3-1-1979

Full Information Estimates of a Nonlinear Macroeconometric Model

Ray C. Fair

William R. Parke

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 519

FULL INFORMATION ESTIMATES OF A NONLINEAR MACROECONOMETRIC MODEL

Ray C. Fair and William R. Parke

March 6, 1979

FULL INFORMATION ESTIMATES OF A NONLINEAR MACROECONOMETRIC MODEL*

by

Ray C. Fair and William R. Parke

1. Introduction

The purpose of this paper is to report on results of estimating the model in Fair [10] by full information maximum likelihood (FIML), nonlinear three stage least squares (3SLS), and nonlinear two stage least squares (2SLS). Ordinary least squares (OLS) estimates are also presented for comparison. Although it has in the past been difficult to compute FIML and 3SLS estimates of large-scale nonlinear models,[1] an algorithm has recently been developed by one of the authors (Parke [20]) that now makes this feasible. The computation of these estimates is discussed in the first part of the paper.

There are a number of ways in which one can examine the differences

---

*The research described in this paper was financed by grant SOC77-03274 from the National Science Foundation. The authors are indebted to Jerry Hausman for helpful comments.

[1]We know of no previously successful attempts to estimate a nonlinear model of the size considered in this study by FIML or 3SLS. An attempt was made in Fair [10, Chapter 3], using traditional algorithms, to estimate the present model by FIML, but the "FIML" estimates presented in [10] are not the true FIML estimates. Since these numbers were published, a much larger value of the likelihood function for this problem (and different coefficient estimates) has been obtained using the algorithm considered in this paper. In Fair [9] a 19-equation model (11 stochastic equations) with 61 unknown coefficients was estimated by FIML using traditional algorithms. This model is not, however, in the category of a large-scale model, and only the recursive part of it is nonlinear. Further results of estimating this model by FIML and 3SLS, again using traditional algorithms, are presented in Belsley [3]. This is the largest model considered by Belsley in both [3] and [4]. The algorithm presented in Dagenais [7] appears capable of estimating only medium-size nonlinear models (about 50 coefficients) by FIML.

among the four sets of estimates once they have been obtained, and the second part of the paper is concerned with this topic. On a strictly statistical level, one can compare the 3SLS and FIML estimates via the Hausman test [17] to test the hypothesis that the error terms are normally distributed. In the same vein, one can compare the 2SLS and FIML estimates via the Hausman test to test the hypothesis that the model is correctly specified. There are, however, as will be discussed, some practical problems that arise when trying to use the Hausman test in the present context, and the current application of the test has only been partly successful. On a more informal level, one can examine the sensitivity of the dymamic prediction accuracy of the model and the sensitivity of policy effects in the model to the alternative estimates. The results of these comparisons are also presented below.

The model and estimation techniques are described in Section II, and the computation of the estimates is discussed in Section III. The coefficient estimates are then presented and discussed in Section IV. The results of the Hausman tests are also discussed in Section IV, and the prediction and policy results are presented in Sections V and VI respectively. Section VII contains a summary of the main conclusions of this study. The algorithm is briefly described in the Appendix.

II. The Model and Estimation Techniques

The Model

The model in Fair [10] has been updated since [10] was published, and the version that has been used in this study is presented in Fair [12]. This version consists of 97 equations, 29 of which are stochastic, and has 182 unknown coefficients to estimate, including 12 first order serial correlation coefficients. The estimation period is 1954I-1978II (98 observations). The model is nonlinear in variables and, as is discussed next, nonlinear in coefficients because of correction for serial correlation of some of the error terms. There is also one nonlinear restriction on the coefficients of two of the equations, which means that there are only 181 freely estimated coefficients.

The Treatment of Serial Correlation

By treating the serial correlation coefficient as a structural coefficient, it is possible to transform an equation with a serially correlated error into an equation without one. This introduces nonlinear restrictions on the coefficients, but otherwise the equation is like any other equation with a non-serially correlated error.[2] This transformation has been made in this study for the relevant equations of the model, and so the model should be thought of as one with nonlinear coefficient restrictions and no serially correlated errors. All references to the covariance matrices of the coefficient estimates in the following discussion are for the coefficient estimates inclusive of the estimates of the serial correlation coefficients.

The Notation

The notation in this paper follows closely the notation in Amemiya [2]. Write the model as

$$(1) \qquad f_i(y_t, x_t, \alpha_i) = u_{it} \ , \quad (i = 1, \ldots, n) \ , \quad (t = 1, \ldots, T) \ ,$$

where y_t is an n-dimensional vector of endogenous variables, x_t is a

[2]See, for example, the discussion in Fair [10, Chapter 3]. This procedure results in the "loss" of the first observation, but this loss has no effect on the asymptotic properties of the estimators.

vector of predetermined variables, and α_i is a vector of unknown coefficients. Assume that the first m equations are stochastic, with the remaining u_{it} $(i = m+1, \ldots, n)$ identically zero for all t. Assume also that $(u_{1t}, \ldots, u_{mt})$ is independently and identically distributed as multivariate $N(0,\Sigma)$. The other assumptions regarding (1) are as in Amemiya [2].

Let J_t be the $n \times n$ Jacobian matrix whose ij element is $\partial f_i/\partial y_{jt}$, $(i, j = 1, \ldots, n)$, and let S be the $m \times m$ matrix whose ij element is s_{ij}, where $s_{ij} = \frac{1}{T}\sum_{t=1}^{T} u_{it}u_{jt}$, $(i, j = 1, \ldots, m)$. Also, let u_i be the T-dimensional vector $(u_{i1}, \ldots, u_{iT})'$, and let u be the m·T-dimensional vector $(u_{11}, \ldots, u_{1T}, \ldots, u_{m1}, \ldots, u_{mT})'$. Assume for now that there are no constraints among the α_i's, and let α denote the k-dimensional vector $(\alpha_1', \ldots, \alpha_m')'$ of all the unknown coefficients. Finally, let G_i' be the $k_i \times T$ matrix whose t^{th} column is $\partial f_i(y_t, x_t, \alpha_i)/\partial\alpha_i$, where k_i is the dimension of α_i, and let G' be the $k \times m{\cdot}T$ matrix

$$\begin{bmatrix} G_1' & 0 & \cdots & 0 \\ 0 & G_2' & & \\ \vdots & & \ddots & \\ 0 & & & G_m' \end{bmatrix},$$

where $k = \sum_{i=1}^{m} k_i$.

Two Stage Least Squares (2SLS)

2SLS estimates of α_i (say $\hat{\alpha}_i$) are obtained by minimizing

$$(2) \qquad u_i'Z_i(Z_i'Z_i)^{-1}Z_i'u_i = u_i'D_iu_i$$

with respect to α_i , where Z_i is a $T \times K_i$ matrix of predetermined variables. Z_i and K_i can differ from equation to equation. An estimate of the covariance matrix of $\hat{\alpha}_i$ (say $\hat{V}_{2ii}$) is

$$(3) \qquad \hat{V}_{2ii} = \hat{\sigma}_{ii}(\hat{G}_i' D_i \hat{G}_i)^{-1} ,$$

where $\hat{G}_i$ is G_i evaluated at $\hat{\alpha}_i$ and $\hat{\sigma}_{ii} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_{it}^2$, $\hat{u}_{it} = f_i(y_t, x_t, \hat{\alpha}_i)$. The 2SLS estimator in this form is presented in Amemiya [1].

If an equation is nonlinear in variables only, standard linear 2SLS packages can be used to obtain $\hat{\alpha}_i$ by merely redefining the variables. If, on the other hand, an equation is nonlinear in coefficients, then in general a nonlinear optimization algorithm must be used to minimize (2). A special case of coefficient nonlinearity occurs when the nonlinearity arises only because of the presence of the first order serial correlation coefficient. In this case (2) can be minimized by an iterative technique like the Cochrane-Orcutt technique [6]. This technique is discussed in Fair [8], and it is the technique that has been used in this study to minimize (2) for those equations that are estimated under the assumption of first order serial correlation of the error term.[3]

---

[3] In Fair [8, p. 514] it was suggested that the covariance matrix of the coefficient estimates inclusive of the estimate of the serial correlation coefficient be estimated by ignoring the correlation between the latter estimate and the other coefficient estimates. Fisher, Cootner, and Baily [14, p. 575, fn. 6], however, have pointed out that one need not ignore this correlation. In terms of the notation in [8], their suggested estimate of the covariance matrix is

$$(i) \qquad \hat{\sigma}_{11}\begin{bmatrix} \hat{Q}_1\hat{Q}_1' & \hat{Q}_1\hat{u}_{1_{-1}}' \\ \hat{u}_{1_{-1}}\hat{Q}_1' & \hat{u}_{1_{-1}}\hat{u}_{1_{-1}}' \end{bmatrix}^{-1} .$$

(continued)

In the discussion of the Hausman tests in Section IV reference will be made to the covariance matrix of all the 2SLS coefficient estimates, i.e., to the $k \times k$ covariance matrix of $\hat{\alpha}$, where $\hat{\alpha} = (\hat{\alpha}_1', \ldots, \hat{\alpha}_m')'$. For the standard linear simultaneous equations model this covariance matrix is presented in Theil [22, pp. 499-500] for the case in which the same set of first stage regressors is used for each equation. For the case considered here, a nonlinear model and a different set of first stage regressors for each equation, it is straightforward to show that this matrix (say V_2) is[4]

$$(4) \qquad V_2 = \begin{pmatrix} V_{211} & \cdots & V_{21m} \\ \vdots & & \vdots \\ V_{2m1} & \cdots & V_{2mm} \end{pmatrix},$$

It can be easily seen that this matrix is the same as $\hat{V}_{2ii}$ in (3). In other words, if in the first order serial correlation case one minimized (2) using some general purpose algorithm and then computed $\hat{V}_{2ii}$ in (3), the same numbers would be obtained (aside from rounding error) as would be obtained if one used the iterative technique in [8] to get the estimates and then computed the matrix in (i). (This is assuming that the exogenous, lagged exogenous, and lagged endogenous variables in the equation being estimated are included in the Z_i matrix. If this is not done, then the technique in [8] leads to inconsistent estimates, whereas minimizing (2) using some general purpose algorithm still results in consistent estimates.) For the results in this study the Fisher, Cootner, and Baily suggestion was followed: the estimated covariance matrix in (i) was used.

[4] The derivation in Theil can be easily modified to incorporate the case of different sets of first stage regressors. Nonlinearity can be handled as in Amemiya [1, Appendix 1], i.e., by a Taylor expansion of each equation. The formal proof that V_2 is as in (4), (5), and (6) is straightforward but lengthy, and it is omitted here. Jorgenson and Laffont [18, p. 636] incorrectly assert that the off-diagonal blocks of V_2 are zero.

where

$$(5) \qquad V_{2ii} = \sigma_{ii}\left[\text{plim}\ \frac{1}{T}G_i'D_iG_i\right]^{-1} ,$$

$$(6) \qquad V_{2ij} = \sigma_{ij}\left[\text{plim}\ \frac{1}{T}G_i'D_iG_i\right]^{-1}\left[\text{plim}\ \frac{1}{T}G_i'D_iD_j'G_j\right]\left[\text{plim}\ \frac{1}{T}G_j'D_jG_j\right]^{-1} .$$

An estimate of V_{2ii} is $\hat{V}_{2ii}$ in (3). An estimate of V_{2ij} (say $\hat{V}_{2ij}$) is:

$$(7) \qquad \hat{V}_{2ij} = \hat{\sigma}_{ij}(\hat{G}_i'D_i\hat{G}_i)^{-1}(\hat{G}_i'D_iD_j'\hat{G}_j)(\hat{G}_j'D_j\hat{G}_j)^{-1} ,$$

where $\hat{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_{it}\hat{u}_{jt}$.

<u>Three Stage Least Squares (3SLS)</u>

3SLS estimates of α (say $\hat{\alpha}$) are obtained by minimizing

$$(8) \qquad u'[\hat{\Sigma}^{-1} \otimes Z(Z'Z)^{-1}Z']u = u'Du$$

with respect to α , where $\hat{\Sigma}$ is a consistent estimate of Σ and Z is a $T \times K$ matrix of predetermined variables. An estimate of the covariance matrix of $\hat{\alpha}$ (say $\hat{V}_3$) is

$$(9) \qquad \hat{V}_3 = (\hat{G}'D\hat{G})^{-1} ,$$

where $\hat{G}$ is G evaluated at $\hat{\alpha}$. $\hat{\Sigma}$ is usually estimated from the 2SLS estimated residuals. This estimator is presented in Jorgenson and Laffont [18]. See also Amemiya [2].

The 3SLS estimator as discussed in [18] and [2] and as presented in (8) uses the same Z matrix for each equation. In small samples this can be a disadvantage of 3SLS relative to 2SLS. It is possible to modify (8) to include the case of different Z_i matrices for each equation, and although this modification was not used in this study, it is of interest to consider. This estimator is the one that minimizes

$$\text{(10)} \qquad u'\left[\begin{pmatrix} Z_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & Z_m \end{pmatrix}\begin{pmatrix} \hat{\sigma}_{11}Z_1'Z_1 & \cdots & \hat{\sigma}_{1m}Z_1'Z_m \\ \vdots & & \vdots \\ \hat{\sigma}_{m1}Z_m'Z_1 & \cdots & \hat{\sigma}_{mm}Z_m'Z_m \end{pmatrix}^{-1}\begin{pmatrix} Z_1' & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & Z_m' \end{pmatrix}\right]u = u'\overline{D}u$$

with respect to $\hat{\alpha}$. An estimate of the covariance matrix of this estimator is $(\hat{G}'\overline{D}\hat{G})^{-1}$. (10) reduces to (8) when $Z_1 = \ldots = Z_m = Z$. The computational problem with this estimator is that it requires inverting the middle matrix in brackets. This matrix is of dimension $K^* = \sum_{i=1}^{m} K_i$, which is generally a large number. In the present application K^* is 350, and it did not appear feasible to invert a matrix this large. In some applications, however, it may be feasible to invert this matrix. This estimator has the advantage that it is the natural full information extension of 2SLS when different sets of first stage regressors are used. The consistency of this estimator can be proved along the lines of the proof (Jorgenson and Laffont [18, pp. 626-628]) that the estimator that minimizes (8) is consistent.

<u>Full Information Maximum Likelihood (FIML)</u>

FIML estimates of α are obtained by maximizing

$$\text{(11)} \qquad L = -\frac{T}{2}\log|S| + \sum_{t=1}^{T}\log|J_t|$$

with respect to α . An estimate of the covariance matrix of these estimates (say $\hat{V}_4$) is

$$\hat{V}_4 = -\left(\frac{\partial^2 L}{\partial\alpha\partial\alpha'}\right)^{-1} , \tag{12}$$

where the derivatives are evaluated at the optimum. FIML is, of course, a well known estimator. See, for example, Chow [5] for a recent discussion in the nonlinear case.

Ordinary Least Squares (OLS)

OLS estimates of α_i are obtained by minimizing (2) for $D_i = I$. For purposes of this study, the estimated covariance matrix of the OLS estimates was taken to be (3) for $D_i = I$. The discussion in the second paragraph under the 2SLS heading is relevant here also. In particular, note that the Cochrane-Orcutt iterative technique can be used to minimize (2) if the nonlinearity in coefficients is due solely to the presence of the serial correlation coefficient.

III. The Computation of the Estimates

As noted in the previous section, the iterative procedure in [8] was used for the 2SLS estimates of the equations that were estimated under the assumption of first order serial correlation of the error terms. Otherwise, a standard 2SLS package was used.[5] The 2SLS technique has been

---

[5]The TSCORC and INST options in the TSP regression program were used for these estimates. It should be noted, however, that the TSCORC option was modified to use the formula in (i) in footnote 2 to compute the covariance matrix. The standard TSCORC option assumes that $\hat{Q}_i\hat{u}'_{i_{-1}}$ is zero when computing the covariance matrix. Also, both the TSCORC and INST options

(continued)

the primary method used to estimate the successively updated versions of the model, and the estimates of the 182 coefficients of the current version presented in [12] are 2SLS estimates. This set of estimates is the starting point for the present study. For this set a different Z_i matrix was used for each of the 26 equations estimated by 2SLS. (Three of the 29 stochastic equations have no right-hand-side endogenous variables and so were estimated by OLS.) The variables used for each Z_i matrix are presented in Table 2-5 in [12]. The number of variables in a given matrix varies from 11 to 31.

Although there are 182 unknown coefficients in the model, only 107 were estimated by 3SLS and FIML in this study. The other 75 coefficients were set equal to their 2SLS estimates. This was done partly to ease the computational burden and partly because, as discussed below, it is not clear that 98 observations are enough to estimate all 182 coefficients by FIML. The 75 non-estimated coefficients include all the coefficients in 13 equations and the coefficients of strike dummy variables in 5 of the remaining 16 equations. It should be noted, however, that even though the structural coefficients of 13 equations were not estimated by 3SLS and FIML, these equations were not dropped from the model. For example, $\sum$ and S were still taken to be 29×29 matrices, and J_t was still taken to be 97×97. This procedure allows the correlations between the error terms in the 13 non-estimated equations and the error terms in the 16 estimated equations to have an effect on the coefficient estimates of the 16

---

divide the sum of squared residuals by $T - k_i$ when computing the estimated variance, where k_i is the number of coefficients estimated, whereas for present purposes all sums of squares have been divided by T .

estimated equations.

The 3SLS estimates were obtained using the algorithm that is described in the Appendix. 58 variables were chosen for the Z matrix. This set of variables was chosen to correspond roughly to the union of the sets of variables in the 16 relevant Z_i matrices, although not every variable in this union was chosen. (Had every variable in the union been chosen, the number of variables in the Z matrix would have been close to the total number of observations.) A list of these variables is available from the authors upon request. The 2SLS residuals were used to compute $\hat{\Sigma}$.

The matrix D in (8) is 2842×2842 ($m \cdot T = 29 \cdot 98 = 2842$) , and so computing $u'Du$ once for a given value of α requires a large number of calculations. Fortunately, however, only a small fraction of these calculations need to be performed most of the time that the algorithm requires a new value of the objective function corresponding to a new value of α . In particular, most of the time the algorithm is changing the coefficients of only one equation between evaluations of the objective function, and recomputing $u'Du$ when only one equation has been affected requires many fewer calculations than are needed when all equations have been affected.

The results of computing the 107 3SLS estimates are presented in the first half of Table 1. The cost of computing $u'Du$ once varies from 0.40 seconds when only one equation has been affected to 2.94 seconds when all equations have been affected. The approximate number of function evaluations per iteration of the algorithm is 432. The algorithm was allowed to run for 28 iterations, starting from the 2SLS estimates, and took about 106 minutes on the IBM 370-158 at Yale. It is clear from the results

TABLE 1

The Cost of the 3SLS and FIML Estimates

107 coefficients estimated
1 nonlinear restriction across 2 equations
16 stochastic equations estimated
13 stochastic equations not estimated
58 identities (97 total equations)
98 quarterly observations (1954I-1978II)

3SLS

F = u'Du in (8)
F at start (2SLS estimates) = 1898.63
F after 28 iterations = 1850.30
Total ΔF = -48.33

Iter. No.	|ΔF|	# > 1%	Iter. No.	|ΔF|	# > 1%	Iter. No.	|ΔF|	# > 1%
1	21.90	67	11	0.72	32	21	0.20	16
2	9.52	63	12	0.55	29	22	0.19	17
3	3.79	56	13	0.56	28	23	0.21	10
4	1.89	51	14	0.37	23	24	0.20	18
5	1.85	47	15	0.39	20	25	0.22	10
6	0.86	38	16	0.27	20	26	0.25	10
7	0.63	31	17	0.36	21	27	0.09	6
8	1.11	34	18	0.23	14	28	0.14	10
9	0.72	25	19	0.25	27			
10	0.63	30	20	0.23	15			

Approximate number of function evaluations per iteration = 432.
Approximate cost per function evaluation = 0.40 - 2.94 seconds.
Approximate total cost of 28 iterations = 106 minutes.

TABLE 1 (continued)

FIML

L = L in (11).
L at start (2SLS estimates) = 2465.95 (two Jacobians), 2569.73 (six Jacobians)
L after 28 iterations = 2508.16 (two Jacobians), 2613.09 (six Jacobians)
L after 43 iterations = 2508.67 (two Jacobians), 2614.14 (six Jacobians)
Total ΔL = 42.72 (two Jacobians), 44.41 (six Jacobians)

Two Jacobians			Two Jacobians			Six Jacobians		
Iter. No.	ΔL	# > 1%	Iter. No.	ΔL	# > 1%	Iter. No.	ΔL	# > 1%
1	19.66	74	15	0.17	21	29	0.22	23
2	8.16	66	16	0.07	26	30	0.08	22
3	3.31	58	17	0.16	30	31	0.04	19
4	1.91	58	18	0.19	33	32	0.04	17
5	1.11	57	19	0.21	32	33	0.04	21
6	1.56	65	20	0.13	36	34	0.06	31
7	0.91	58	21	0.13	32	35	0.08	36
8	0.90	68	22	0.14	35	36	0.07	24
9	0.65	60	23	0.10	30	37	0.07	26
10	0.50	50	24	0.06	22	38	0.07	36
11	0.79	58	25	0.04	22	39	0.06	31
12	0.61	63	26	0.05	21	40	0.06	23
13	0.36	40	27	0.03	17	41	0.05	23
14	0.25	35	28	0.05	19	42	0.03	16
						43	0.04	14

Approximate number of function evaluations per iteration = 432.
Approximate cost per function evaluation = 0.20 - 0.64 seconds (two Jacobians),
0.37 - 0.85 seconds (six Jacobians).
Approximate total cost of 43 iterations = 121 minutes.

Notes: # > 1% = number of coefficients that changed by more than 1.0 percent from the previous iteration.
Approximate cost of one minute = $12.48 without discounts.
80% discount given for large overnight jobs.

in Table 1 that the algorithm had not found the exact optimum after 28 iterations, although the objective function and coefficient estimates were not changing very much by this time. The estimates at this point were taken to be the 3SLS estimates.

With respect to the 3SLS covariance matrix, the algorithm does not compute $\hat{G}$, and so extra work is involved at the end to obtain the covariance matrix. For present purposes $\hat{G}$ was computed numerically and then $(\hat{G}'D\hat{G})^{-1}$ was obtained. The total time involved in these calculations was about 3.2 minutes.

The FIML estimates were obtained using the same algorithm. Computing L in (11) once for a given value of α also requires a large number of calculations, but there are again cost savings that can be made. These savings are as follows. First, when the coefficients of only one equation are changed by the algorithm, which is most of the time, only one row and one column of S are affected. The average cost of computing S is thus much less than it would be if all the rows and columns had to be computed anew each time a new value of L was needed. Second, the Jacobian matrix J_t is very sparse (333 nonzero elements out of 9409), and so considerable saving can be achieved by using a sparse matrix routine to take its determinant. Third, it turns out, as reported in Fair [10, Chapter 3], that a fairly good approximation to $\sum_{t=1}^{T} \log|J_t|$ is $\frac{T}{2}(\log|J_1| + \log|J_T|)$. This approximation obviously saves an enormous amount of time, since only 2 determinants have to be computed instead of 98. Unlike the first two savings, however, this third saving does require that an approximation be made. The exact value of L is not being computed by the algorithm, and the hope is that the error involved in this approximation is nearly constant for different sets of coefficient values. As will be discussed, this seems to be the case from the present results.

The results of computing the FIML estimates are presented in the second half of Table 1. For the first 28 iterations two Jacobians (J_1 and J_{98}) were computed per evaluation of L , and for the remaining 15 iterations six Jacobians $(J_1, J_{20}, J_{39}, J_{59}, J_{78}, J_{98})$ were computed. When six Jacobians were computed, $\sum_{t=1}^{T} \log|J_t|$ was approximated by first computing the six values of $\log|J_t|$, $t = 1, 20, 39, 59, 78, 98$, and then interpolating in the appropriate way between pairs. When two Jacobians are used, the cost of computing L once varies from 0.20 seconds when only one equation has been affected to 0.64 seconds when all equations have been affected. When six Jacobians are used, the corresponding numbers are 0.37 seconds and 0.85 seconds. The total time for the 43 iterations was about 121 minutes. Again, the algorithm had not quite found the optimum at the time it was stopped, but it seemed fairly close. The estimates after the 43^{rd} iteration were taken to be the FIML estimates.

With respect to the Jacobians, the results of switching from two to six Jacobians after iteration 28 suggest that little is lost by using only two. The change in L when the six-Jacobian approximation replaced the two-Jacobian approximation, while about 100 points, merely reflects a different bias of the six-Jacobian approximation. The stability of the bias is more important than its absolute value because adding a constant bias has no effect on the likelihood maximization. The close agreement of the two- and six-Jacobian results can be seen in the small change (only 0.22 points) on iteration 29, the first using six Jacobians. If the change to six Jacobians were important, the likelihood change would have been larger and the coefficients would have changed much more.

The second derivatives that are needed for the FIML covariance matrix in (12) were computed numerically. The total time involved in this was about 52.6 minutes. Computing the second derivatives turned out not to be a straightforward task, and the exact way that this was finally done is explained in the Appendix.

Before concluding the discussion of the FIML estimates, the identification issue should be mentioned. For a linear model Sargan [21] has proved Klein's [19] conjecture that the FIML estimator is unidentified if the number of observations is less than the number of endogenous plus predetermined variables. For a nonlinear model the exact conditions for identification are not known, but in the present case it seemed unlikely that 98 observations were enough to estimate the complete model by FIML. In the 29 stochastic equations there are 140 different variables (endogenous plus predetermined), counting different nonlinear functional forms of the same variable as different variables.[6] In the restricted version of the model, on the other hand, there are only 83 different variables; and it seemed likely in this case that 98 observations were enough for identification. This is the main reason for cutting the problem down from 182 to 107 unknown coefficients.

In order to make the OLS results comparable to the 3SLS and FIML results, only the 107 coefficients were estimated by OLS. This means that

---

[6]For the serial correlation coefficients only one extra variable was counted per coefficient. In other words, the additional lagged variables that serial correlation introduces (after the transformation mentioned in Section II) were only counted as one extra variable.

the values of the other 75 coefficients are the same across all four sets of estimates.

Two further points about the estimates should be noted. The first concerns the treatment of the nonlinear restriction across the two equations. For the 3SLS and FIML estimates the restriction is easy to handle. One can merely solve for one of the coefficients in terms of the others and then substitute this solution for the coefficient wherever it appears. The dimension of the optimization problem is then reduced from 107 to 106. Likewise, $\hat{V}_3$ in (9) and $\hat{V}_4$ in (12) are reduced from 107×107 to 106×106. For the 2SLS and OLS estimates the restriction is less straightforward to handle because it is across two equations. It was handled in these two cases by first estimating one of the equations (the price equation) unrestricted and then using these coefficient estimates and the restriction to eliminate one of the coefficients from the other equation (the wage equation). This way of accounting for the restriction, which is discussed in more detail in Fair [12, pp. 11-13], affects only the coefficient estimates of the wage equation. The dimension of the optimization problem for the wage equation was reduced from 6 to 5, and the estimated covariance matrix was reduced from 6×6 to 5×5.

The second point to note is that even though all 181 unrestricted coefficients were estimated by 2SLS, the relevant covariance matrix to compare to the 3SLS and FIML matrices is 106×106, not 181×181. Therefore, V_2 in (4), and likewise V_1 for the OLS estimates, should be considered for present purposes to be 106×106. In particular, note that although the coefficients of strike dummy variables in 5 of the 16 equations were estimated by 2SLS, these coefficient estimates were taken as fixed for purposes of computing the covariance matrices of the remaining coefficient estimates in the equations.

IV. The Coefficient Estimates and the Hausman Tests

The estimates of the 106 unrestricted coefficients and their estimated standard errors are presented in Table 2. The coefficient estimates in Table 2 are not in themselves very useful for descriptive purposes because they require knowledge of the model, and an explanation of the model is beyond the scope of this paper. Of more interest for present purposes are the last three columns in Table 2, and these will be discussed along with the discussion of the Hausman tests.

The Hausman m-statistic provides a useful way of exmaining the differences among the estimates, although, as will be seen, there are some problems with applying the Hausman tests in practice. Consider two estimators, $\hat{\beta}_0$ and $\hat{\beta}_1$, where under some null hypothesis both estimators are consistent, but only $\hat{\beta}_0$ attains the asymptotic Cramer-Rao bound, while under the alternative hypothesis only $\hat{\beta}_1$ is consistent. Let $\hat{q} = \hat{\beta}_1 - \hat{\beta}_0$, and let $\hat{V}_0$ and $\hat{V}_1$ denote consistent estimates of the asymptotic covariance matrices (V_0 and V_1) of $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively. Hausman's m-statistic is $\hat{q}'(\hat{V}_1 - \hat{V}_0)^{-1}\hat{q}$, and he has shown that it is asymptotically distributed as χ^2 with k degrees of freedom, where k is the dimension of $\hat{q}$. Note that under the null hypothesis, $V_1 - V_0$ is positive definite.

Consider now comparing the FIML and 3SLS estimates. Under the null hypothesis of correct specification and normally distributed errors, both estimates are consistent, but only the FIML estimates attain the asymptotic Cramer-Rao bound, while under the alternative hypothesis of correct specification and nonnormality, only the 3SLS estimates are consistent. (See Amemiya [2].) Let $\hat{\alpha}^{(3)}$ and $\hat{\alpha}^{(4)}$ denote the 3SLS and FIML estimates of α respectively, and let $\hat{q} = \hat{\alpha}^{(3)} - \hat{\alpha}^{(4)}$. The m-statistic in this

TABLE 2

The Four Sets of Coefficient Estimates

Eq. No. in [12]	Coef. No. in [12]	2SLS $Coef_2$	2SLS SE_2	3SLS $Coef_3$	3SLS SE_3	FIML $Coef_4$	FIML SE_4	OLS $Coef_1$	OLS SE_1	$\frac{SE_2}{SE_3}$	$\frac{SE_3}{SE_4}$	$m = \frac{(Coef_2-Coef_3)^2}{(SE_2^2-SE_3^2)}$
1	2	0.9761	0.0273	0.9709	0.0217	0.9617	0.0266	0.9689	0.0267	1.26	0.82	0.10
1	3	0.01493	0.00898	0.01529	0.00740	0.01753	0.00902	0.01341	0.00878	1.21	0.82	0.00
1	6	0.02114	0.01368	0.02300	0.01087	0.02735	0.01288	0.02289	0.01340	1.26	0.84	0.05
1	7	-0.008177	0.006024	-0.009434	0.004818	-0.008162	0.006071	-0.007532	0.005746	1.25	0.79	0.12
1	8	-0.007056	0.002167	-0.005478	0.001702	-0.005760	0.002051	-0.005046	0.002003	1.27	0.83	1.38
1	9	0.04045	0.02963	0.02569	0.02379	0.02746	0.03281	0.03050	0.02782	1.25	0.73	0.70
1	1	0.08526	0.09583	0.06206	0.07606	0.03387	0.09352	0.03405	0.09262	1.26	0.81	0.16
2	11	0.4678	0.0894	0.4851	0.0628	0.5111	0.0724	0.3419	0.0783	1.42	0.87	0.07
2	12	0.07293	0.02585	0.06464	0.02015	0.06469	0.02735	0.05137	0.02449	1.28	0.74	0.26
2	13	-0.1321	0.0534	-0.1263	0.0377	-0.1152	0.0492	-0.2165	0.0465	1.42	0.77	0.02
2	14	0.1672	0.0560	0.1719	0.0393	0.1862	0.0545	0.2513	0.0491	1.42	0.72	0.01
2	15	0.02145	0.02370	0.02704	0.01740	0.05602	0.02193	0.03444	0.02232	1.36	0.79	0.12
2	16	0.02352	0.02607	0.01038	0.01945	-0.03087	0.02768	0.01210	0.02478	1.34	0.70	0.57
2	17	0.1890	0.0911	0.2224	0.0711	0.3308	0.0930	0.2294	0.0879	1.28	0.76	0.34
2	18	0.2981	0.0692	0.2969	0.0508	0.2720	0.0588	0.3711	0.0594	1.36	0.86	0.00
2	167	-0.003556	0.003143	-0.003061	0.002322	-0.004077	0.002575	-0.000250	0.002785	1.35	0.90	0.05
2	10	-3.553	0.722	-3.513	0.502	-3.402	0.599	-4.690	0.626	1.44	0.84	0.01
3	20	0.9702	0.0296	0.9704	0.0229	0.9796	0.0256	0.9550	0.0283	1.29	0.89	0.00
3	21	-0.04037	0.01863	-0.02871	0.01385	-0.01116	0.01453	-0.03382	0.01505	1.35	0.95	0.87
3	22	0.05463	0.02134	0.04901	0.01602	0.03661	0.01758	0.05829	0.01840	1.33	0.91	0.16
3	24	-0.02697	0.00506	-0.02892	0.00387	-0.02983	0.00508	-0.02029	0.00458	1.31	0.76	0.36
3	25	0.1382	0.0274	0.1304	0.0188	0.1076	0.0261	0.1333	0.0216	1.45	0.72	0.15
3	182	0.007849	0.007031	0.009893	0.005545	0.015904	0.006437	0.009826	0.006349	1.27	0.86	0.22
3	183	0.04311	0.03853	0.03740	0.03059	0.04053	0.03850	0.05578	0.03883	1.26	0.79	0.06
3	30	0.6132	0.0860	0.6630	0.0697	0.6586	0.0863	0.6761	0.0917	1.23	0.81	0.98
3	19	-0.1408	0.2243	-0.1219	0.1682	-0.0238	0.1866	-0.2384	0.2039	1.33	0.90	0.02
4	32	0.9431	0.0265	0.9514	0.0158	0.9482	0.0140	0.9431	0.0265	1.67	1.13	0.15
4	33	-0.004101	0.006797	-0.004706	0.005109	-0.006757	0.005365	-0.004101	0.006797	1.33	0.95	0.02
4	34	0.01966	0.01083	0.01855	0.00753	0.02044	0.00752	0.01966	0.01083	1.44	1.00	0.02
4	35	-0.006389	0.002875	-0.006470	0.002133	-0.006215	0.002275	-0.006389	0.002875	1.35	0.94	0.00
4	36	-0.004142	0.002568	-0.004618	0.001897	-0.003912	0.002131	-0.004142	0.002568	1.35	0.89	0.08
4	37	0.02636	0.01071	0.01270	0.00785	0.00333	0.00919	0.02636	0.01071	1.36	0.85	3.52
4	38	0.4489	0.2527	0.3440	0.1916	0.3218	0.2046	0.4489	0.2527	1.32	0.94	0.41
4	168	0.004538	0.003833	0.002663	0.003005	0.002642	0.003721	0.004538	0.003833	1.28	0.81	0.62
4	39	0.8279	0.0641	0.7599	0.0457	0.6816	0.0516	0.8279	0.0641	1.40	0.89	2.29
4	31	-0.2765	0.1490	-0.2391	0.0911	-0.2602	0.0812	-0.2765	0.1490	1.64	1.12	0.10
5	41	0.5902	0.0878	0.5814	0.0722	0.5261	0.0817	0.5970	0.0871	1.22	0.88	0.03
5	42	0.01379	0.00835	0.01452	0.00702	0.01873	0.00829	0.01269	0.00816	1.19	0.85	0.03
5	43	-0.006225	0.005111	-0.006585	0.004309	-0.008833	0.004998	-0.005630	0.005020	1.19	0.86	0.02
5	44	0.08245	0.02069	0.08289	0.01743	0.09176	0.01885	0.08155	0.02064	1.19	0.92	0.00
5	40	-0.06397	0.05318	-0.06806	0.04481	-0.09176	0.05213	-0.05765	0.05220	1.19	0.86	0.02
6	46	0.8757	0.0433	0.9147	0.0257	0.9325	0.0332	0.8796	0.0425	1.68	0.77	1.25
6	47	-0.04097	0.01476	-0.02318	0.00841	-0.01690	0.01126	-0.03991	0.01449	1.75	0.75	2.15
6	48	0.06103	0.02009	0.04039	0.01157	0.03190	0.01542	0.05883	0.01970	1.74	0.75	1.58
6	50	0.1318	0.0340	0.1071	0.0224	0.1033	0.0276	0.1402	0.0325	1.52	0.81	0.94
6	45	-0.3288	0.1111	-0.2027	0.0627	-0.1537	0.0849	-0.3187	0.1088	1.77	0.74	1.89
8	57	0.8578	0.0584	0.8456	0.0474	0.8286	0.0511	0.7959	0.0545	1.23	0.93	0.13
8	58	-0.01535	0.00539	-0.01649	0.00426	-0.01730	0.00493	-0.01620	0.00484	1.26	0.86	0.12
8	59	0.1944	0.0888	0.1898	0.0717	0.2026	0.0817	0.2945	0.0825	1.24	0.88	0.01
8	60	-0.001233	0.000799	-0.001067	0.000646	-0.001114	0.000765	-0.002143	0.000744	1.24	0.84	0.12
8	56	0.4300	0.2860	0.3049	0.2320	0.2771	0.2867	0.7330	0.2688	1.23	0.81	0.56
9	62	0.8299	0.0199	0.8190	0.0127	0.8092	0.0148	0.8323	0.0194	1.56	0.86	0.51
9	63	0.06139	0.00476	0.06516	0.00310	0.06718	0.00378	0.06112	0.00474	1.54	0.82	1.09
9	64	0.06188	0.01307	0.06161	0.00823	0.06494	0.00839	0.06013	0.01208	1.59	0.98	0.00
9	65	0.007898	0.005385	0.012312	0.003370	0.013473	0.003451	0.008249	0.004893	1.60	0.98	1.10
9	66	-0.003349	0.001335	-0.002581	0.000855	-0.002481	0.000880	-0.003249	0.001278	1.56	0.97	0.56
9	67	-0.005097	0.001063	-0.004051	0.000727	-0.003395	0.000903	-0.005211	0.001027	1.46	0.80	1.82
9	169	0.1699	0.0993	0.1008	0.0647	0.0819	0.0711	0.1760	0.0984	1.53	0.91	0.84
9	61	-0.1505	0.0219	-0.1471	0.0142	-0.1513	0.0163	-0.1498	0.0219	1.54	0.87	0.04

TABLE 2 (continued)

Eq. No. in [12]	Coef. No. in [12]	2SLS Coef$_2$	2SLS SE$_2$	3SLS Coef$_3$	3SLS SE$_3$	FIML Coef$_4$	FIML SE$_4$	OLS Coef$_1$	OLS SE$_1$	$\frac{SE_2}{SE_3}$	$\frac{SE_3}{SE_4}$	$m = \frac{(Coef_2 - Coef_3)^2}{(SE_2^2 - SE_3^2)}$
10	69	0.2393	0.0769	0.2398	0.0402	0.2520	0.0465	0.2337	0.0490	1.91	0.86	0.00
10	70	0.8811	0.0832	0.9060	0.0436	0.8848	0.0514	0.8874	0.0505	1.91	0.85	0.12
10	71	-0.1476	0.0259	-0.1796	0.0222	-0.1680	0.0318	-0.1484	0.0248	1.17	0.70	5.71
10	75	0.3961	0.1143	0.4319	0.0811	0.4049	0.1133	0.4001	0.1012	1.41	0.72	0.20
10	68	0.1693	0.0358	0.2086	0.0316	0.1934	0.0405	0.1700	0.0351	1.13	0.78	5.42
11	76	-0.005246	0.001770	-0.005301	0.001499	-0.005319	0.001668	-0.006424	0.001629	1.18	0.90	0.00
11	77	0.08609	0.02057	0.10261	0.01378	0.09850	0.01736	0.11382	0.01434	1.49	0.79	1.17
11	78	0.04443	0.01655	0.05466	0.01344	0.05871	0.01748	0.03294	0.01514	1.23	0.77	1.12
11	79	0.05132	0.01592	0.04251	0.01361	0.04287	0.01813	0.04888	0.01557	1.17	0.75	1.14
11	80	0.05555	0.01607	0.05329	0.01352	0.03986	0.01833	0.04976	0.01549	1.19	0.74	0.07
11	81	-0.02291	0.01298	-0.02790	0.01133	-0.02266	0.01364	-0.01931	0.01261	1.15	0.83	0.62
12	85	-0.09550	0.03535	-0.06001	0.02053	-0.05265	0.02019	-0.09710	0.03567	1.72	1.02	1.52
12	86	0.0001700	0.0000522	0.0001283	0.0000321	0.0001153	0.0000310	0.0001721	0.0000527	1.62	1.04	1.03
12	87	0.2919	0.0499	0.2795	0.0209	0.2225	0.0292	0.3033	0.0353	2.38	0.72	0.07
12	88	0.1776	0.0420	0.1835	0.0231	0.1747	0.0309	0.1767	0.0419	1.82	0.75	0.03
12	89	0.04297	0.03810	0.04208	0.02035	0.04449	0.03040	0.04178	0.03800	1.87	0.67	0.00
12	92	0.4187	0.1066	0.3809	0.0580	0.2685	0.0914	0.4248	0.1057	1.84	0.63	0.18
12	84	-0.6013	0.2212	-0.3792	0.1285	-0.3324	0.1263	-0.6114	0.2232	1.72	1.02	1.52
13	94	-0.2770	0.0695	-0.2957	0.0356	-0.3193	0.0417	-0.3177	0.0711	1.95	0.85	0.10
13	95	-0.05756	0.01849	-0.06238	0.00945	-0.06677	0.01113	-0.07002	0.01842	1.96	0.85	0.09
13	96	-0.0002309	0.0000579	-0.0002420	0.0000310	-0.0002621	0.0000361	-0.0002591	0.0000597	1.87	0.86	0.05
13	97	0.1552	0.0288	0.1580	0.0124	0.1499	0.0166	0.1119	0.0225	2.32	0.75	0.01
13	98	-0.2999	0.1121	-0.3172	0.0553	-0.3309	0.0663	-0.2589	0.1181	2.03	0.83	0.03
13	93	1.379	0.344	1.466	0.182	1.586	0.212	1.556	0.355	1.89	0.86	0.09
15	103	0.7976	0.0397	0.7867	0.0236	0.7811	0.0272	0.7944	0.0380	1.68	0.87	0.12
15	104	0.001673	0.000295	0.001679	0.000178	0.001703	0.000211	0.001699	0.000278	1.66	0.84	0.00
15	106	-0.002194	0.001709	-0.002087	0.001084	-0.002478	0.001466	-0.002262	0.001669	1.58	0.74	0.01
15	180	-0.3134	0.1075	-0.3795	0.0581	-0.4003	0.0693	-0.2991	0.1006	1.85	0.84	0.53
15	102	0.1681	0.0353	0.1828	0.0208	0.1885	0.0236	0.1706	0.0342	1.70	0.88	0.27
16	108	0.8637	0.0537	0.8928	0.0436	0.8565	0.0500	0.8648	0.0533	1.23	0.87	0.86
16	109	0.09192	0.03197	0.07605	0.02599	0.09847	0.02967	0.08670	0.03152	1.23	0.88	0.73
16	110	-0.01515	0.00838	-0.01649	0.00684	-0.01903	0.00787	-0.00956	0.00770	1.23	0.87	0.08
16	107	0.09830	0.06536	0.06613	0.05412	0.09843	0.06089	0.11278	0.06455	1.21	0.89	0.77
24	146	-0.2212	0.0711	-0.2266	0.0562	-0.2514	0.0942	-0.2226	0.0703	1.27	0.60	0.02
24	147	0.5044	0.1600	0.5441	0.1281	0.6014	0.2165	0.5078	0.1582	1.25	0.59	0.17
24	148	0.5245	0.1452	0.5368	0.1107	0.5184	0.1196	0.4967	0.1303	1.31	0.93	0.02
24	186	0.6322	0.0726	0.6011	0.0603	0.5918	0.0917	0.6426	0.0681	1.20	0.66	0.59
24	156	0.2513	0.1263	0.2510	0.1090	0.1948	0.1766	0.2397	0.1225	1.16	0.62	0.00
24	145	0.09979	0.63898	-0.10178	0.47863	-0.29390	0.61388	0.00962	0.59560	1.31	0.78	0.24
90	172	0.8379	0.0578	0.8335	0.0486	0.8189	0.0516	0.8395	0.0553	1.19	0.94	0.02
90	173	0.04318	0.02794	0.05329	0.02384	0.06121	0.02863	0.04129	0.02719	1.17	0.83	0.48
90	174	0.04085	0.01201	0.04033	0.00865	0.03501	0.01163	0.03248	0.00946	1.39	0.74	0.00
90	175	0.04989	0.02387	0.04159	0.01137	0.02818	0.01563	0.02953	0.01186	2.10	0.73	0.16
90	176	0.01343	0.01305	0.02188	0.01104	0.02635	0.01413	0.01677	0.01263	1.18	0.78	1.48
90	177	0.03456	0.01188	0.03472	0.01026	0.03568	0.01232	0.03603	0.01160	1.16	0.83	0.00
90	178	0.2644	0.1201	0.2294	0.1045	0.2119	0.1153	0.2489	0.1166	1.15	0.91	0.35
90	171	-12.94	3.82	-12.78	2.71	-11.01	3.71	-10.19	2.97	1.41	0.73	0.00
										AVE= 1.44	AVE= 0.83	

case is $\hat{q}'(\hat{V}_3 - \hat{V}_4)^{-1}\hat{q}$, where the estimated covariance matrices $\hat{V}_3$ and $\hat{V}_4$ are defined in (9) and (12) respectively. In principle, therefore, the hypothesis of normality can be tested by computing m and comparing it to, say, the critical χ^2 value at the 95 percent confidence level. In the present case, however, $\hat{V}_3 - \hat{V}_4$ is not positive definite. This can be easily seen from the second to last column in Table 2. Each number in this column is the square root of the ratio of a diagonal element of $\hat{V}_3$ to the corresponding diagonal element of $\hat{V}_4$. A necessary condition for $\hat{V}_3 - \hat{V}_4$ to be positive definite is that all these numbers be greater than one, and this is clearly not the case. In fact, only 5 of the 106 numbers are greater than one, with the average value of all the numbers being 0.83. In other words, on average the 3SLS standard errors are less than the FIML standard errors.

There are at least two possible explanations for this somewhat puzzling result. One explanation is that the error terms are in fact not normally distributed, in which case there is no presumption that $V_3 - V_4$ is positive definite. The other explanation is based on a small sample argument. As noted above, 58 variables were used in the Z matrix for the 3SLS estimates, which with only 98 observations means that quite good fits are obtained in the first stage regressions. In other words, the predicted values of the endogenous variables from the first stage regressions are quite close to the actual values. In the case of the FIML estimates, on the other hand, we know from Hausman's [16] interpretation of the FIML estimator as an instrumental variables estimator that FIML takes into account the nonlinear restrictions on the reduced form coefficients in forming the instruments. This means that in small samples the instruments that FIML forms are likely to be based on worse first-stage fits of the endogenous variables than are the instruments that 3SLS forms. In a loose sense this situation is analogous to the fact that in the 2SLS case the more variables that are used in the first stage regressions the better is the fit in the second stage regression. If this second explanation is true, then the present results indicate that many more observations are needed before the 3SLS and FIML estimates can be used to test the normality hypothesis.[7]

Consider next comparing the FIML and 2SLS estimates. Under the null hypothesis of normally distributed errors and correct specification, both estimates are consistent, but only the FIML estimates attain the asymptotic Cramer-Rao bound. Under the alternative hypothesis of normality and incorrect specification of some subset of all the equations, all the FIML estimates are inconsistent, but only the 2SLS estimates of the incorrectly specified subset are inconsistent. The Hausman test can thus be applied one or more equations at a time to test the hypothesis that the rest of the model is correctly specified. If for some subset of the equations the m-statistic exceeds the critical value, then the test would indicate that there is misspecification somewhere in the rest of the model. Unfortunately, however, in the present case many diagonal blocks of $\hat{V}_2 - \hat{V}_4$ are not positive definite, as can be seen from Table 2, where many of the 2SLS standard errors are less than the corresponding FIML standard errors. It is thus not possible to use the Hausman test in this case.

The situation is more favorable for comparing the 3SLS and 2SLS

---

[7]We are not the first to find the FIML standard errors on average larger than the 3SLS standard errors. Although Hausman does not discuss this, for 10 of the 12 estimated coefficients of Klein's Model I in Table 1 in Hausman [15, p. 649], the FIML standard error is larger than the corresponding 3SLS standard error.

estimates, where every diagonal element of $\hat{V}_2$ is greater than the corresponding diagonal element of $\hat{V}_3$. This can be most easily seen from the third-to-last column in Table 2. In this case, however, because of the use by 2SLS of some first stage regressors not used by 3SLS, 3SLS is not necessarily asymptotically more efficient than 2SLS. The Hausman test is thus not, strictly speaking, applicable, and in fact $\hat{V}_2 - \hat{V}_3$ is not positive definite in the present case. For 4 of the 16 estimated equations, the relevant diagonal block of $\hat{V}_2 - \hat{V}_3$ is not positive definite, and so the entire matrix is obviously not positive definite.

In spite of the above problem, we have used the 2SLS and 3SLS estimates to compute the m-statistic for each of the 106 coefficients one at a time and for each of the 12 equations for which the diagonal block of $\hat{V}_2 - \hat{V}_3$ is positive definite.[8] The m-values for the 106 coefficients are presented in the last column of Table 2. The critical χ^2 value at the 95 percent confidence level for these numbers (one degree of freedom) is 3.84, and as can be seen from the table, only two of the numbers exceed this value. The null hypothesis of correct specification is thus accepted in 104 of the 106 cases. (Remember that the alternative hypothesis in each of these cases is that there is misspecification somewhere in the model other than in the particular equation that includes the coefficient.) Similar results were achieved when the test was applied one equation at a time (rather than one coefficient at a time). In none of the 12 cases

---

[8]Note that none of these tests require that the off diagonal blocks of $\hat{V}_2$ be computed. Since we knew from examining the diagonal blocks alone that $\hat{V}_2 - \hat{V}_3$ and $\hat{V}_2 - \hat{V}_4$ were not positive definite, no purpose would have been served by computing the entire V_2 matrix in (4).

was the m-value greater than the critical χ^2 value at the 95 percent confidence level. These results are thus encouraging regarding the specification of the model, but because of the above problem, they must be interpreted with considerable caution. It appears that many more observations are needed before the Hausman test can be used with much confidence for models like the present one.

V. Dynamic Prediction Accuracy

Since macroeconometric models are used to make predictions more than one period ahead, it is of some interest to examine the sensitivity of the dynamic prediction accuracy of the model to the four sets of estimates. For present purposes both static and dynamic predictions for the four sets were made for two periods, the estimation period (1954I-1978II) and the last 10 quarters of the estimation period (1976I-1978II).[9] The root mean squared errors (RMSEs) from these predictions for 6 selected variables are presented in Table 3.[10] As can be seen from the table, the results differ very little across estimators for the static predictions. The results are also fairly close for the dynamic predictions, although

---

[9] Because of the possible small sample problem for the FIML estimator discussed in Section III, no observations were excluded from the estimation period to be used for outside-sample predictions. Therefore, all the RMSEs in Table 3 are for within-sample predictions.

[10] In Fair [11] an alternative procedure to the RMSE procedure is proposed for estimating the predictive accuracy of a model. This procedure has certain advantages over the RMSE procedure, such as accounting for the fact that variances of forecast errors are not constant across time, but because it requires successive reestimation of the model, it was not used in this study. This procedure also provides a quite different way of examining the effects of misspecification than does the Hausman test, and given the problems encountered in this study in applying the Hausman test, the procedure in [11] may turn out to be more practical.

TABLE 3

Root Mean Squared Errors

	1954I-1978II (98 obs.)	GNPR	GNPD	UR	RBILL	M1	WFF
Static:	2SLS	0.64	0.31	0.27	0.45	0.84	0.58
	3SLS	0.66	0.31	0.27	0.44	0.84	0.58
	FIML	0.66	0.32	0.27	0.45	0.85	0.59
	OLS	0.66	0.31	0.28	0.44	0.85	0.58
Dynamic:	2SLS	2.05	1.62	1.19	0.92	3.41	2.13
	3SLS	2.16	1.61	1.19	0.93	3.69	2.12
	FIML	2.07	1.53	1.13	0.95	3.49	2.09
	OLS	2.02	1.77	1.24	0.95	3.42	2.26
	1976I-1978II (10 obs.)						
Static:	2SLS	0.70	0.39	0.36	0.21	0.78	0.38
	3SLS	0.73	0.43	0.35	0.21	0.73	0.40
	FIML	0.68	0.45	0.33	0.21	0.72	0.40
	OLS	0.70	0.39	0.37	0.21	0.81	0.38
Dynamic:	2SLS	1.38	0.57	0.57	0.65	1.68	0.51
	3SLS	1.53	0.65	0.46	0.56	1.51	0.71
	FIML	1.32	0.70	0.48	0.52	1.29	0.72
	OLS	1.57	0.58	0.58	0.67	1.30	0.53

Notes: GNPR = real GNP
GNPD = GNP deflator
UR = unemployment rate
RBILL = bill rate
M1 = money supply
WFF = wage rate

The RMSEs for GNPR, GNPD, M1, and WFF were computed from percentage errors. A percentage error for a given quarter is defined to be the absolute error divided by the actual value of the variable.

The RMSEs for UR and RBILL are in the natural units of the variables (percentage points).

The simulations were deterministic, with all error terms set equal to zero.

there is somewhat more variance across estimators in this case. Even in this case, however, there is no obviously superior estimator.

The fact that the results in Table 3 do not discriminate between the 2SLS and 3SLS estimates is consistent with the Hausman test results discussed in the previous section. The one perhaps surprising result in Table 3 is that the OLS RMSEs are so close to the others. In spite of some fairly large differences between the OLS coefficient estimates and the others in Table 2, this has little effect on the errors in Table 3. The main conclusion from this exercise thus appears to be that RMSE results like those in Table 3 are not good at discriminating among alternative estimators.

VI. Policy Effects

Since macroeconometric models are also used for policy purposes, it is of interest to examine the sensitivity of policy effects in the model to the four sets of estimates. Results that pertain to this issue are presented in Table 4. The numbers in this table were constructed for each set of estimates as follows. First, a base forecast was made for the 1978IV-1982IV period, with guessed values used for the exogenous variables. The same exogenous values were used for each set of estimates. From this base path the real value of government purchases of goods (XG) was increased by 10 billion dollars at an annual rate and a new forecast was generated. The effects of this change on two variables, real GNP and the GNP deflator, are presented in Table 4. Each number in the table is the difference between the predicted value of the variable after the change and the predicted value before the change.[11]

TABLE 4

Effects of a Permanent Increase in XG of 10.0 Billion Dollars at an Annual Rate

	1978	1979				1980				1981				1982				Sum over the 17
	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	Quarters
GNPR (Real GNP) (billions of 1972 dollars at an annual rate)																		
2SLS	9.4	12.2	12.9	13.0	12.3	11.4	10.3	9.2	8.3	7.4	6.7	6.1	5.6	5.2	4.9	4.7	4.5	36.0
3SLS	9.8	12.8	13.4	13.2	12.3	11.1	9.9	8.8	7.8	7.1	6.4	5.9	5.5	5.2	5.0	4.8	4.6	36.0
FIML	9.5	12.6	13.3	13.2	12.4	11.4	10.3	9.4	8.6	7.9	7.3	6.8	6.4	6.1	5.8	5.5	5.3	37.9
OLS	9.9	12.9	13.8	14.1	13.6	12.7	11.7	10.7	9.8	9.0	8.3	7.6	7.1	6.7	6.3	6.0	5.7	41.5
GNPD (GNP Deflator) (1972 = 100)																		
2SLS	0.069	0.129	0.183	0.231	0.271	0.305	0.330	0.350	0.362	0.372	0.378	0.382	0.383	0.384	0.383	0.383	0.381	
3SLS	0.068	0.125	0.174	0.218	0.255	0.287	0.313	0.334	0.350	0.363	0.374	0.384	0.391	0.397	0.403	0.408	0.412	
FIML	0.060	0.107	0.147	0.183	0.214	0.242	0.266	0.286	0.303	0.318	0.332	0.344	0.354	0.364	0.373	0.381	0.389	
OLS	0.070	0.134	0.194	0.248	0.292	0.330	0.361	0.384	0.401	0.414	0.424	0.430	0.434	0.437	0.438	0.439	0.438	

Notes: Each number is the difference between the predicted value of the variable after the change and the predicted value before the change.

The number in the last column for GNPR for each row is the sum of the other numbers in the row divided by 4.

The OLS results for real GNP in Table 4 are more expansionary than are the results for the other three estimators. The sum of the real GNP increases over the 17 quarters is 41.5 billion dollars for OLS, compared to 36.0, 36.0, and 37.9 for 2SLS, 3SLS, and FIML, respectively. The OLS estimates of the real GNP multipliers thus appear to be biased upwards, a conclusion that is consistent with simple textbook examples of the simultaneity bias of OLS estimates. Although not shown in the table, a similar result shows up for the predictions of the money supply (M1) . The sum of the M1 increases over the 17 quarters was 42.8 for OLS, compared to 23.3, 19.2, and 22.5 for 2SLS, 3SLS, and FIML, respectively. With respect to the results for the GNP deflator in Table 4, the OLS results are slightly more inflationary than are the others.

Since the OLS estimates are the only inconsistent estimates of the four sets (assuming correct specification and normality of the error terms), it is encouraging that the policy effects from the OLS estimates differ more from the others than do the others from themselves. In other words, the results in Table 4 do appear to discriminate against OLS, something which was not true of the RMSE results in Table 3.

VI. Summary and Conclusion

This study has demonstrated that it is feasible to obtain full information estimates of a fairly large nonlinear model. As can be seen from Table 1, these estimates are still not cheap, but the algorithm that

---

[11] See Fair [13] for a more detailed discussion of this experiment. The 2SLS results in Table 4 are the same as the XG results in Table 2 in [13] except that the results in [13] are based on stochastic rather than (as in Table 4) deterministic simulation of the model. It should also be noted that although the simulation period used for these results is outside of the estimation period, this need not have been the case. The experiment in Table 4 could have used a within-sample period.

has been used in this study does appear to have greatly increased their computational feasibility.

Due possibly to small sample problems, the present attempt to use the Hausman test to examine the differences among the 2SLS, 3SLS, and FIML estimates was at best only partly successful. Neither the difference between the 3SLS and FIML estimated covariance matrices nor the difference between the 2SLS and 3SLS estimated covariance matrices was positive definite. The former result was possibly due to better first-stage fits for 3SLS than for FIML. The latter result was due to the fact that the sample size prevented all the variables that were used in the first stage regressions for the 2SLS estimates from being included in the one regressor matrix (the Z matrix) for the 3SLS estimates. It is possible, as noted in Section II, to modify the 3SLS estimator to use a different set of regressors for each equation, but for the present model this estimator was not computationally feasible. It thus appears that more observations are needed before the Hausman tests can be applied with much confidence in situations like the present one.

The RMSE results in Table 3 did not reveal important differences between the OLS estimates and the others, but the policy results in Table 4 did. Judging from the results in Table 4, the OLS estimates do appear to show some simultaneity bias.

We hope that the results in this study will encourage further work on the full information estimation of models. Given the recent theoretical interest in nonlinear full information estimators and the computational results in this paper, the time for full information estimators may have finally arrived.

APPENDIX

I. The Algorithm

Since the algorithm that was used to obtain the 3SLS and FIML estimates is described in detail in Parke [20], it will only be briefly discussed here. It belongs to the class of relaxation (no derivative) algorithms. It takes advantage of two characteristics of macroeconometric models that we have observed to exist in practice. The first and most important is that the 3SLS and FIML estimates of each equation almost satisfy the property that the means of the estimated residuals is zero. For OLS and 2SLS this property is true exactly, and so it is not surprising that it is almost true for other estimators. The second characteristic is that the correlations of coefficient estimates across equations are generally less than the correlations within equations. General purpose algorithms do not take advantage of this specific structure of the 3SLS and FIML optimization problems, and this appears to be the reason they have not been successful when applied to large problems.

Many different directions are searched per iteration of the algorithm. A quadratic interpolation is used to find the maximum for each direction. The directions are generated in four basis ways: (1) For each equation the objective function is maximized by changing one coefficient at a time. For coefficients that are not constant terms or serial correlation coefficients, the constant term is at the same time adjusted so that the mean of the equation's residuals is unchanged. (The restriction across the two equations in the present model requires an adjustment to both equations' constant terms for the coefficients of these two equations.)

(2) After step (1) has been completed for a given equation, the vector of beginning coefficients is subtracted from the vector of ending coefficients, and the difference vector is taken as the next direction. This is a likely direction of increase for the whole equation. (3) After steps (1) and (2) have been completed for all the equations, the coefficient vector at the start of the iteration is subtracted from the current coefficient vector, yielding the direction of total change through the iteration. This direction, a likely direction of increase for the whole model, is then searched. Similarly, the 2SLS coefficient vector is subtracted from the current coefficient vector, and this direction of long-run change is searched. The changes since other past coefficient vectors may also be used for further searches at this point. (4) The directions from steps (2) and (3) are repeated. This completes the calculations for one iteration.

Note that the algorithm spends much of its time in steps (1) and (2), i.e., in examining one equation at a time. As alluded to above, this turns out to be an efficient use of the algorithm's time. Note also, because the constant term is changed by itself in step (1), that the algorithm does not force the estimates to satisfy the property of zero residual means. This would, of course, be wrong. With respect to step (3), the use of directions from past coefficient vectors, such as the starting vector, has in practice been of considerable help in increasing the rate of convergence of the algorithm. This type of searching is usually not done by other algorithms. Finally, note from Table 1 that the algorithm has the characteristic of increasing the objective function by a large amount in the first few iterations. This allows one to get very cheaply a good idea of what the final estimates will be like.

II. Estimating the FIML Covariance Matrix

The procedure we used to obtain the estimated FIML covariance matrix arose from an unexpected difficulty. In taking the second partial derivatives of L in (12) numerically, we found that the resulting matrix was not positive definite, in spite of several tests of alternative differentiation strategies. To avoid this problem, we instead calculated the covariance matrix for a transformed set of coefficients, using a transformation equivalent to the directions in step (1) above. This covariance matrix was positive definite, and it was also not sensitive to alternative differentiation strategies. Solving back through the transformation yielded a positive definite covariance matrix for the original coefficients. Comparing the inverse of this matrix with the unsuccessful second-derivative matrices confirmed that the straightforward approach is very sensitive to slight errors in the second derivative approximations. Again, this technique and its motivation are discussed in detail in Parke [20].

REFERENCES

[1] Amemiya, Takeshi, "The Nonlinear Two-Stage Least-Squares Estimator," Journal of Econometrics, II (July 1974), 105-110.

[2] __________, "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model," Econometrica, XLV (May 1977), 955-968.

[3] Belsley, David A., "On the Computational Competitiveness of Full Information Maximum Likelihood and Three Stage Least Squares in the Estimation of Nonlinear, Simultaneous-Equations Models," Journal of Econometrics, forthcoming.

[4] __________, "On the Efficient Computation of the Nonlinear Full-Information Maximum Likelihood Estimator," mimeo, January 1979.

[5] Chow, Gregory C., "On the Computation of Full-Information Maximum Likelihood Estimates for Nonlinear Equation Systems," The Review of Economics and Statistics, LV (February 1973), 104-109.

[6] Cochrane, D. and G. H. Orcutt, "Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms," Journal of the American Statistical Association, XLIV (March 1949), 32-61.

[7] Dagenais, Marcel G., "The Computation of FIML Estimates as Iterative Generalized Least Squares Estimates in Linear and Nonlinear Simultaneous Equations Models," Econometrica, XLVI (November 1978), 1351-1362.

[8] Fair, Ray C., "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors," Econometrica, XXXVIII (May 1970), 507-516.

[9] __________, "On the Robust Estimation of Econometric Models," Annals of Economic and Social Measurement, III (October 1974), 667-677.

[10] __________, A Model of Macroeconomic Activity. Volume II: The Empirical Model (Cambridge: Ballinger Publishing Co., 1976).

[11] __________, "Estimating the Expected Predictive Accuracy of Econometric Models," Cowles Foundation Discussion Paper No. 480, revised October 1978.

[12] __________, "The Fair Model as of November 1, 1978," mimeo.

[13] __________, "An Estimate of the Uncertainty of Policy Effects in a Macroeconometric Model," Cowles Foundation Discussion Paper No. 511, 1979.

[14] Fisher, Franklin M., Paul H. Cootner, and Martin Neil Baily, "An Econometric Model of the World Copper Industry," The Bell Journal, III (Autumn 1972), 568-609.

[15] Hausman. Jerry A., "Full Information Instrumental Variables Estimation of Simultaneous Equations Systems," Annals of Economic and Social Measurement, III (October 1974), 641-652.

[16] __________, "An Instrumental Variable Approach to Full-Information Estimators for Linear and Certain Nonlinear Econometric Models," Econometrica, XLIII (July 1975), 727-738.

[17] __________, "Specification Tests in Econometrics," Econometrica, XLVI (November 1978), 1251-1271.

[18] Jorgenson, Dale W. and Jean-Jacques Laffont, "Efficient Estimation of Nonlinear Simultaneous Equations with Additive Disturbances," Annals of Economic and Social Measurement, III (October 1974), 615-640.

[19] Klein, Laurence R., "Forecasting and Policy Evaluation Using Large Scale Econometric Models: The State of the Art," in M. D. Intriligator (ed.), Frontiers of Quantitative Economics (Amsterdam: North Holland, 1971).

[20] Parke, William R., "An Algorithm for Full Information Estimation of Large Nonlinear Models," forthcoming.

[21] Sargan, J. D., "Asymptotic Theory and Large Models," International Economic Review, XVI (February 1975), 75-91.

[22] Theil, Henri, Principles of Econometrics (New York: John Wiley & Sons, 1971).