A NOTE ON AMEMIYA'S NONLINEAR TWO-STAGE LEAST SQUARES ESTIMATORS*

Richard E. Quandt

Econometric Research Program
Princeton University
Research Memorandum No. 178

May 1975

A NOTE ON AMEMIYA'S NONLINEAR TWO-STAGE LEAST SQUARES ESTIMATORS*

by

Richard E. Quandt

1. Introduction

Consider the estimation of any or all the equations in a system given by

$$y_{it} = f_{it}(Y_{it}, \theta_i) + u_{it} \quad \begin{array}{l} i=1,\ldots,g \\ t=1,\ldots,T \end{array} \tag{1-1}$$

where i indexes equations and endogenous variables, t indexes observations, y_{it} and u_{it} are random variables, Y_{it} is the vector of random variables $(y_{1t},\ldots,y_{i-1,t}, y_{i+1,t},\ldots,y_{gt})$ and θ_i is a vector of parameters. Two-stage least squares estimators for θ_i have been suggested in [3] and [4] for systems nonlinear in the variables but linear in the parameters. The suggested approach estimates θ_i by first regressing the endogenous functions appearing on the right hand side of (1-1) on some polynomials in the exogenous variables and using the predictions for these endogenous functions from the first stage in a second stage in which $\sum_t (y_{it} - f_{it}(\hat{Y}_{it}, \theta_i))^2$ is minimized; a procedure analogous to TSLS in the linear case. For general nonlinear systems, Amemiya introduced the general nonlinear two-stage least squares estimator [2]. Writing y_i and f_i for the vectors containing y_{it} and f_{it} as elements, Amemiya's class of estimators is obtained by finding the θ_i that minimizes

$$(y_i - f_i)'D(D'D)^{-1}D'(y_i - f_i) \tag{1-2}$$

where D is a matrix of suitably chosen constants. Various members of the

* I am indebted to NSF Grant GS43747X for support.

class are obtained by defining D in alternative ways. Possible choices for D are (a) $D = X$, the matrix of exogenous variables in the entire system of equations. This yields what is called the standard nonlinear TSLS estimator (SNTSLS). (b) $D = E(\partial f_i/\partial\theta)'$, which is denoted as the best nonlinear TSLS (BNTSLS) estimator. Other possibilities are (c) to employ as the columns of D not only the columns of X but also higher order polynomials of these columns. Depending on the degree of the polynomial, these estimates will be referred to as POLY-2, POLY-3 and POLY-4. Amemiya further devises (d) the nonlinear limited information maximum likelihood estimator which is the standard LIML estimator (NLIML) for θ_i, the coefficients of the ith equation, on the assumption that the equations other than the ith are linear and given by $Y_i = X\Pi + V_i$. A final estimator of interest is (e) the modified nonlinear LIML estimator (MNLIML) which results from the first step in an iterative computation of the nonlinear LIML. It is shown in [1] that the asymptotic covariance matrices cov( ) of all these estimators obey

$$\text{cov(SNTSLS)} > \text{cov(BNTSLS)} > \text{cov(MNLIML)} > \text{cov(NLIML)}$$

where $\text{cov}(*) > \text{cov}(**)$ means that $\text{cov}(*) - \text{cov}(**)$ is a positive semidefinite matrix.

The purpose of this note is to examine the finite sample behavior of several of these estimators and to compare them with the full information maximum likelihood estimator and the ordinary (nonlinear) least squares estimator in a particular two equation model in which one equation is nonlinear only in variables, and the other is nonlinear in both variables and coefficients. The estimators explicitly computed and compared are (1) OLS which is the estimator obtained by minimizing the residual sum of squares without attending to the simultaneity (i.e., the estimator obtained by minimizing (1-2) with $D = I$) of the system; (2) SNTSLS, and other polynomial estimators where D alternately contains 1st, 2nd, 3rd or 4th degree polynomials of the columns of X (but no cross-product terms); (3) NLIML; (4) MNLIML; (5) FIML. The BNTSLS estimator is omitted from consideration since $E(\partial f/\partial\theta)'$ will often be difficult to compute and may not even exist.

2. Description of the Model

The structural equations of the model are

$$y_{1t} + b_1 \log y_{2t} + b_2 z_{1t} + b_3 = u_{1t} \tag{2-1}$$

$$z_{3t}^{b_4} y_{1t} + y_{2t} + b_5 z_{2t} = u_{2t} \tag{2-2}$$

where (u_{1t}, u_{2t}) is i.i.d. as $N(0,\Sigma)$. In order to avoid difficulties in generating observation as well as in obtaining estimates we wish to restrict the true values of the parameters so that, for all t, there exists a unique solution for y_{1t}, y_{2t} with y_{2t} positive. Eliminating y_{1t} from (2-1) and (2-2) we have

$$y_{2t} + k_{1t} \log y_{2t} + k_{2t} = 0 \tag{2-3}$$

where $k_{1t} = -b_1 z_{3t}^{b_4}$ and $k_{2t} = u_{1t}(z_{3t}^{b_4} - b_2 z_{1t} - b_3) + b_5 z_{2t} - u_{2t}$. A sufficient condition for our requirement is that $k_{1t} > 0$.[1] This in turn requires that $b_1 < 0$ and $z_{3t} > 0$ for all t. The values of the other coefficients may be set without regard to these considerations.

In each experiment a set of values was chosen for the exogenous variables. The latter are identical in repeated samples and were generated independently

[1] It is easy to verify that if $k_{1t} < 0$, there will exist two solutions, one solution, or no solution according to whether $\log(-k_{1t}) >, =,$ or $< -k_{2t}$.

from uniform distributions over specified ranges. The values of the structural parameters and the ranges of the exogenous variables as well as the sample size are given in Table 1. Each experiment or case consisted of 50 successful replications; on some occasions more than 50 replications were necessary to generate 50 successful ones because of computational failures.[2] It is clear from Table 1 that Case 2 is the same as Case 1 and Case 4 is the same as Case 3 except for sample size. Case 5 differs from Case 1 by having a true correlation between u_{1t} and u_{2t} of .84 rather than .43. Cases 6 and 7 differ in that selected b-coefficients have different values from Case 1. In Case 8 the range of z_3 is increased.

3. Results of Experiments

Very few computational failures were encountered in the experiments. If the computation of any method failed for whatever reason, the sample was discarded and a new one was generated. In nine instances, scattered over Cases 1, 2, 5 and 6, failures occurred in computing the polynomial estimators because a local minimum to (1-2) could not be achieved. FIML failed only in Case 5, in which case, however it failed 32 times as a result of apparently very flat segments of the likelihood function. Since Case 5 is the one in which the covariance matrix of errors is (intentionally) nearly singular, this is not very surprising. For the remaining cases the failure rate is negligible.

Mean square errors are displayed in Table 2. The superiority of FIML is evident. Out of 40 possible comparisons (8 cases, 5 coefficients), FIML has the lowest MSE in 31 instances. NLIML is next with 8 instances and MNLIML is best in only one instance. No other estimator ever has lowest MSE for any coefficient. On the other hand, FIML is worst in terms of MSE 4 times and NLIML 13 times. As expected OLS performs badly, being worst 15 times. SNTLS is

---

*See Section 3.

Table 1

Characteristics of Experiments
(Blanks indicate figures identical with corresponding figures in Case 1)

	Case 1	2	3	4	5	6	7	8
b_1	-6.0					-2.0		
b_2	2.0							
b_3	-2.0							
b_4	2.0						0.5	
b_5	-2.0							
σ_1^2	6.0							
σ_2^2	8.0							
σ_{12}	3.0				5.8			
Range z_1	(-10.0,5.0)							
Range z_2	(15.0,20.0)		(10.0,30.0)	(10.0,30.0)				
Range z_3	(0.8,1.2)							(0.7,1,3)
N	30	60		60				

worst 5 times, POLY-2 once and POLY-3 twice. Although FIML seems clearly best on these criteria, there is no obvious second best. NLIML is second in the number of times it is best but it is also second (to OLS) in the number of times it is worst, suggesting the frequent presence of outliers. POLY-4 is never best but also never worst. A somewhat similar conclusion emerges from Table 3. For each case and each replication we ranked the estimators in terms of their absolute deviations from the true value for each coefficient. These ranks were then averaged in each Case over the replications and (to save space in the table) over coefficients as well. These mean ranks are displayed in Table 3. They are obviously robust in comparison with MSE's and indicate that FIML has the lowest rank in every case. OLS has the worst mean rank in 5 Cases and NLIML in 3. The differences, however, are not staggering and even FIML is not that far ahead. In particular, the performance of the methods other than FIML seems bunched closely together. As a final generalization of this type we observe that higher order polynomial estimators tend to do somewhat better than the lower order ones, as can be confirmed from both Tables 2 and 3.

The comparison of Case 1 and 2 on the one hand and 3 and 4 on the other reveals the expected decrease of MSE's with increasing sample size. The magnitudes of the decreases are not uniform over estimators or coefficients but occur even for OLS. Increasing the range of variation for either z_2 or z_3 substantially reduces the MSE's for the coefficients in Equation (2-1) and increasing the range of variation of z_3 also has a substantial effect on the MSE of b_4. Reducing the true value of b_4 (Case 7) has the same effect as would reducing the range of variation of z_3, as one would expect. The absolute advantage of FIML over some particular alternative estimator is highly variable, not only over coefficients but from Case to Case. Thus, for example, for b_1 , the ratio of the MSE of FIML to that of OLS ranges from .25 to .75, whereas for b_2 the same ratios range from .30 to .98.

Table 4 contains the ratios for each Case, coefficient and estimator of the MSE to the mean over the replications of the estimated asymptotic variance. For larger samples and consistent estimators these ratios may be expected to be near unity. Although no rigorous test is applied to judge closeness to one, many of the ratios are in the heuristically acceptable range of .8 to 1.2. Notable exceptions are MNLIML, NLIML and OLS and Cases 1, 7 and 8. Coefficient b_4 is particularly badly estimated in this respect. Moreover, increasing the sample size, as occurs between Cases 1 and 2 and between Cases 3 and 4 improves the ratios in less than one half the cases; an unusual result for estimators other than OLS. The fraction of cases in which improvement occurs is just over one half if MNLIML, NLIML and OLS are omitted. Table 5 presents the number of coefficients for each Case and estimator for which we reject the null hypothesis that (estimated coefficients-true value)/ $MSE^{1/2}$ is distributed as $N(0,1)$. The Table is based on the Kolmogorov-Smirnov test and is not altered markedly if the divisor is not $MSE^{1/2}$ but (asymptotic variance)$^{1/2}$. In general, normality is rejected for MNLIML, NLIML, and OLS but not for the other methods. The fit to normality does not improve as sample size is increased, confirming the anomaly noted in Table 4. This anomaly often characterizes situations in which convergence to the true asymptotic distribution has adequately taken place for practical purposes. In such situations increasing the sample size with relatively few replications such as 50 may exhibit only random oscillations about the true distribution.

4. Conclusion

The various estimators examined are not massively different in performance. On the whole

(1) FIML is the best estimator;

(2) OLS is the worst estimator;

(3) The evidence concerning the further ranking of the various estimators introduced by Amemiya is somewhat ambiguous;

(4) The results appear to be highly dependent on the particular model and on the particular true value assigned to the coefficients. This latter point is well known but deserves to be stressed, for in some cases estimators other than FIML may do almost as well as FIML but in other cases can do substantially worse. From the point of view of obtaining estimates that are best behaved in small samples the risk-averse strategy is obviously to choose FIML. It is clear, however, that in models of larger size than the one investigated here the relative computational cost of FIML will rise, since the upper bound on the number of parameters in the optimization problems involved in the other methods is the upper bound of the number of coefficients in a single equation. Whether an improvement in the MSE's of, say, a factor of 2 is worth it must be left to the investigator.

BIBLIOGRAPHY

[1] Amemiya, T., "The Nonlinear Limited Information Maximum Likelihood Estimator and the Modified Nonlinear Two-Stage Least Squares Estimator," Technical Report No. 147, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1974.

[2] ____________, "The Nonlinear Two-Stage Least Squares Estimator," Journal of Econometrics, 2(1974), 105-110.

[3] Goldfeld, S.M. and R.E. Quandt, Nonlinear Methods in Econometrics, North Holland, 1972.

[4] Kelejian, H., "Two Stage Least Squares and Econometric Systems Linear in Parameter but Nonlinear in Endogenous Variables," Journal of the American Statistical Association, 64(1971), 373-374.

Table 2

Mean Square Errors

		Case 1	2	3	4	5	6	7	8
b_1	FIML	.8832	.3948	.2098	.1670	.6159	.4440	3.1053	.5361
	SNTSLS	1.3639	.6317	.3161	.2125	1.4040	.8661	4.9628	.8705
	POLY-2	1.3502	.6553	.2982	.1858	1.2573	.9352	4.5994	.8325
	POLY-3	1.2022	.6631	.2971	.1875	1.0594	.9084	3.8235	.7152
	POLY-4	1.1961	.6996	.2953	.1916	1.0033	.8137	3.8648	.7239
	MNLIML	1.0934	.6979	.3671	.2638	.9247	.6676	2.8141	.7109
	NLIML	1.1222	.7278	.4062	.2887	.9274	.7564	2.7685	.7441
	OLS	1.6312	1.5699	.3387	.2230	.9259	1.2223	5.8504	.9268
b_2	FIML	.0264	.0101	.0101	.0074	.0116	.0128	.0541	.0216
	SNTSLS	.0447	.0131	.0120	.0092	.0368	.0188	.0882	.0354
	POLY-2	.0441	.0137	.0109	.0095	.0333	.0222	.0853	.0349
	POLY-3	.0409	.0141	.0107	.0095	.0298	.0201	.0764	.0324
	POLY-4	.0411	.0158	.0105	.0095	.0285	.0176	.0783	.0327
	MNLIML	.0407	.0171	.0107	.0108	.0293	.0171	.0636	.0343
	NLIML	.0411	.0177	.0113	.0112	.0294	.0177	.0630	.0350
	OLS	.0528	.0340	.0103	.0102	.0294	.0264	.1130	.0402
b_3	FIML	6.8208	2.7085	1.8543	1.3930	4.1895	4.8585	23.6434	4.09744
	SNTSLS	10.7255	4.4976	2.9595	1.6259	10.2438	8.9519	38.2350	6.6410
	POLY-2	10.6275	4.5906	2.8160	1.4626	9.1840	9.9974	35.6417	6.3695
	POLY-3	9.6187	4.7252	2.7880	1.4810	7.9044	9.7204	30.2668	5.6369
	POLY-4	9.5611	5.0304	2.7577	1.5083	7.4461	8.5344	30.6158	5.6954
	MNLIML	9.1149	5.0824	3.3063	2.0679	7.1110	7.4699	22.9328	5.9058
	NLIML	9.3494	5.3162	3.6307	2.2722	7.1340	8.4501	22.6216	6.1695
	OLS	13.1360	11.6729	3.0587	1.7496	7.1182	13.2451	46.6891	7.5283
b_4	FIML	.0584	.0262	.0435	.0220	.0309	.1099	.0514	.0277
	SNTSLS	.0690	.0319	.0491	.0333	.0540	.1486	.0654	.0325
	POLY-2	.0682	.0323	.0495	.0338	.0536	.1401	.0657	.0321
	POLY-3	.0677	.0329	.0479	.0334	.0504	.1356	.0651	.0315
	POLY-4	.0670	.0322	.0447	.0316	.0513	.1271	.0642	.0313
	MNLIML	.0723	.0430	.0489	.0332	.0598	.1084	.0596	.0357
	NLIML	.0867	.0638	.0505	.0360	.1193	.1056	.0583	.0437
	OLS	.0635	.0284	.0476	.0276	.0484	.1125	.0581	.0301
b_5	FIML	.769E-3	.449E-3	.561E-3	.300E-3	.776E-3	.976E-3	.717E-3	.739E-3
	SNTSLS	.749E-3	.448E-3	.565E-3	.324E-3	.832E-3	1.056E-3	.696E-3	.730E-3
	POLY-2	.741E-3	.447E-3	.573E-3	.324E-3	.829E-3	1.056E-3	.698E-3	.717E-3
	POLY-3	.743E-3	.447E-3	.582E-3	.324E-3	.826E-3	1.068E-3	.700E-3	.720E-3
	POLY-4	.724E-3	.444E-3	.572E-3	.324E-3	.810E-3	1.053E-3	.688E-3	.703E-3
	MNLIML	.713E-3	.443E-3	.558E-3	.323E-3	.825E-3	.971E-3	.686E-3	.682E-3
	NLIML	.711E-3	.443E-3	.556E-3	.322E-3	.889E-3	.956E-3	.680E-3	.688E-3
	OLS	.741E-3	.445E-3	.569E-3	.323E-3	.798E-3	1.001E-3	.696E-3	.709E-3

Table 3

Mean Ranks

	Case 1	2	3	4	5	6	7	8
FIML	3.70	3.64	3.92	3.80	3.60	3.60	3.98	3.66
SNTSLS	4.54	3.96	4.55	4.28	4.89	4.69	4.77	4.42
POLY-2	4.65	4.24	4.56	4.39	4.88	4.95	4.90	4.68
POLY-3	4.39	4.49	4.32	4.46	4.26	4.90	4.62	4.29
POLY-4	4.36	4.44	4.16	4.33	4.27	4.62	4.31	4.38
MNLIML	4.43	4.51	4.79	4.92	4.69	3.96	4.08	4.56
NLIML	4.75	4.96	5.13	5.38	5.11	4.15	4.01	4.88
OLS	5.18	5.77	4.56	4.43	4.31	5.14	5.34	5.14

Table 4

Ratios of Mean Square Errors to Mean Asymptotic Variances

		Case 1	2	3	4	5	6	7	8
b_1	FIML	1.10	.78	.80	1.13	1.17	.61	1.26	1.10
	SNTSLS	1.16	.80	.92	1.04	1.29	.55	1.19	1.27
	POLY-2	1.30	.93	.92	.96	1.29	.82	1.35	1.34
	POLY-3	1.24	.98	.95	.98	1.16	.89	1.25	1.24
	POLY-4	1.29	1.06	.95	1.01	1.11	.84	1.38	1.29
	MNLIML	1.43	1.43	1.29	1.64	1.16	.82	1.23	1.52
	NLIML	1.49	1.51	1.43	1.81	1.16	.94	1.22	1.60
	OLS	2.13	3.34	1.17	1.37	1.14	1.68	2.94	1.95
b_2	FIML	1.27	.84	.88	1.16	1.00	.72	1.35	1.30
	SNTSLS	1.52	.75	.84	1.17	1.35	.58	1.33	1.54
	POLY-2	1.63	.85	.79	1.23	1.32	.86	1.53	1.62
	POLY-3	1.58	.94	.78	1.24	1.23	.83	1.51	1.57
	POLY-4	1.64	1.03	.77	1.24	1.19	.75	1.67	1.62
	MNLIML	1.85	1.36	.81	1.49	1.32	.82	1.60	1.89
	NLIML	1.88	1.42	.85	1.55	1.32	.85	1.60	1.93
	OLS	2.41	2.78	.78	1.40	1.32	1.35	3.18	2.19
b_3	FIML	1.13	.70	.78	1.10	1.07	.63	1.26	1.13
	SNTSLS	1.23	.75	.96	.93	1.26	.54	1.21	1.30
	POLY-2	1.37	.86	.96	.88	1.26	.83	1.38	1.38
	POLY-3	1.33	.91	.98	.90	1.15	.90	1.30	1.31
	POLY-4	1.37	1.00	.97	.92	1.10	.83	1.44	1.36
	MNLIML	1.58	1.36	1.27	1.48	1.18	.86	1.31	1.67
	NLIML	1.64	1.43	1.40	1.63	1.18	.99	1.31	1.75
	OLS	2.27	3.23	1.16	1.23	1.17	1.71	3.07	2.09
b_4	FIML	1.45	1.34	1.27	1.30	1.29	1.30	1.46	1.46
	SNTSLS	1.46	1.38	1.15	1.53	1.31	1.20	1.47	1.49
	POLY-2	1.46	1.42	1.17	1.57	1.31	1.18	1.49	1.49
	POLY-3	1.47	1.45	1.16	1.56	1.25	1.24	1.51	1.49
	POLY-4	1.48	1.43	1.10	1.49	1.29	1.21	1.51	1.50
	MNLIML	1.90	2.06	1.41	1.74	1.86	1.41	1.58	2.09
	NLIML	2.34	3.09	1.49	1.91	3.88	1.41	1.62	2.60
	OLS	1.45	1.36	1.23	1.43	1.26	1.24	1.41	1.49
b_5	FIML	.88	1.11	.93	1.00	1.03	1.10	.87	.83
	SNTSLS	.84	1.10	.93	1.08	1.07	1.13	.84	.80
	POLY-2	.84	1.10	.94	1.08	1.07	1.14	.85	.79
	POLY-3	.84	1.10	.96	1.08	1.07	1.17	.85	.80
	POLY-4	.82	1.10	.95	1.08	1.05	1.16	.84	.78
	MNLIML	.93	1.09	1.03	1.08	1.25	1.25	.89	.91
	NLIML	.94	1.08	1.03	1.07	1.37	1.24	.90	.92
	OLS	.84	1.10	.95	1.08	1.05	1.13	.85	.79

Table 5

Tests of Normality
Number of Coefficients for Which
Normality is Rejected at .05 Level

	Case 1	2	3	4	5	6	7	8
FIML	0	1	0	0	0	0	1	0
SNTSLS	0	2	0	0	0	1	0	0
POLY-2	0	2	0	0	0	0	0	0
POLY-3	0	3	1	0	0	0	2	1
POLY-4	2	4	1	0	0	2	3	1
MNLIML	4	4	3	4	1	1	4	4
NLIML	4	4	3	4	2	1	4	4
OLS	4	3	3	3	1	4	4	4