Journal of Econometrics 2 (1974) 105–110. © North-Holland Publishing Company

THE NONLINEAR TWO-STAGE LEAST-SQUARES ESTIMATOR

Takeshi AMEMIYA*

Stanford University, Stanford, Calif. 94305, U.S.A.

Received August 1973

1. Introduction

In this paper we consider estimation of the parameters of a single equation of a simultaneous equations model which is nonlinear both in variables and paarmeters. Such a model has never been analyzed in the literature to the best of our knowledge. Models in which the nonlinearity appears only in variables or only in parameters have been previously considered. For the former case see Kelejian (1971) and other references cited in Goldfeld and Quandt (1972), and for the latter case see, for example, Zellner, Huang and Chau (1965.)

We define the nonlinear two-stage least-squares estimator (NL2SLS) for our model and derive its asymptotic distribution. Our estimator reduces to the NL2SLS of Kelejian if the nonlinearity exists only in variables, to the NL2SLS of Zellner and others if the nonlinearity exists only in parameters, and to the usual 2SLS estimator if the regression function is linear both in variables and parameters. We show that the well-known optimality properties of 2SLS extend to NL2SLS in the model that is linear in variables and nonlinear in parameters. The question of whether they extend to NL2SLS in the general nonlinear model is left for further study.

2. Main results

We consider the nonlinear regression equation

$$y_t = f(z_t, \beta) + u_t, \tag{1}$$

where y_t is a scalar random variable, u_t is a scalar random variable with zero mean and constant variance σ^2, z_t is an H-component vector consisting partly of endogenous variables (that is, random variables correlated with u_t) and partly of exogenous variables (that is, known constants), β is a G-component

*This research was supported by National Science Foundation Grant GS-2635 to the Institute for Mathematical Studies in the Social Sciences, Stanford University. The author wishes to thank Arthur Goldberger for helpful suggestions.

vector of unknown parameters, and f is a possibly nonlinear function in both z and β having continuous first and second derivatives with respect to β. Our aim is to estimate β and σ^2 on the basis of observations on y_t and z_t, $t = 1, 2, \ldots, T$. We may regard eq. (1) as a single equation of a complete nonlinear simultaneous equations model which generates all the endogenous variables that appear in eq. (1).

We define the nonlinear two-stage least-squares estimator (NL2SLS) as follows:

Definition. The NL2SLS estimator of β in model (1), denoted $\hat{\beta}$, is the value of β that minimizes

$$\Phi(\beta) = (y-f)'X(X'X)^{-1}X'(y-f), \tag{2}$$

where y and f are T-component vectors whose tth elements are y_t and $f(z_t, \beta)$, respectively, and X is a $T \times K$ matrix of certain constants with rank K.

We have not specified X because we will consider it later. It is clear that for appropriate choices of X, the NL2SLS defined above reduces to Kelejian's NL2SLS, Zellner's NL2SLS, and Theil's 2SLS. In Kelejian's case, X consists of the low-order polynomials of all the exogenous variables of the system, and in Zellner's and Theil's cases, X consists of all the exogenous variables of the system.

We prove the following theorem in appendix 1:

Theorem. Let $\hat{\beta}$ be the NL2SLS estimator defined above. Then:

(i) $\hat{\beta}$ converges in probability to the true value β_0,

and

(ii) $\sqrt{T}(\hat{\beta}-\beta_0)$ converges in distribution to

$$N\left\{0, \sigma^2\left[\operatorname{plim} \frac{1}{T}\frac{\partial f'}{\partial \beta}\bigg|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f}{\partial \beta'}\bigg|_{\beta_0}\right]^{-1}\right\},$$

if the following assumptions are satisfied:

(A) The parameter space is compact.

(B) $\{u_t\}$ is i.i.d.

(C) $\lim (1/T)X'X$ exists and is nonsingular.

(D) $(1/T)(\partial f'/\partial\beta)X$ converges in probability to a constant matrix of rank G uniformly in β.

(E) $(1/T)(\partial^2 f'/\partial\beta_i\partial\beta)X$ converges in probability to a constant matrix uniformly in β for $i = 1, 2, \ldots, G$, where β_i is the ith element of β.

The minimization of eq. (2) must be done by an iterative procedure. The standard procedure is the Gauss–Newton method defined by

$$\hat{\beta}_{(n)} = \hat{\beta}_{(n-1)} + \left[\frac{\partial f'}{\partial \beta} X(X'X)^{-1}X' \frac{\partial f}{\partial \beta'}\right]^{-1} \frac{\partial f'}{\partial \beta} X(X'X)^{-1}X'(y-f), \quad (3)$$

where f and $\partial f'/\partial \beta$ in the right-hand side are evaluated at $\hat{\beta}_{(n-1)}$, the estimate obtained in the $(n-1)$th iteration. Note that in Theil's and Kelejian's cases the right-hand side becomes independent of $\hat{\beta}_{(n-1)}$ and eq. (3) reduces to Theil's 2SLS and Kelejian's NL2SLS respectively. For the convergence properties of the Gauss–Newton method in general and for its various modifications see Draper and Smith (1966) or Jacoby, Kowalik, and Pizzo (1972).

3. The efficiency

It is well-known that in the linear case 2SLS has the same asymptotic distribution as the limited information maximum likelihood estimator (LIML) and has the smallest (in matrix sense) asymptotic variance–covariance matrix in the class of instrumental variables method estimators. In our general nonlinear model we cannot make similar statements since it is usually difficult to evaluate the asymptotic variance–covariance matrix of the maximum likelihood estimator or to find a meaningful criterion by which to choose the optimal X. These same difficulties apply in Kelejian's case as well. Thus, the most we can do is merely to say, like Kelejian, that it may be a good idea to let X consist of the low-order polynomials of all the exogenous variables of the system. This point deserves further research.

We will show below that we *can* extend the well-known efficiency properties of 2SLS to the case where the regression function is nonlinear only in the parameters.

In this case eq. (1) becomes, in vector notation,

$$y = Z\alpha(\beta) + u, \quad (4)$$

where Z is the matrix whose tth row is z_t, α is an H-component vector ($H > G$) each element of which is a function of β. From (ii) of the Theorem, the asymptotic variance–covariance matrix of NL2SLS is

$$\sigma^2 \left[\left.\frac{\partial \alpha'}{\partial \beta}\right|_{\beta_0} \operatorname{plim} \frac{Z'X(X'X)^{-1}X'Z}{T} \left.\frac{\partial \alpha}{\partial \beta'}\right|_{\beta_0}\right]^{-1}. \quad (5)$$

Hence, it is obvious that (5) is minimized (in matrix sense) when X is the matrix of all the exogenous variables of the system, denoted X_0.

The log likelihood function concentrated in β is

$$\log L = \text{const.} + \frac{T}{2} \log \frac{(y-Z\alpha)'M(y-Z\alpha)}{(y-Z\alpha)'(y-Z\alpha)}, \quad (6)$$

where $M = I - X_0(X_0'X_0)^{-1}X_0'$. The LIML estimator of β, denoted β^*, is the value of β that maximizes eq. (6). The first and second derivatives of $\log L$ with respect to β are given in appendix 2. Using the same line of proof as in appendix 1, we can prove that under the same assumptions as those of the Theorem, the limit distribution of $\sqrt{T}(\beta^* - \beta_0)$ is normal with zero mean and variance–covariance matrix equal to eq. (5). Thus, NL2SLS has the same asymptotic distribution as LIML in this case.

Appendix 1

Proof of the theorem

We need the following lemma, which is proved in Amemiya (1972).

Lemma. Let $Q_T(\omega, \theta)$ be a measurable function on a measurable space Ω and for each ω in Ω a continuous function for θ in a compact set Θ. If $Q_T(\omega, \theta)$ converges to $Q(\theta)$ in probability uniformly for all θ in Θ, and if $\hat{\theta}_T(\omega)$ converges to θ_0 in probability, then $Q_T[\omega, \hat{\theta}_T(\omega)]$ converges to $Q(\theta_0)$ in probability.

Proof of (i). By a Taylor expansion we have

$$X'f(\hat{\beta}) - X'f(\beta_0) = X'\frac{\partial f}{\partial \beta'}\bigg|_{\beta^+}(\hat{\beta} - \beta_0), \tag{7}$$

where β^+ lies between $\hat{\beta}$ and β_0. Substituting $y-u$ for $f(\beta_0)$ and premultiplying eq. (7) by $(TX'X)^{-\frac{1}{2}}$, we obtain

$$v + w = (TX'X)^{-\frac{1}{2}}X'\frac{\partial f}{\partial \beta'}\bigg|_{\beta^+}(\hat{\beta} - \beta_0),$$

where

$$v = (TX'X)^{-\frac{1}{2}}X'u$$

and

$$w = (TX'X)^{-\frac{1}{2}}[X'f(\hat{\beta}) - X'y].$$

But plim $v = 0$ because of assumptions (B) and (C), and plim $w = 0$ because

$$w'w = \frac{1}{T}\Phi(\hat{\beta}) \leqq \frac{1}{T}\Phi(\beta_0) = v'v.$$

Therefore, plim $\hat{\beta} = \beta_0$ because of assumptions (C) and (D).

Proof of (ii). We have

$$\sqrt{T}(\hat{\beta} - \beta_0) = -\left[\frac{1}{T}\frac{\partial^2 \Phi}{\partial \beta \partial \beta'}\bigg|_{\bar{\beta}}\right]^{-1}\frac{1}{\sqrt{T}}\frac{\partial \Phi}{\partial \beta}\bigg|_{\beta_0}, \tag{8}$$

where $\tilde{\beta}$ lies between $\hat{\beta}$ and β_0. We also have

$$\frac{\partial \Phi}{\partial \beta} = -2\frac{\partial f'}{\partial \beta}X(X'X)^{-1}X'(y-f) \tag{9}$$

and

$$\frac{\partial^2 \Phi}{\partial \beta \partial \beta'} = 2\frac{\partial f'}{\partial \beta}X(X'X)^{-1}X'\frac{\partial f}{\partial \beta'} - 2A, \tag{10}$$

where A is the matrix whose ith row is

$$(y-f)'X(X'X)^{-1}X'\frac{\partial^2 f}{\partial \beta_i \partial \beta'}.$$

Therefore, by assumptions (B), (C) and (D), using Corollary 2.6.1 of Anderson (1971),

$$-\frac{1}{2}\frac{1}{\sqrt{T}}\frac{\partial \Phi}{\partial \beta}\bigg|_{\beta_0} \to N\left[0, \sigma^2 \operatorname{plim} \frac{1}{T}\frac{\partial f'}{\partial \beta}\bigg|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f}{\partial \beta'}\bigg|_{\beta_0}\right] \tag{11}$$

and by assumptions (B), (C), (D) and (E),

$$\frac{1}{2}\operatorname{plim}\frac{1}{T}\frac{\partial^2 \Phi}{\partial \beta \partial \beta'}\bigg|_{\beta_0} = \operatorname{plim}\frac{1}{T}\frac{\partial f'}{\partial \beta}\bigg|_{\beta_0} X(X'X)^{-1}X'\frac{\partial f}{\partial \beta'}\bigg|_{\beta_0}. \tag{12}$$

But, by assumptions (A), (B), (C), (D) and (E), the consistency of $\hat{\beta}$, and the Lemma, we have

$$\operatorname{plim}\frac{1}{T}\frac{\partial^2 \Phi}{\partial \beta \partial \beta'}\bigg|_{\tilde{\beta}} = \operatorname{plim}\frac{1}{T}\frac{\partial^2 \Phi}{\partial \beta \partial \beta'}\bigg|_{\beta_0}. \tag{13}$$

Thus, (ii) follows from eqs. (8), (11), (12) and (13).

Appendix 2

The first and second derivatives of eq. (6)

From eq. (6) we have

$$\frac{\partial \log L}{\partial \beta} = \frac{T}{(y-Z\alpha)'M(y-Z\alpha)}\frac{\partial \alpha'}{\partial \beta}(Z'MZ\alpha - Z'My)$$

$$-\frac{T}{(y-Z\alpha)'(y-Z\alpha)}\frac{\partial \alpha'}{\partial \beta}(Z'Z\alpha - Z'y) \tag{14}$$

and

$$\frac{\partial^2 \log L}{\partial\beta\partial\beta'} = \frac{T}{[(y-Z\alpha)'M(y-Z\alpha)]^2}\left\{(y-Z\alpha)'M(y-Z\alpha)\right.$$
$$\times\left[\frac{\partial\alpha'}{\partial\beta}Z'MZ\frac{\partial\alpha}{\partial\beta'}-B\right]$$
$$\left.-2\frac{\partial\alpha'}{\partial\beta}Z'M(y-Z\alpha)(y-Z\alpha)'MZ\frac{\partial\alpha}{\partial\beta'}\right\} \tag{15}$$
$$-\frac{T}{[(y-Z\alpha)'(y-Z\alpha)]^2}\left\{(y-Z\alpha)'(y-Z\alpha)\right.$$
$$\left.\times\left[\frac{\partial\alpha'}{\partial\beta}Z'Z\frac{\partial\alpha}{\partial\beta'}-C\right]-2\frac{\partial\alpha'}{\partial\beta}Z'(y-Z\alpha)(y-Z\alpha)'Z\frac{\partial\alpha}{\partial\beta'}\right\},$$

where B is the matrix whose ith column is

$$\frac{\partial^2\alpha'}{\partial\theta_i\partial\theta}Z'M(y-Z\alpha),$$

and C is the matrix whose ith column is

$$\frac{\partial^2\alpha'}{\partial\theta_i\partial\theta}Z'(y-Z\alpha).$$

References

Amemiya, T., 1972, Regression analysis when the dependent variable is truncated normal, Technical Report no. 59 (Institute for Mathematical Studies in the Social Sciences, Stanford University) scheduled to appear in Econometrica.

Anderson, T.W., 1971, The statistical analysis of time series (John Wiley, New York).

Draper, N.R. and H. Smith, 1966, Applied regression analysis (John Wiley, New York).

Goldfeld, S.M. and R.E. Quandt, 1972, Nonlinear methods in econometrics (North-Holland, Amsterdam).

Jacoby, S.L.S., J.S. Kowalik and J.T. Pizzo, 1972, Iterative methods for nonlinear optimization problems (Prentice-Hall, Englewood Cliffs, N.J.).

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

Zellner, A., D.S. Huang and L.C. Chau, 1965, Further analysis of the short run consumption function with emphasis on the role of liquid assets, Econometrica 33, 571–581.