respect to a set of instruments. This is quite important since, as it was shown by Hausman (1975) and Amemiya (1977), both minimum distance and maximum likelihood estimation procedures can be viewed as instrumental variable (IV) methods. According to Gallant the j-th equation of (1) is said to be identified with respect to the instrument Z if the only solution of the almost sure limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} z_i q_j(y_i, x_i, \theta_j) = 0 \tag{2}$$

is $\theta_j = \theta_j^*$ where θ_j^* is the true value of θ_j.

It is easily understood that the above definition requires the existence of unique consistent estimators for the structural parameters. Gallant (1977) provided sufficient conditions for the existence of the limit in (2). These conditions included the existence of an integrable function dominating $zq_j(y,x,\theta_j)$ and required (x_i, u_i) to form a Casàro summable sequence. However, for a complicated nonlinear system it is not easy to verify these conditions and hence, Gallant's definition is not very useful in spite of its applicability to general models and its intuitive appeal.

3.2 Estimation

A number of estimation techniques are available for NLSES. Here we briefly discuss nonlinear minimum distance (NLMD) and full information likelihood (FIML) procedures. In Subsection 5.2, some further discussion is there for models nonlinear only in variables.

Let

$$q_j(\theta_j) = [q_j(y_1,x_1,\theta_j),q_j(y_2,x_2,\theta_j),\ldots,q_j(y_N,x_N,\theta_j)]' \quad (N\times 1)$$

and

$$q(\theta) = [q_1'(\theta_1),q_2'(\theta_2),\ldots,q_m'(\theta_m)]'. \qquad (Nm\times 1)$$

Definition 1. NLMD estimator for θ, denoted by $\hat{\theta}_{MD}$, is defined as follows:

$$\hat{\theta}_{MD} : \mathcal{Y} \to \Theta$$

such that

$$S(\hat{\theta}_{MD}) = \inf_{\Theta} S(\theta)$$

where $S(\theta) = q'(\theta)Dq(\theta)$ for a suitably chosen matrix D.

Different choices of D will lead to different estimators, e.g., if we take [see Fair and Parke (1980)]

$$D = \begin{bmatrix} z_1 & 0 & \ldots & 0 \\ 0 & Z_2 & \ldots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \ldots & Z_m \end{bmatrix} \left[\begin{pmatrix} z_1' & 0 & \ldots & 0 \\ 0 & Z_2' & \ldots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \ldots & Z_m' \end{pmatrix} (\hat{\Sigma} \otimes I) \begin{pmatrix} Z_1 & 0 & \ldots & 0 \\ 0 & Z_2 & \ldots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \ldots & Z_m \end{pmatrix} \right]^{-1} \begin{bmatrix} Z_1' & 0 & \ldots & 0 \\ 0 & Z_2' & \ldots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \ldots & Z_m' \end{bmatrix} \quad (3$$

where $\hat{\Sigma}$ is a consistent estimator for Σ and z_j is instrument for j-th equation, then we will get nonlinear three-stage least squares (NL3SLS) estimator, and if $\hat{\Sigma}$ in (3) is replaced by an identity matrix,

$\hat{\theta}_{MD}$ will reduce to nonlinear two-stage least squares (NL2SLS) estimator.

Gallant (1977) established strong consistency and asymptotic normality for both NL2SLS and NL3SLS when certain nonlinear parametric restrictions were present across equations and also showed that NL3SLS estimator is asymptotically more efficient than NL2SLS estimator.

Assumption 4. In Assumption 3, the general assumption about the error distribution is now specified as normal with mean vector 0 and variance covariance matrix Σ.

Under the above assumption the log-likelihood function can be written as

$$\ell(\theta,\Sigma) = -\frac{Nm}{2}\ln 2\pi - \frac{N}{2}\ln|\Sigma| + \sum_{i=1}^{N} \ln\|\partial q(y_i,x_i,\theta)/\partial y_i\|$$

$$-\frac{1}{2}\sum_{i=1}^{N} q(y_i,x_i,\theta)'\Sigma^{-1}q(y_i,x_i,\theta). \quad (4)$$

Now,

$$\left.\frac{\partial \ell}{\partial \Sigma}\right|_{\hat{\Sigma}} = 0 \Rightarrow \hat{\Sigma} = \sum_{i=1}^{N} q(y_i,x_i,\theta)q'(y_i,x_i,\theta)/N.$$

Putting this value of $\hat{\Sigma}$ in (4) we get the concentrated log-likelihood function

$$\ell^*(\theta) = -\frac{N}{2}(\ln 2\pi+1) + \sum_{i=1}^{N} \ln \| \partial q(y_i,x_i,\theta)/\partial y_i \|$$

$$- \frac{N}{2} \ln |q(y_i,x_i,\theta)q'(y_i,x_i,\theta)/N|. \qquad (5)$$

Definition 2. FIML estimator for θ, denoted by $\hat{\theta}_{ML}$, is defined as

$$\hat{\theta}_{ML} : \mathcal{Y} \to \Theta$$

such that

$$\ell^*(\hat{\theta}_{ML}) = \sup_{\Theta} \ell^*(\theta).$$

Amemiya (1977) showed that FIML estimator is consistent, asymptotically normal and more efficient than NLMD estimator.[1]

4. Linearized Estimation for NLSES (Limited Information)

4.1 System nonlinear both in parameters and variables

We rewrite the system (1) as

$$q_j(y_i,x_i,\theta_j) = u_{ij}, \qquad i = 1,2,\ldots,N, \; j = 1,2,\ldots,m \quad (6)$$

with earlier definitions of all the variables. Expanding $q_j(y_i,x_i,\theta_j)$ around a "feasible" point $(y_i^o,x_i^o) \in \mathcal{Y} \times \mathcal{X}$, we get

$$q_j(y_i,x_i,\theta_j) + [\partial q_j(y_i,x_i,\theta_j)/\partial y_i \Big|_{(y_i^o,x_i^o)}] \; (y_i-y_i^o)$$

$$+ [\partial q_j(y_i,x_i,\theta_j)/\partial x_i \Big|_{(y_i^o,x_i^o)}] \; (x_i-x_i^o)$$

$$+ r_j(y_i,x_i,y_i^o,x_i^o) = u_{ij} \qquad (7)$$

where $r_j(y_i,x_i,y_i^o,x_i^o)$ is the remainder.

In the expansion, for simplicity we take terms only up to first order and the derivatives are taken with respect to nonconstant variables only.

Collecting terms appropriately (7) can be rewritten as

$$y_i'\Gamma_j + x_i'\Delta_j + r_{ij} = u_{ij} \tag{8}$$

where $r_{ij} = r_j(y_i, x_i, y_i^o, x_i^o)$ and Γ_j, Δ_j are functions of θ_i, y_i^o and x_i^o. Writing (8) in system form

$$Y\Gamma + X\Delta + R = U \tag{9}$$

and putting $R = 0$, we get standard linear system

$$Y\Gamma + X\Delta = U. \tag{10}$$

Definition 3. j-th equation of (6) is said to be partially identified if the usual rank and order condition is satisfied by j-th equation of (10).

Ifj-th equation is partially identified then putting $r_j = 0$, we can get some (inconsistent initial) estimates of Γ_j and Δ_j.

Normalizing (8) with respect to one endogenous variable in each equation, we have[2]

$$y_{ij} = Y_{ij}'\gamma_j + X_{ij}'\delta_j + \bar{r}_{ij} + u_{ij} \tag{11}$$

where Y_{ij} is the vector of included endogenous variables (except y_{ij}), and X_{ij} is the vector of included exogenous variables in j-th equation, γ_j and δ_j are subvectors of respectively Γ_j and Δ_j multiplied by (-1), and $\bar{r}_{ij} = -r_{ij}$.

Definition 4. j-th equation of (6) is said to be fully identified if there exist[3] a differentiable function $h(\cdot)$ such that

$$\theta_j = h\begin{bmatrix}\gamma_j \\ \delta_j\end{bmatrix}. \tag{12}$$

It is not very clear how this definition[4] is related to the definitions of Fisher and Gallant, discussed in Section 3. Apparently our definition seems to be stronger--this will be clear from the model for our simulation study in the next section. If a system (or a part of it) is fully identified we can apply linearized full information (LFI) [or linearized limited information (LLI)] method of estimation. We now discuss LLI method, and LFI method will be discussed in Section 6.

Let us take the first equation of (11) and write it as

$$y_1 = Y_1\gamma_1 + X_1\delta_1 + \bar{r}_1 + u_1 = \hat{Y}_1\gamma_1 + X_1\delta_1 + \bar{r}_1 + e_1 \tag{13}$$

where $\hat{Y}_1$ is obtained from a regression on instruments consisting of X_1 and other variables, and $e_1 = [u_1 + (Y_1-\hat{Y}_1)\gamma_1]$.

We rewrite (13) as

$$y_1 = Z_1\beta_1 + \bar{r}_1 + e_1 \tag{14}$$

where $z_1 = [\hat{Y}_1, X_1]$ and $\beta_1 = \begin{bmatrix}\gamma_1 \\ \delta_1\end{bmatrix}$. It is clear that $\underset{N\to\infty}{\text{plim}} \frac{1}{N} Z_1' e_1 = 0$.

Assumption 5. $Z_1'Z_1$ is positive definite.

An initial estimate of β can be obtained from (14) by putting $\bar{r}_1$ = 0 and then applying OLS, i.e.,

$$\hat{\beta}_1 = (Z_1'Z_1)^{-1}Z_1'y_1.$$

Again, using (12) and (7) we can write

$$\hat{\theta}_1 = h(\hat{\beta}_1)$$

and

$$\hat{\bar{r}}_1 = -[q_1(\hat{\theta}_1) - (y_1 - \hat{Y}_1\hat{\gamma}_1 - X_1\hat{\delta}_1)].$$

Now $\hat{\beta}_1$ can be improved upon using $\hat{\bar{r}}_1$, i.e.,

$$\hat{\hat{\beta}}_1 = (Z_1'Z_1)^{-1}Z_1'(y_1 - \hat{\bar{r}}_1).$$

In general, the iteration procedure at the n-th step will look like

$$\beta_{1(n)} = (Z_1'Z_1)^{-1}Z_1'(y_1 - \bar{r}_{1(n-1)}) \qquad (15)$$

and

$$\theta_{1(n)} = h(\beta_{1(n)})$$

where $\bar{r}_{1(n-1)} = -[q_1(\theta_{1(n-1)} - (y_1 - \hat{Y}_1\gamma_{1(n-1)} - X_1\delta_{1(n-1)})]$ and "(n)" is the subscript for the n-th interation. Putting the value of $\bar{r}_{1(n-1)}$ in (15), we have

$$\beta_{1(n)} = \beta_{1(n-1)} + (Z_1'Z_1)^{-1}Z_1'u_{1(n-1)} \tag{16a}$$

and

$$\theta_{1(n)} = h(\beta_{1(n)}) \tag{16b}$$

where $u_{1(n-1)} = q_1(\theta_{1(n-1)})$.

To ensure convergence of such interation scheme and consistency of the resultant estimator, we have to make a further assumption.

Assumption 6. (i) $Z_1'Z_1 \pm Z_1' \frac{\partial \overline{r}_1}{\partial \beta_1}$ is positive definite in the neighborhood of β_1^*, the true value of β_1.

(ii) $\lim\limits_{N\to\infty} \frac{1}{N} Z_1'Z_1 = P$ and $\operatorname{plim}\limits_{N\to\infty} \left[-\frac{1}{N} Z_1' \left.\frac{\partial \overline{r}_1}{\partial \beta_1}\right|_{\beta_1^*} \right] = Q_1$ exist in the neighborhood of β_1^* and $[I-P^{-1}Q_1]$ is non-singular.

An equivalent form of (16a) is

$$\beta_{1(n)} - \beta_1^* = (Z_1'Z_1)^{-1}Z_1'e_1 - (Z_1'Z_1)^{-1}Z_1' \left.\frac{\partial \overline{r}_1}{\partial \beta_1}\right|_{\overline{\beta}_1} (\beta_{1(n-1)} - \beta_1^*)$$

where $\overline{\beta}_1 \in (\beta_{1(n-1)}, \beta_1^*)$. Now suppose our starting value of β_1 is very close to β_1^*. Then Assumption 6(i) implies that eigenvalues of $(Z_1'Z_1)^{-1}Z_1' \frac{\partial \overline{r}_1}{\partial \beta_1}$ lie in $(-1, 1)$. This ensures the convergence of our algorithm [see Proposition 1 of Byron and Bera (1983)].

For a given sample of size N, let

$$\lim_{n\to\infty} \theta_{1(n)} = \theta_1^N.$$

Proposition 1. Under the Assumptions 1, 2, 3, 5 and 6

$$\operatorname*{plim}_{N\to\infty} \theta_1^N = \theta_1^*.$$

Proof is almost same as given in Proposition 2 of Byron and Bera (1983) for the single equation case.

4.2 System nonlinear only in variables

In this subsection we consider the estimation problems for a special class of nonlinear system where nonlinearity appears only through variables. Such models have been discussed extensively in the literature [e.g., Kelejian (1971), Edgerton (1972), Goldfeld and Quandt (1972), Bowden (1978) and Bowden and Turkington (1981)].

Following Bowden (1978, p. 58) we write the first equation of the system as

$$y_1 = g(Y_1, X_1)\theta_1^1 + X_1\theta_1^2 + u_1 \tag{17}$$

where the symbols have standard interpretation.

For notational simplicity (17) will be referred to as

$$Y_1 = Z_o\theta + u_1 \tag{18}$$

where $Z_o = [g, X_1]$ and $\theta = \begin{bmatrix} \theta_1^1 \\ \theta_1^2 \end{bmatrix}$.

Before discussing our LLI method we briefly mention two of the available procedures, namely, NL2SLS and naive instrumental variables (NIV) methods [see Bowden (1978, Ch. 2)] for a comparative study discussed in the next section.

1. Nonlinear two-stage least squares method: It is basically an IV procedure [see Edgerton (1972)], originally suggested by Kelejian (1971), where instrument for Z_o is

$$\bar{Z} = [\hat{g}, X_1].$$

Here g can be obtained by regressing the entire function g(•) on a "low order polynomial" of X variables. It can be shown that [see Amemiya (1974)] NL2SLS estimators are consistent and

$$\sqrt{N}(\hat{\theta}_{NL2SLS} - \theta^*) \overset{\mathcal{D}}{\rightarrow} N\left[0, \sigma_{11} \text{ plim } [\frac{\bar{Z}'\bar{Z}}{N}]^{-1}\right],$$

where $\overset{\mathcal{D}}{\rightarrow}$ denotes limiting distribution.

Efficiency of this procedure is questionable since it is very difficult to get a proper instrument for g(•). We can increase the efficiency by taking higher degree polynomials while forming the instruments, but then computationally it will become burdensome and "loss of degrees of freedom" will be quite substantial [see Bowden and Turkington (1981)].

2. Naive instrumental variables method: This was suggested by Bowden and Turkington (1981). Here the instrument for Z_o is

$$\tilde{Z} = [g(\hat{Y}_1, X_1), X_1]$$

where $\hat{Y}_1$ can be obtained by regressing Y_1 on the exogenous variables. They showed that under certain conditions this technique gives consistent estimators and

$$\sqrt{N}(\hat{\theta}_{NIV}-\theta^*) \overset{\mathcal{D}}{\to} N\left[0, \sigma_{11} \text{ plim } [\frac{\tilde{Z}'Z_o}{N}]^{-1} \text{ plim } [\frac{\tilde{Z}'\tilde{Z}}{N}] \text{ plim } [\frac{Z_o'\tilde{Z}}{N}]^{-1}\right].$$

Computationally it is quite convenient, and performance of this technique is also quite good as reported in Bowden and Turkington (1981).

Another way to get instruments is to replace g(•) by its conditional expectation given X. However, the use of this technique is very limited because it is not always possible to get a close expression for conditional expectation, and even if we get, the expression will be a function of the structural parameters. Bowden (1978) has successfully applied this technique to estimate disequilibrium econometric models.

3. Linearized limited information method: After linearizing with respect to the variables (18) can be written as

$$y_1 = \underline{Z}H\theta + r + u_1 \tag{19}$$

with

$$\underline{Z} = [Y_1, X_1] - G$$

where G and H are metrices of known constants, and r is the remainder term. If there is a constant term in the equation then G will be a null matrix, and H is nonsingular if the equation is identified.

Taking the instruments as

$$Z = [\hat{Y}_1, X_1] - G$$

and denoting $\beta = H\theta$, the iteration scheme can be written as

$$\beta_{(n)} = \beta_{(n-1)} + (Z'Z)^{-1}Z'u_{1(n-1)} \tag{20}$$

and

$$\theta_{(n)} = H^{-1}\beta_{(n)}$$

where $u_{1(n-1)} = y_1 - Z_o\theta_{(n)} = y_1 - Z_oH^{-1}\beta_{(n)}$.

Since we are dealing with models that are linear in parameters we can modify (20) by putting an explicit expression for a step length at each iteration step. This step length, say d_n, can be obtained by minimizing $u_{1(n)}'u_{1(n)}$ at the n-th step, i.e., choose d_n for the iteration procedure

$$\beta_{(n)} = \beta_{(n-1)} + d_n(Z'Z)^{-1}Z'u_{1(n-1)} \tag{21}$$

such that $u_{1(n)}'u_{1(n)}'$ is minimum. It is easy to verify that the solution is

$$d_n = \frac{u_{1(n-1)}'Z(Z'Z)^{-1}H^{-1'}Z_o'u_{1(n-1)}}{u_{1(n-1)}'Z(Z'Z)^{-1}H^{-1'}Z_o'Z_oH^{-1}(Z'Z)^{-1}Z'u_{1(n-1)}} . \tag{22}$$

It is worthwhile to check the convergence of this modified algorithm and the consistency of the resultant estimator. From (21) we can write

$$\beta_{(n)} = (Z'Z)^{-1}Z'[Z\beta_{(n-1)} + d_nu_{1(n-1)}]$$

$$= (Z'Z)^{-1}Z'[Z\beta_{(n-1)} + d_n(y_1-Z_o\theta_{(n-1)})]$$

$$= (Z'Z)^{-1}Z'[Z\beta_{(n-1)} + d_n(\underline{Z}\beta^*+r+u_1-Z_o\theta_{(n-1)})]$$

$$= (Z'Z)^{-1}Z'[Z\beta_{(n-1)} + d_n(Z\beta^*+r+e_1-Z_o\theta_{(n-1)})]$$

where

$$e_1 = (\underline{Z}-Z)\beta^* + u_1.$$

Since $r_{(n-1)} = Z_o\theta_{(n-1)} - Z\beta_{(n-1)}$, we have

$$\beta_{(n)} = (Z'Z)^{-1}Z'[Z\beta_{(n-1)} + d_n(Z\beta^*+r+e_1-Z\beta_{(n-1)}-r_{(n-1)})]$$

$$= (Z'Z)^{-1}Z'[Z\beta_{(n-1)} + d_n\{Z\beta^*+e_1-Z\beta_{(n-1)} - \frac{\partial r}{\partial\beta}\Big|_{\bar\beta}(\beta_{(n-1)}-\beta^*)\}]$$

since $r - r_{(n-1)} = -\frac{\partial r}{\partial\beta}\Big|_{\bar\beta}(\beta_{(n-1)}-\beta^*)$ where $\bar\beta \in (\beta_{(n-1)}, \beta^*)$.
Therefore,

$$\beta_{(n)} = d_n(Z'Z)^{-1}Z'e_1 - d_n(Z'Z)^{-1}Z'\frac{\partial r}{\partial\beta}\Big|_{\bar\beta}(\beta_{(n-1)}-\beta^*) + (1-d_n)\beta_{(n-1)} + d_n\beta^*$$

or

$$\beta_{(n)} - \beta^* = d_n(Z'Z)^{-1}Z'e_1 - d_n(Z'Z)^{-1}Z'\frac{\partial r}{\partial\beta}\Big|_{\bar\beta}(\beta_{(n-1)}-\beta^*) + (1-d_n)(\beta_{(n-1)}-\beta^*)$$

i.e.,

$$\beta_{(n)} - \beta^* = d_n(Z'Z)^{-1}Z'e_1 - [d_n(Z'Z)^{-1}Z'\frac{\partial r}{\partial\beta}\Big|_{\bar\beta} - (1-d_n)I]\,(\beta_{(n-1)}-\beta^*).$$

So the convergence of this algorithm depends on $\lambda = \max|\lambda_i|$, where λ_i's are the eigenvalues of $[d_n(Z'Z)^{-1}Z'\frac{\partial r}{\partial\beta}\Big|_{\bar\beta} - (1-d_n)I]$. If μ_i's are the eigenvalues of $(Z'Z)^{-1}Z'\frac{\partial r}{\partial\beta}\Big|_{\bar\beta}$ then

$$\lambda_i = d_n\mu_i - (1-d_n).$$

That means for $\mu_i \in (-1,1)$ and $d_n \in (0,1)$[5] for all n, $\lambda_i \in (-1,1)$. Even in certain cases where $\mu_i \notin (-1,1)$, λ_i might lie in $(-1,1)$ interval and this will help the algorithm to converge. Consistency is also guaranteed since

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} Z'e_1 = \operatorname*{plim}_{N\to\infty} \frac{1}{N} Z'u_1 = 0.$$

It is difficult to compare analytically the performances of the various estimation techniques discussed above. Therefore, in the next section we evaluate their relative performance through a simulation study. Since nonlinearity in a system can occur in a number of ways and any simulation study is specific to the model considered, the results of our study provide only some rough indications.

5. Simulation Study

We took the following familiar artificial model for our simulation study [see Bowden and Turkington (1981)]

$$y_{1i} = \theta_{11} \ln y_{2i} + \theta_{12} + \theta_{13}x_i + u_{i1} \tag{23}$$

$$y_{2i} = \theta_{21} \exp(y_{1i}) + \theta_{22}x_i + u_{i2}. \tag{24}$$

According to Fisher's criterion both the equations are identified.

Linearizing these equations by taking a Taylor series expansion around $y_1 = 0$ and $y_2 = 1$, we have

$$y_1 = \theta_{11}y_2 + (\theta_{12}-\theta_{11}) + \theta_{13}x + r_1 + u_1$$

$$y_2 = \theta_{21}y_1 + \theta_{21} + \theta_{22}x + r_2 + u_2 \tag{25}$$

where r_1 and r_2 are the remainder terms. Applying Definition 1, we observe that the first equation is not partially identified whereas the second one is, under the conditions that $\theta_{11} \neq (\theta_{12}-\theta_{11})$. But the set over which $\theta_{11} = (\theta_{12}-\theta_{11})$ has measure 0. Therefore, the equation (25) is partially identified almost everywhere. Again, from (25) we see that (24) is fully identified. So LLI technique can be applied to estimate θ_{21} and θ_{22}.

Observations on the exogenous variable were generated from a uniform distribution with mean 16 and different variances. They were kept fixed throughout 100 replications we performed. The structural disturbances were generated from a bivariate normal distribution with two different variance-covariance matrices

$$\Sigma_1 = \begin{bmatrix} .50 & .25 \\ .25 & 2.00 \end{bmatrix} \text{ and } \Sigma_2 = \begin{bmatrix} .50 & .50 \\ .50 & 10.00 \end{bmatrix} .$$

We took the following values for the structural coefficients: $\theta_{11} = 1$, $\theta_{12} = -10$, $\theta_{13} = .5$, $\theta_{21} = .5$ and $\theta_{22} = 15$. With these parameter values we obtained explicit expressions for y_1 and y_2, in terms of x, u_1 and u_2:

$$y_2 = (15x+u_2)/[1-.5 \exp(-10+.5x+u_1)]$$

and

$$y_1 = \ln y_2 - 10 + .5x + u_1.$$

Data for y_1 and y_2 were generated according to the above expressions.

Five estimation techniques - OLS, NIV, NL2SLS1, NL2SLS2 and LLI were selected for comparison. For NIV, NL2SLS1 and LLI the first stage estimations were done by taking only linear function of x, whereas for NL2SLS2 we took both linear and quadratic terms. For LLI we used the following stopping rule [see Gallant (1975, p. 76)]: stop at the (n+1)-th step if

$$\| \theta_{(n)} - \theta_{(n+1)} \| < \varepsilon_1 (\| \theta_{(n)} \| + \varepsilon_2)$$

and

$$| ESS(\theta_{(n)}) - ESS(\theta_{(n+1)}) | < \varepsilon_1 [ESS(\theta_{(n)}) + \varepsilon_2]$$

where ESS is the error sum of squares and we took ε_1 = .000001 and ε_2 = .0001. Three different criteria were used for comparison - mean bias, MSE and variance. They were calculated in the following way. Let δ^r be the estimate of a parameter δ in the r-th replication, r = 1,2,...,100. Then

$$\text{(i) Mean bias} = \frac{1}{100} \sum_{r=1}^{100} (\delta^r - \delta)$$

$$\text{(ii) MSE} = \frac{1}{100} \sum_{r=1}^{100} (\delta^r - \delta)^2$$

and

$$\text{(iii) Variance} = \frac{1}{100} \sum_{r=1}^{100} (\delta^r - \bar{\delta})^2$$

where $\bar{\delta} = (1/100) \Sigma_{r=1}^{100} \delta^r$.

For sample size 30 and variance covariance matrix Σ_1 results are reported in Table 1. Given the numerical magnitudes of θ_{21} and θ_{22}, in terms of MSE and variance estimates of θ_{22}, the parameter that is associated with a linear function of the exogenous variable, are always better than those of θ_{21} which is associated with a nonlinear function of one endogenous variable. As the variance of x increases, the MSEs and variances of all the techniques decrease systematically but for the biases there is no systematic improvement. Also, as expected NL2SLS2 is always better than NL2SLS1. Performance of NIV is always better than NL2SLS1. Similar results were also obtained by Bowden and Turkington (1981).

Insert Table 1 Here

In terms of MSE and variance OLS is the best. In certain cases OLS also performs quite well in terms of mean bias. Comparing the MSEs and the variances of the last four techniques that give consistent estimators, for θ_{21} - LLI is always superior to others, and for θ_{22} - LLI is superior to both NIV and NL2SLS1 when V(x) = 1. It seems that an increase in V(x) does not improve the performance of LLI very much compared to the improvements in the other three techniques. When the noise level is changed to Σ_2, individual performances of all the five methods, as it should be, become worse (but not in terms of mean bias), but their relative performance remains unchanged. These results are reported in Table 2.

Insert Table 2 Here

In Tables 3 and 4 we give the results for sample size 80. Making pairwise comparison with the corresponding quantities in Tables 1 and 2, we observe that in all cases there is an improvement in terms of MSE and variance. But rather surprisingly mean biases do not decrease as expected [similar behavior of mean bias was also noticed by Goldfeld and Quandt (1972, p. 244)]. We also note that the relative performance of LLI remains the same. Therefore, in conclusion of this simulation study we may say that LLI performs quite favorably compared to the other methods.

Insert Tables 3 and 4 Here

6. Linearized Estimation Method (Full Information)

Let us write the linearized system of equations as [see equation (11)]

$$y_j = Y_j\gamma_j + X_j\delta_j + \overline{r}_j + u_j, \qquad j = 1, 2, \ldots, m.$$

Now let

$$Z = \begin{bmatrix} Z_1 & 0 & \cdots & 0 \\ 0 & Z_2 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & Z_m \end{bmatrix}$$

where $Z_j = [\hat{Y}_j, X_j]$, $u = [u_1', u_2', \ldots, u_m']'$, and $\beta = [\beta_1', \beta_2', \ldots, \beta_m']'$ with $\beta_j = [\gamma_j', \delta_j']'$.

We assume all the equations are identified according to our Definitions 3 and 4. So there exists a differentiable function h(·) such that

$$\theta = h(\beta).$$

Then our earlier algorithm [see Equation (16)] can be generalized straightforwardly as

$$\beta_{(n)} = \beta_{(n-1)} + d_n[Z'(\hat{\Sigma}^{-1}_{(n-1)} \otimes I)Z]^{-1} Z'(\hat{\Sigma}^{-1}_{(n-1)} \otimes I)u_{(n-1)} \quad (26a)$$

and

$$\theta_{(n)} = h(\beta_{(n)}) \quad (26b)$$

where (i,j)-th element of $\hat{\Sigma}_{(n)}$ is $[u'_{i(n)} u_{j(n)}]/N$, $i,j = 1,2,\ldots,m$ and $d_n \varepsilon (-1,1)$ is chosen in such a way that $\theta_{(n)}$ minimizes $\ell^*(\theta)$, given in Equation (5).

We now present an empirical example to illustrate our procedure. We take Bodkin and Klein's (1967, p. 33) model

$$x_{i1} - \theta_{11} 10^{\theta_{12} x_{i2}} \left[\theta_{13} y_{i1}^{-\theta_{14}} + (1-\theta_{13}) y_{i2}^{-\theta_{14}} \right]^{-\theta_{15}/\theta_{14}} = u_{i1} \quad (27)$$

and

$$x_{i3} - \frac{\theta_{13}}{1-\theta_{13}} \left[\frac{y_{i1}}{y_{i2}} \right]^{-(1+\theta_{14})} = u_{i2} \quad (28)$$

where x_1 is real output, x_2 is time, x_3 is the ratio of price of capital services and wage rate, y_1 is the capital input and y_2 is the labor input. This model was also considered by Bard (1974, p. 133) and we use the same data. This particular model does not fit into our specification of NLSES since θ_{13} and θ_{14} appear in both the equations. This makes our model overidentifeid and this will be clear from the equation (31).

After linearizing around $y_1 = y_2 = x_2 = x_3 = 1$ and $x_2 = 0$, from (27) and (28), we have

$$y_1 = \delta_{1o} + \delta_{11}x_1 + \delta_{12}x_2 + \gamma_{12}y_2 + \bar{r}_1 + u_1 \qquad (29)$$

and

$$y_2 - y_1 = \delta_{2o} + \delta_{21}x_3 + \bar{r}_2 + u_2 \qquad (30)$$

where

$$\delta_{1o} = -(1-\theta_{15})/\theta_{15}\theta_{13}, \ \delta_{11} = 1/\theta_{11}\theta_{15}\theta_{13}$$

$\delta_{12} = -\theta_{12} \ln 10/\theta_{15}\theta_{13}$, $\gamma_{12} = -(1-\theta_{13})/\theta_{13}$, $\delta_{2o} = -1/(1+\theta_{14})$
$\delta_{21} = (1-\theta_{13})/\{\theta_{13}(1+\theta_{14})\}$ and $\bar{r}_1$, $\bar{r}_2$, u_1, u_2 have usual meanings. From (29) and (30) it is easily seen that both the equations are identified with the following θ to β correspondence:

$$\theta_{11} = (1-\gamma_{12}-\delta_{1o})/\delta_{11}, \ \theta_{12} = -\delta_{12}/\{\ln 10(1-\gamma_{12}-\delta_{1o})\}$$

$$\theta_{13} = 1/(1-\gamma_{12}) \text{ or } \delta_{2o}/(\delta_{2o}-\delta_{21})$$

$$\theta_{14} = -(1+\delta_{2o})/\delta_{2o} \text{ and } \theta_{15} = (1-\gamma_{12})/(1-\gamma_{12}-\delta_{1o}). \qquad (31)$$

To estimate u_1 we used the first value of θ_{13} and the second one was used to estimate u_2.

Bard (1974) derived the FIML estimator by minimizing the generalized residual variance in the reduced-form equations. This was possible because for this particular model both the equations can be solved explicitly to get reduced-form equations and there is a one-to-one correspondence between the reduced-form and the structural parameters. His procedure was equivalent to maximizing the concentrated log-likelihood function, derived from the reduced-form equations, assuming that the reduced-form errors are also normal (which is a strong assumption). Our concentrated log-likelihood function $\ell^*(\theta)$ was derived by assuming normality for structural errors, but there were two difficulties: firstly, the initial values that we got by putting the remainder terms as zero were far away from the optimum values; secondly, $\ell^*(\theta)$ was seen having a number of local maxima. Therefore, it was essential to choose initial values lying in the immediate neighborhood of the optimum values, and using Bard's results we took the following initial values: $\delta_{1o} = .4$, $\delta_{11} = 2.5$, $\delta_{12} = -.02$, $\gamma_{12} = -.7$, $\delta_{2o} = -.5$ and $\delta_{21} = .4$. In Table 5 we report the results for two cases: general and diagonal Σ. Estimates in brackets are due to Bard (1974, p. 138). Iteration schemes converged with small number iterations, and the Bard's and our estimates are very close. The values of the concentrated log-likelihood functions obtained by putting Bard's estimates in $\ell^*(\theta)$ are slightly higher than those of ours. This slight difference may be due to the flatness of $\ell^*(\theta)$ in the neighborhood of the optimum values.

Insert Table 5 Here

As noted earlier our $\ell^*(\theta)$ is based on the normality of the disturbances u_{i1} and u_{i2}. For the validity of our estimates, we test this assumption using four test statistics, $C_j(j=1,2,3,4)$ of Bera and John (1983). The statistics values along with their corresponding degrees of freedom and finite sample critical points are given in Table 6. We observed that all the statistics are less than their corresponding critical values except C_2 which exceeds the 10 percent critical value. However, this rejection of normality is not at all strong, and the omnibus test statistics C_3 and C_4 do not reject normality. Therefore, on the basis of the above results, we may conclude that the disturbances in the Bodkin and Klein's model follow a bivariate normal distribution.

Insert Table 6 Here

7. Concluding Remarks

A method has been suggested for the estimation of NLSES which is computationally straightforward and is capable of producing consistent estimates. This was achieved by linearization on variables rather than parameters, as is the case with other algorithms. In certain cases, as our simulation study shows, this procedure can also provide more efficient estimates than NL2SLS. Where it gives less efficient estimates, these can be used as starting values for more efficient procedures such as FIML. However, our procedure is not applicable to all NLSES. It is applicable to systems that satisfy our identification condition which requires a correspondence between the structural parameters and the linearized parameters. Further work needs to be done to characterize the class of functions where we can apply this estimation procedure. Also it would be interesting to study the properties of the linearized estimator analytically and compare them with those of the commonly used estimators.

FOOTNOTES

[1]Amemiya (1977) asserted that consistency of $\hat{\theta}_{ML}$ crucially depends on Assumption 4. However, through a counter example, Phillips (1982) demonstrated that normality is not necessary for $\hat{\theta}_{ML}$ to be consistent.

[2]For simplicity, we assume that $\Gamma_{jj} = 1$ for all j. However, our analysis can be modified to the case when $\Gamma_{jj} \neq 1$, but that will introduce some complicated algebra.

[3]If such a function does not exist, we can call the equation under-identified. This can be tackled by taking a higher order approximation. When θ_j is not uniquely recoverable there will be some loss of information unless we impose some restrictions on γ_j and δ_j.

[4]Strictly speaking our definition has much to do with estimability rather than identifiability.

[5]From the expression (22) it is clear that d_n may not always lie in the interval (0,1). Therefore, it is worthwhile to check whether at each interation step $d_n \in (0,1)$.

REFERENCES

Amemiya, T., 1974, "The nonlinear two-stage least square estimator," Journal of Econometrics 2, 105-110.

Amemiya, T., 1977, "The maximum likelihood estimator and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model," Econometrica 45, 955-968.

Bard, Y., 1974, Nonlinear Parameter Estimation (Academic Press, New York).

Bera, A. K. and S. John, 1983, "Test for multivariate normality with Pearson alternatives," Communication in Statistics, Theory and Method 12, 103-117.

Bodkin, R. G. and L. R. Klein, 1967, "Nonlinear estimation of aggregate production functions," Review of Economics and Statistics 49, 28-44.

Bowden, R. J., 1978, The Econometrics of Disequilibrium (North Holland, Amsterdam).

Bowden, R. J. and D. A. Turkington, 1981, "A comparative study of instrumental variable estimators for nonlinear simultaneous models," Journal of the American Statistical Association 76, 988-995.

Byron, R. P. and A. K. Bera, 1983, "Linearized estimation of nonlinear single equation functions," International Economic Review 24, 237-248.

Edgerton, D. L., 1972, "Some properties of two-stage least squares as applied to nonlinear models," International Economic Review 13, 26-32.

Fair, R. C. and W. R. Parke, 1980, "Full information estimates of a nonlinear macroeconometric model," Journal of Econometrics 13, 269-291.

Fisher, F. M., 1966, The Identification Problem in Econometrics (McGraw-Hill, New York).

Gallant, A. R., 1975, "Nonlinear regression," American Statistican 29, 73-81.

Gallant, A. R., 1977, "Three-stage least squares estimation for a system of simultaneous, nonlinear, implicit equations," Journal of Econometrics 5, 71-88.

Goldfeld, S. M. and R. E. Quandt, 1972, Nonlinear Methods in Econometrics (North Holland, Amsterdam).

Hausman, J., 1975, "An instrumental variable approach to full information estimators for linear and certain nonlinear econometric models," Econometrica 43, 727-738.

Kelejian, H., 1971, "Two-stage least squares and econometric systems linear in parameters but nonlinear in variables," Journal of American Statistical Association 66, 373-374.

Phillips, P. C. B., 1982, "On the consistency of nonlinear FIML," Econometrica 50, 1307-1324.

Table 1

Mean Biases, Mean Square Errors and Variances of Different Estimators[a]
(Sample Size 30 and $V(u) = \Sigma_1$)

	Parameter $\theta_{21} = .5$			Parameter $\theta_{22} = 15.0$		
Method	Mean Bias	MSE	Var	Mean Bias	MSE	Var
			$V(x) = 1.0$			
OLS	$-.19292^{-2}$	$.15700^{-2}$	$.15663^{-2}$	$.62420^{-2}$	$.54550^{-3}$	$.50537^{-3}$
NIV	$.10139^{-3}$	$.82673^{-2}$	$.82673^{-2}$	$.65134^{-2}$	$.19112^{-2}$	$.18750^{-2}$
NL2SLS1	$.50298^{-3}$	$.90311^{-2}$	$.90308^{-2}$	$.60313^{-2}$	$.20065^{-2}$	$.19678^{-2}$
NL2SLS2	$-.78258^{-3}$	$.73380^{-2}$	$.73374^{-2}$	$.60244^{-2}$	$.16365^{-2}$	$.15991^{-2}$
LLI	$.11639^{-2}$	$.31187^{-2}$	$.31173^{-2}$	$.58906^{-2}$	$.16899^{-2}$	$.16553^{-2}$
			$V(x) = 3.0$			
OLS	$-.11866^{-2}$	$.65210^{-3}$	$.65070^{-3}$	$.64006^{-2}$	$.54550^{-3}$	$.50781^{-3}$
NIV	$.17347^{-3}$	$.12135^{-2}$	$.12134^{-2}$	$.64285^{-2}$	$.73624^{-3}$	$.69092^{-3}$
NL2SLS1	$-.88775^{-3}$	$.16402^{-2}$	$.16394^{-2}$	$.66651^{-2}$	$.86594^{-3}$	$.82031^{-3}$
NL2SLS2	$.25779^{-4}$	$.11698^{-2}$	$.11698^{-2}$	$.64050^{-2}$	$.69427^{-3}$	$.65430^{-3}$
LLI	$-.54645^{-3}$	$.78662^{-3}$	$.78632^{-3}$	$.66221^{-2}$	$.78583^{-3}$	$.74219^{-3}$
			$V(x) = 5.0$			
OLS	$-.68546^{-3}$	$.28220^{-3}$	$.28173^{-3}$	$.63074^{-2}$	$.49972^{-3}$	$.46142^{-3}$
NIV	$.29337^{-3}$	$.42634^{-3}$	$.42625^{-3}$	$.62972^{-2}$	$.54932^{-3}$	$.51025^{-3}$
NL2SLS1	$-.87132^{-3}$	$.66057^{-3}$	$.65982^{-3}$	$.67772^{-2}$	$.71335^{-2}$	$.66894^{-3}$
NL2SLS2	$.13958^{-3}$	$.41520^{-3}$	$.41518^{-3}$	$.64042^{-2}$	$.54932^{-3}$	$.50537^{-3}$
LLI	$-.54841^{-3}$	$.31759^{-3}$	$.31729^{-3}$	$.67489^{-2}$	$.69427^{-3}$	$.64941^{-3}$

[a]Power refers to 10^{-K}. Thus, $-.19292^{-2} = -.0019292$.

Table 2

Mean Biases, Mean Square Errors and Variances of Different Estimators (Sample Size 30 and $V(u) = \Sigma_2$)

	Parameter $\theta_{21} = .5$			Parameter $\theta_{22} = 15.0$		
Method	Mean Bias	MSE	Var	Mean Bias	MSE	Var
			$V(x) = 1.0$			
OLS	$-.10503^{-1}$	$.80704^{-2}$	$.79601^{-2}$	$.16558^{-1}$	$.25329^{-2}$	$.22558^{-2}$
NIV	$.16839^{-2}$	$.41680^{-1}$	$.41677^{-1}$	$.13787^{-1}$	$.94223^{-2}$	$.92310^{-2}$
NL2SLS1	$.27535^{-2}$	$.45545^{-1}$	$.45537^{-1}$	$.12643^{-1}$	$.99335^{-2}$	$.97729^{-2}$
NL2SLS2	$.37258^{-2}$	$.36987^{-1}$	$.36973^{-1}$	$.12397^{-1}$	$.79765^{-2}$	$.78247^{-2}$
LLI	$-.55730^{-3}$	$.15819^{-1}$	$.15818^{-1}$	$.12869^{-1}$	$.81940^{-2}$	$.80249^{-2}$
			$V(x) = 3.0$			
OLS	$-.56355^{-2}$	$.33215^{-2}$	$.32898^{-2}$	$.15977^{-1}$	$.24681^{-2}$	$.22095^{-2}$
NIV	$.88011^{-3}$	$.61016^{-2}$	$.61008^{-2}$	$.13967^{-1}$	$.33379^{-2}$	$.31396^{-2}$
NL2SLS1	$-.15642^{-2}$	$.82580^{-2}$	$.82555^{-2}$	$.14554^{-1}$	$.40780^{-2}$	$.38647^{-2}$
NL2SLS2	$.35951^{-3}$	$.58895^{-2}$	$.58894^{-2}$	$.14106^{-1}$	$.32120^{-2}$	$.30151^{-2}$
LLI	$-.29733^{-2}$	$.39946^{-2}$	$.39858^{-2}$	$.14683^{-1}$	$.37422^{-2}$	$.35254^{-2}$
			$V(x) = 5.0$			
OLS	$-.32147^{-2}$	$.14307^{-2}$	$.14204^{-2}$	$.15337^{-1}$	$.22545^{-2}$	$.20166^{-2}$
NIV	$.11015^{-2}$	$.21437^{-2}$	$.21427^{-2}$	$.13703^{-1}$	$.25520^{-2}$	$.23657^{-2}$
NL2SLS1	$-.16729^{-2}$	$.33246^{-2}$	$.33218^{-2}$	$.14852^{-1}$	$.32921^{-2}$	$.30713^{-2}$
NL2SLS2	$.49131^{-3}$	$.20831^{-2}$	$.20828^{-2}$	$.14037^{-1}$	$.25063^{-2}$	$.23096^{-2}$
LLI	$-.22256^{-2}$	$.16128^{-2}$	$.16078^{-2}$	$.14911^{-1}$	$.31052^{-2}$	$.28833^{-2}$

Table 3

Mean Biases, Mean Square Errors and Variances of Different Estimators
(Sample Size 80 and $V(u) = \Sigma_1$)

	Parameter θ_{21} = .5			Parameter θ_{22} = 15.0		
Method	Mean Bias	MSE	Var	Mean Bias	MSE	Var
			V(x) = 1.0			
OLS	$.19873^{-3}$	$.40716^{-3}$	$.40712^{-3}$	$.69671^{-2}$	$.22125^{-3}$	$.17334^{-3}$
NIV	$.34398^{-2}$	$.30205^{-2}$	$.30087^{-2}$	$.63680^{-2}$	$.74768^{-3}$	$.71045^{-3}$
NL2SLS1	$.36037^{-2}$	$.33429^{-2}$	$.33300^{-2}$	$.62520^{-2}$	$.79727^{-3}$	$.75928^{-3}$
NL2SLS2	$-.28821^{-2}$	$.29791^{-2}$	$.29708^{-2}$	$.88470^{-2}$	$.77057^{-3}$	$.69580^{-3}$
LLI	$.42855^{-2}$	$.93436^{-3}$	$.91600^{-3}$	$.59782^{-2}$	$.64850^{-3}$	$.61279^{-3}$
			V(x) = 3.0			
OLS	$.50887^{-4}$	$.23137^{-3}$	$.23136^{-3}$	$.70786^{-2}$	$.24795^{-3}$	$.19775^{-3}$
NIV	$.55303^{-3}$	$.48720^{-3}$	$.48690^{-3}$	$.75234^{-2}$	$.34332^{-3}$	$.28564^{-3}$
NL2SLS1	$.63360^{-4}$	$.65392^{-3}$	$.65392^{-3}$	$.77949^{-2}$	$.39673^{-3}$	$.33691^{-3}$
NL2SLS2	$-.44966^{-3}$	$.51129^{-3}$	$.51109^{-3}$	$.81419^{-2}$	$.35858^{-3}$	$.29541^{-3}$
LLI	$.10783^{-2}$	$.26003^{-3}$	$.25887^{-3}$	$.76518^{-2}$	$.36240^{-3}$	$.30029^{-3}$
			V(x) = 5.0			
OLS	$.78395^{-4}$	$.10448^{-3}$	$.10447^{-3}$	$.70559^{-2}$	$.21744^{-3}$	$.16601^{-3}$
NIV	$.25976^{-3}$	$.16436^{-3}$	$.16430^{-3}$	$.76008^{-2}$	$.26321^{-3}$	$.20508^{-3}$
NL2SLS1	$-.35500^{-3}$	$.25984^{-3}$	$.25972^{-3}$	$.81215^{-2}$	$.31662^{-3}$	$.25146^{-3}$
NL2SLS2	$-.17136^{-3}$	$.17875^{-3}$	$.17871^{-3}$	$.81043^{-2}$	$.28610^{-3}$	$.22217^{-3}$
LLI	$.37276^{-3}$	$.10567^{-3}$	$.10553^{-3}$	$.80405^{-2}$	$.31281^{-3}$	$.24902^{-3}$

Table 4

Mean Biases, Mean Square Errors and Variances of Different Estimators
(Sample Size 80 and $V(u) = \Sigma_2$)

	Parameter $\theta_{21} = .5$			Parameter $\theta_{22} = 15.0$		
Method	Mean Bias	MSE	Var	Mean Bias	MSE	Var
			V(x) = 1.0			
OLS	$-.47728^{-2}$	$.20548^{-2}$	$.20320^{-2}$	$.17791^{-1}$	$.10071^{-2}$	$.69092^{-3}$
NIV	$.80783^{-2}$	$.15017^{-1}$	$.14951^{-1}$	$.13806^{-1}$	$.34752^{-2}$	$.32886^{-2}$
NL2SLS1	$.83934^{-2}$	$.16636^{-1}$	$.16566^{-1}$	$.13590^{-1}$	$.37880^{-2}$	$.36060^{-2}$
NL2SLS2	$-.71600^{-2}$	$.14865^{-1}$	$.14814^{-1}$	$.20332^{-1}$	$.36888^{-2}$	$.32764^{-2}$
LLI	$.63167^{-2}$	$.46346^{-2}$	$.45944^{-2}$	$.13382^{-1}$	$.29793^{-2}$	$.28027^{-2}$
			V(x) = 3.0			
OLS	$-.23280^{-2}$	$.11451^{-2}$	$.11397^{-2}$	$.17188^{-1}$	$.98419^{-3}$	$.68603^{-3}$
NIV	$.14605^{-2}$	$.24199^{-2}$	$.24178^{-2}$	$.16438^{-1}$	$.14305^{-2}$	$.11621^{-2}$
NL2SLS1	$.35032^{-3}$	$.32560^{-2}$	$.32558^{-2}$	$.17087^{-1}$	$.17509^{-2}$	$.14600^{-2}$
NL2SLS2	$-.12523^{-2}$	$.25431^{-2}$	$.25415^{-2}$	$.18254^{-1}$	$.15335^{-2}$	$.12012^{-2}$
LLI	$.87410^{-3}$	$.13014^{-2}$	$.13006^{-2}$	$.16945^{-1}$	$.15526^{-2}$	$.12671^{-2}$
			V(x) = 5.0			
OLS	$-.11680^{-2}$	$.51508^{-3}$	$.51371^{-3}$	$.16743^{-1}$	$.86594^{-3}$	$.59082^{-3}$
NIV	$.73868^{-3}$	$.81465^{-3}$	$.81410^{-3}$	$.16625^{-1}$	$.10528^{-2}$	$.77637^{-3}$
NL2SLS1	$-.63256^{-3}$	$.12929^{-2}$	$.12925^{-2}$	$.17826^{-1}$	$.14038^{-2}$	$.10889^{-2}$
NL2SLS2	$-.33391^{-3}$	$.88811^{-3}$	$.88800^{-3}$	$.17837^{-1}$	$.11825^{-2}$	$.86182^{-3}$
LLI	$-.63106^{-5}$	$.53063^{-3}$	$.53063^{-3}$	$.17737^{-1}$	$.13123^{-2}$	$.99854^{-3}$

Table 5

LFI and Bard's Estimates for Bodkin and Klein's Model[a]

Cases	$\ell^*(\theta)$		Parameter Estimates					Number of Iterations
	Starting Value	After Convergence	θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}	
General Σ	196.29320	222.80308 (223.02315)	.55995 (.54600)	.00632 (.00636)	.55634 (.57520)	.99919 (.99800)	1.31564 (1.26500)	9
Diagonal Σ	196.27808	217.31200 (218.88407)	.55551 (.54170)	.00635 (.00577)	.55599 (.56290)	.99958 (.94600)	1.31508 (1.35700)	5

[a]Bard's estimates are in parenthesis.

Table 6

Results on Test for Multivariate Normality

	Test Statistics	Degrees of Freedom	Finite Sample Critical Points	
			10 Percent	5 Percent
C_1	1.5159	2	3.6050	4.6842
C_2	4.8899	3	4.3500	5.4358
C_3	4.0614	4	5.6638	6.9087
C_4	6.4059	5	6.7914	8.1364