A COMPARISON OF FIML AND ROBUST ESTIMATES OF A NONLINEAR MACROECONOMETRIC MODEL

Ray C. Fair*

Working Paper No. 15

COMPUTER RESEARCH CENTER FOR ECONOMICS AND MANAGEMENT SCIENCE
National Bureau of Economic Research, Inc.
575 Technology Square
Cambridge, Massachusetts 02139

October 1973

*Department of Economics, Princeton University, and NBER Computer Research Center. Research supported in part by National Science Foundation Grant GJ-1154X2 to the National Bureau of Economic Research, Inc.

Abstract

The prediction accuracy of six estimators of econometric models are compared. Two of the estimators are ordinary least squares (OLS) and full-information maximum likelihood (FIML). The other four estimators are robust estimators in the sense that they give less weight to large residuals. One of the four estimators is approximately equivalent to the least-absolute-residual (LAR) estimator, one is a combination of OLS for small residuals and LAR for large residuals, one is an estimator proposed by John W. Tukey, and one is a combination of FIML and LAR. All of the estimators account for first-order serial correlation of the error terms.

The main conclusion is that robust estimators appear quite promising for the estimation of econometric models. Of the robust estimators considered in the paper, the one based on minimizing the sum of the absolute values of the residuals performed the best. The FIML estimator and the combination of the FIML and LAR estimators also appear promising.

Acknowledgments

The author is indebted to David Belsley, Gregory Chow, David Hoaglin, Paul Holland, and Edwin Kuh for helpful comments and to David Jones and Rod Gretlin for research assistance.

Contents

Tables

I. Introduction

Many recent studies of robust estimation techniques have been Monte Carlo studies and have been concerned with estimating a small number of parameters.[1] The purpose of this paper is to examine the usefulness of such techniques for the estimation of econometric models. Six estimators are compared. Each estimator is first used to estimate the stochastic equations of the model described in Fair [4]. Then for each set of estimates, within-sample predictions (both static and dynamic) of the endogenous variables are generated. The estimators are compared in terms of the accuracy of the within-sample predictions. Some outside-sample predictions are also analyzed.

The methodology of this paper is similar to the methodology in Fair [6], where ten estimators were compared. The study [6] dealt only with the eight-equation linear subset of the model in [4], however, while this paper considers the nonlinear part of the model as well. The results in [6] indicate that accounting for first-order serial correlation of the error terms is quite important, and so all six

[1] See, for example, the studies of Andrews et al. [2], Andrews [1], and Hughes [9].

estimators in this paper have been modified to account for first-order serial correlation.[2]

Two of the six estimators are ordinary least squares (OLS) and full-information maximum likelihood (FIML). The other four estimators can be considered to be robust estimators in the sense that they give less weight to large residuals. One of the four estimators is approximately equivalent to the least-absolute-residual (LAR) estimator, one is a combination of OLS for small residuals and LAR for large residuals, one is an estimator proposed by John W. Tukey, and one is a combination of FIML and LAR.

The present model is both nonlinear in coefficients (after adjusting for serial correlation) and nonlinear in variables. Consequently, the standard way of obtaining LAR estimates of a linear model by converting the problem to a linear programming problem could not be used in this study, and the available programs for obtaining FIML estimates of a linear model could not be used. The procedures that were employed to obtain these estimates are described in Sections III and IV.

II. The Model

The model is described in [4] and will not be discussed in any detail here. For present purposes, the monthly housing starts sector has not been used, and housing starts have been taken to be exogenous. The

---

[2]To be consistent with the notation in [6], "AUTO1" should be added to the name of each estimator, but since all estimators in this paper are "AUTO1" estimators, this will not be done.

equations of the model are listed in Table 1. There are a few differences between the equations in Table 1 and the equations in Table 11-4 in [4], and these differences are discussed at the end of Table 1. Dummy variables $D644_t$, $D651_t$, $D704_t$, and $D711_t$ have been added to the CD_t, V_t-V_{t-1}, and M_t equations and dummy variables $D704_t$ and $D711_t$ have been added to the IP_t equation to account for the effects of the two auto strikes. These four equations were the ones most affected by the strikes. The sample period used for the estimation and simulation was 1960 II - 1973 I, a total of 52 observations.

Each stochastic equation of the model except the price equation is assumed to have a first-order serially correlated error term. For each of the six estimation techniques, first-order serial correlation was handled by transforming each equation into one with a non-serially correlated error term and then treating the resulting equation as nonlinear in the coefficients. If, for example, the equation to be estimated is:

$$(1) \quad y_t = b_1 + b_2x_t + b_3y_{t-1} + u_t,$$

where

$$(2) \quad u_t = \rho u_{t-1} + \varepsilon_t,$$

ε_t not being serially correlated, the equation can be written:

$$(3) \quad y_t = \rho y_{t-1} + b_1(1-\rho) + b_2(x_t-\rho x_{t-1}) + b_3(y_{t-1}-\rho y_{t-2}) + \varepsilon_t,$$

which is a standard nonlinear equation in the coefficients.

Table 1. The Equations of the Model

Stochastic Equations		Serial Correlation Parameter
(3.3)	$CD_t = \beta_{11} + \beta_{12}GNP_t + \beta_{13}MOOD_{t-1} + \beta_{14}MOOD_{t-2} + \beta_{15}D644_t + \beta_{16}D651_t + \beta_{17}D704_t + \beta_{18}D711_t$	ρ_1
(3.7)	$CN_t = \beta_{21}GNP_t + \beta_{22}CN_{t-1} + \beta_{23}MOOD_{t-2}$	ρ_2
(3.11)	$CS_t = \beta_{31}GNP_t + \beta_{32}CS_{t-1} + \beta_{33}MOOD_{t-2}$	ρ_3
(4.4)	$IP_t = \beta_{41} + \beta_{42}GNP_t + \beta_{43}PE2_t + \beta_{44}D704_t + \beta_{45}D711_t$	ρ_4
(5.5)	$IH_t = \beta_{51} + \beta_{52}GNP_t + \beta_{53}HSQ_t + \beta_{54}HSQ_{t-1} + \beta_{55}HSQ_{t-2}$	ρ_5
(6.15)	$V_t - V_{t-1} = \beta_{61} + \beta_{62}(CD_{t-1}+CN_{t-1}) + \beta_{63}V_{t-1} + \beta_{64}(CD_{t-1}+CN_{t-1}-CD_t-CN_t) + \beta_{65}D644_t + \beta_{66}D651_t + \beta_{67}D704_t + \beta_{68}D711_t$	ρ_6
(10.7)	$PD_t - PD_{t-1} = \beta_{71} + \beta_{72}\frac{1}{20}\sum_{i=1}^{20} GAP2_{t-i+1}$	
(9.8)	$\log M_t - \log M_{t-1} = \beta_{81} + \beta_{82}t + \beta_{83}(\log M_{t-1} - \log M_{t-1}H_{t-1}) + \beta_{84}(\log Y_{t-1} - \log Y_{t-2}) + \beta_{85}(\log Y_t - \log Y_{t-1}) + \beta_{86}D644_t + \beta_{87}D651_t + \beta_{88}D704_t + \beta_{89}D711_t$	ρ_8
(9.10)	$D_t = \beta_{91} + \beta_{92}t + \beta_{93}M_t$	ρ_9
(9.11)	$\frac{LF_{1t}}{P_{1t}} = \beta_{10,1} + \beta_{10,2}t$	ρ_{10}
(9.12)	$\frac{LF_{2t}}{P_{2t}} = \beta_{11,1} + \beta_{11,2}t + \beta_{11,3}\frac{M_t+MA_t+MCG_t+AF_t}{P_{1t}+P_{2t}}$	ρ_{11}

Table 1 (continued)

Identity Equations

Income Identity $GNP_t = CD_t + CN_t + CS_t + IP_t + IH_t + V_t - V_{t-1} + EX_t - IMP_t + G_t$

(10.5) $GAP2_t = GNPR^*_t - GNPR_{t-1} - (GNP_t - GNP_{t-1})$

(10.8) $GNPR_t = 100 \dfrac{GNP_t - GG_t}{PD_t} + YG_t$

(10.9) $Y_t = GNPR_t - YA_t - YG_t$

(9.2) $M_t H_t = \dfrac{1}{\alpha_t} Y_t$

(9.9) $E_t = M_t + MA_t + MCG_t - D_t$

(9.14) $UR_t = 1 - \dfrac{E_t}{LF_{1t} + LF_{2t} - AF_t}$

Table 1 (continued)

Definition of Symbols

CD_t	= Consumption expenditures for durable goods, SAAR
CN_t	= Consumption expenditures for nondurable goods, SAAR
CS_t	= Consumption expenditures for services, SAAR
$\dagger EX_t$	= Exports of goods and services, SAAR
$\dagger G_t$	= Government expenditures plus farm residential fixed investment, SAAR
GNP_t	= Gross National Product, SAAR
$\dagger HSQ_t$	= Quarterly nonfarm housing starts, seasonally adjusted at quarterly rates in thousands of units
IH_t	= Nonfarm residential fixed investment, SAAR
$\dagger IMP_t$	= Imports of goods and services, SAAR
IP_t	= Nonresidential fixed investment, SAAR
$\dagger MOOD_t$	= Michigan Survey Research Center index of consumer sentiment in units of 100
$\dagger PE2_t$	= Two-quarter-ahead expectation of plant and equipment investment, SAAR
$V_t - V_{t-1}$	= Change in total business inventories, SAAR
$\dagger AF_t$	= Level of the armed forces in thousands
D_t	= Difference between the establishment employment data and household survey employment data, seasonally adjusted in thousands of workers
E_t	= Total civilian employment, seasonally adjusted in thousands of workers
$\dagger GG_t$	= Government output, SAAR
$GNPR_t$	= Gross National Product, seasonally adjusted at annual rates in billions of 1958 dollars
$\dagger GNPR_t^*$	= Potential GNP, seasonally adjusted at annual rates in billions of 1958 dollars
LF_{1t}	= Level of the primary labor force (males 25–54), seasonally adjusted in thousands
LF_{2t}	= Level of the secondary labor force (all others over 16), seasonally adjusted in thousands
M_t	= Private nonfarm employment, seasonally adjusted in thousands of workers
$\dagger MA_t$	= Agricultural employment, seasonally adjusted in thousands of workers
$\dagger MCG_t$	= Civilian government employment, seasonally adjusted in thousands of workers
$M_t H_t$	= Man-hour requirements in the private nonfarm sector, seasonally adjusted in thousands of man-hours per week
$\dagger P_{1t}$	= Noninstitutional population of males 25–54 in thousands
$\dagger P_{2t}$	= Noninstitutional population of all others over 16 in thousands
PD_t	= Private output deflator, seasonally adjusted in units of 100
UR_t	= Civilian unemployment rate, seasonally adjusted
Y_t	= Private nonfarm output, seasonally adjusted at annual rates in billions of 1958 dollars
$\dagger YA_t$	= Agricultural output, seasonally adjusted at annual rates in billions of 1958 dollars
$\dagger YG_t$	= Government output, seasonally adjusted at annual rates in billions of 1958 dollars
$\dagger D644_t$	= Dummy variable: 1 in 1964 IV, 0 otherwise
$\dagger D651_t$	= Dummy variable: 1 in 1965 I, 0 otherwise
$\dagger D704_t$	= Dummy variable: 1 in 1970 IV, 0 otherwise
$\dagger D711_t$	= Dummy variable: 1 in 1971 I, 0 otherwise

Notes: † Exogenous variable.
SAAR = Seasonally adjusted at annual rates in billions of current dollars.

Table 1 (continued)

Differences between present model and model in Fair [4], Table 11-4

1. Housing starts (HSQ_t) exogenous.
2. Imports (IMP_t) exogenous.
3. Price equation (10.7) linear and length of lag is 20 rather than 8.
4. In equation (9.12), $M_t + MA_t + MCG_t$ replaces E_t.
5. Strike dummy variables added to equations (3.3), (4.4), (6.5), and (9.8).

III. The Computation of the FIML Estimates

Write the g^{th} equation of the model at time t as:

$$(4) \quad \phi_g(y_{1t},\ldots, y_{Gt}, x_{1t},\ldots,x_{Nt}, \beta_g) = u_{gt} \qquad \begin{matrix}(g=1,\ldots,G)\\(t=1,\ldots,T)\end{matrix} ,$$

where the y_{it} are endogenous variables, the x_{it} are predetermined variables, β_g is a vector of unknown parameters, and u_{gt} is an error term. The FIML estimates of the unknown parameters in (4) are obtained by maximizing

$$(5) \quad L = -\frac{1}{2} T \log |S| + \sum_{t=1}^{T} \log |J_t|$$

with respect to the unknown parameters,[3] where

$$(6) \quad S = (s_{gh});\ s_{gh} = \frac{1}{T} \sum_{t=1}^{T} u_{gt}u_{ht};\ J_t = \left(\frac{\partial \phi_g}{\partial y_{ht}}\right),\ g, h = 1,\ldots, G.$$

If G-M of the G equations are identities, then S is MxM, but J_t remains GxG.

There are a number of approaches that can be tried to maximize L. The results in Fair [5] indicate that quite large unconstrained maximization problems can be solved using algorithms that either do not require derivatives or for which derivatives are obtained numerically. The approach in [5] is the approach taken in this paper. Three algorithms were used: the 1964 algorithm of Powell [11], which does not require any derivatives; a member of the class of gradient algorithms considered by Huang [8], which requires first derivatives; and the quadratic hill-climbing algorithm of Goldfeld, Quandt, and Trotter [7], which requires both first and second derivatives. All derivatives were obtained numerically. See [5] for more discussion of these algorithms and for a discussion of the computation of numeric derivatives.

---

[3] See, for example, Chow [3].

The model in Table 1 decomposes naturally into two blocks: a linear, simultaneous block and a nonlinear, recursive block. FIML estimates were first obtained for the two blocks separately, using the ordinary least squares estimates as starting points, which required estimating 38 and 23 coefficients, respectively. FIML estimates of all 61 coefficients were then obtained, using the FIML estimates of the two blocks as starting points. In contrast to the work in [5], no systematic attempt was made in this study to compare the various algorithms, and so no results using alternative algorithms will be presented here. Powell's no-derivative algorithm was usually used first to obtain an answer, and then this answer was checked by starting the gradient and quadratic-hill-climbing algorithms from the answer to see if a larger value of the likelihood function could be found. In some cases a larger value was found using the other two algorithms, and in some cases the quadratic-hill-climbing algorithm found a larger value than did the gradient algorithm. In general it appeared that the FIML computational problem here was not as well behaved nor as robust to the use of different algorithms as was the optimal control problem in [5].

The present approach to obtaining the FIML estimates has the advantage of requiring little human effort. Given that algorithm and numeric-derivative programs are available, one needs only to write a simple program to compute the value of L for a given vector of coefficients. In the present case J_t can be factored into two parts: one that is a function of some of the coefficients but not of time and one that is a function of time but not of any coefficients. Consequently, the determinant of J has to be computed only once per evaluation of L rather than the T times required for the more general case. The more general case can be handled by the present approach, however, since all the more general case does is increase the computer time required per evaluation of L. The extra programming effort required for the more general case is quite small.

IV. The Computation of the Robust Estimates

Least-absolute-residual (LAR) estimates of equation (4) are obtained by minimizing

$$(7) \quad Q = \sum_{t=1}^{T} |u_{gt}|$$

with respect to the unknown parameters. Since in the present case u_{gt} is a nonlinear function of the unknown parameters because of the serial correlation assumption, Q cannot be minimized through the solution of a linear programming problem. An attempt was made in this study to minimize Q by using the approach and algorithms discussed in Section III, but this attempt failed. The algorithms were not in general successful in finding global optima. Often they converged to different answers for different starting points, and many times different algorithms converged to different answers from the same starting point.

LAR estimates can, however, be obtained, at least approximately, by converting the problem to a weighted-least-squares problem. Rewrite Q as:

$$(8) \quad Q = \sum_{t=1}^{T} \frac{(u_{gt})^2}{|u_{gt}|} \quad .$$

The problem of minimizing Q in (8) is merely a weighted-least-squares problem if the denominator is known. An iterative procedure can thus be used to minimize Q. Initial estimates of the residuals are first obtained, say by ordinary least squares, and are then used as weights to obtain new estimates of the parameters and residuals by weighted least squares. These new residual estimates are then used as new weights to obtain new parameter and residual estimates, and so on. In the present case, unweighted ordinary-least-squares estimates were used to begin the iteration, and the program was allowed to iterate four times thereafter. The estimates usually changed only slightly after the first or second weighted-least-squares estimates (the first or second iteration following the initial ordinary-least-squares estimates). The problem of zero residual estimates (making weighted-least-squares estimates on the next iteration impossible to obtain) was avoided by setting residual estimates less than a small number ε in absolute value equal to ε. For present purposes, ε was taken to be .00001.

Both the unweighted- and weighted-least-squares problems in the present case are nonlinear problems, and the estimates had to be obtained by a nonlinear technique. The degree of nonlinearity, however, is not great, being due only to the presence of the serial correlation parameter, and hence the problems could be easily solved using standard algorithms.[4] Because the program was allowed only four iterations and because of the ε treatment of very small residuals, the estimates obtained by the

---

[4] The algorithm used in this case is the algorithm programmed into TROLL at the Computer Research Center of the National Bureau of Economic Research. This same algorithm was also used in the computation of the WLS-II and WLS-III estimates described below.

above procedure are not exactly LAR estimates, but for practical purposes they should be quite close. This estimator will be called WLS-I.

The second weighted-least-squares estimator considered is a combination of OLS for small residuals and LAR for large residuals. For this estimator the denominator in (8) was still taken to be $|u_{gt}|$ if $|u_{gt}| \geq k$, but was taken to be k if $|u_{gt}| < k$. The value of k was taken to be a robust estimate of the standard error of the regression, namely $\hat{m}/.6745$, where $\hat{m}$ is the median of the absolute value of the estimated residuals.[5] The WLS-I estimates were used as starting points, and the program was allowed to run for four iterations. The median of the absolute value of the residual estimates was reestimated at each iteration, and the value of k was changed from iteration to iteration. This estimator will be called WLS-II.

The third weighted-least-squares estimator considered weights each residual as[6]

$$\left[1 - \left(\frac{z}{k_1}\right)^2\right]^2 \quad \text{if} \quad |z| \leq k_1$$

and 0 otherwise, where

$$z = \frac{u_{gt}}{k_2} \quad .$$

---

[5]See Andrews et al. [2] for a use of this estimator.

[6]The weights used for this estimator are to be compared to $1/|u_{gt}|$ for the WLS-I estimator and $1/|u_{gt}|$ or $1/k$ for the WLS-II estimator.

This estimator is attributed to John W. Tukey by Andrews [1]. Values for k_1 of both 6 and 9 have been proposed, and the value of 6 was used for this study. The value of k_2 was taken to be $\hat{m}/.6745$, where again $\hat{m}$ is the median of the absolute value of the residuals. The WLS-I estimates were used as starting points, and the program was allowed to run for four iterations. The value of k_2 was changed from iteration to iteration. This estimator will be called WLS-III.

All three of the weighted-least-squares estimators in this section are single-equation estimators and do not take into account the problems associated with estimating systems of equations.

V. The Combination of the FIML and Robust Estimators

Considering robust estimators as weighted-least-squares estimators, it is quite straightforward to combine the FIML and robust estimators. Consider, for example, the WLS-I estimator, which in the single-equation case weights each residual by $1/|u_{gt}|$. The natural extension to the FIML case is to consider maximizing

$$(9) \qquad L^* = -\frac{1}{2} T \log |S^*| + \sum_{t=1}^{T} \log |J_t| \quad ,$$

where

$$(10) \qquad S^* = (s^*_{gh}); \quad s^*_{gh} = \frac{1}{T} \sum_{t=1}^{T} \frac{u_{gt} u_{ht}}{\sqrt{|u_{gt}|}\sqrt{|u_{ht}|}}, \; g, h = 1, \ldots, G,$$

and where J_t is the same as in (6). Given an initial set of residual estimates to be used as weights, L^* can be maximized with respect to the unknown parameters. In the maximization process each residual is weighted by one over the square root of the absolute value of the initial residual estimate. Weighting schemes other than the one used for WLS-I could also be proposed, which would merely change the computation of s^*_{gh} in (10).

For purposes here only the WLS-I weighting scheme was combined with FIML. The weights were taken from the WLS-I residual estimates, with residual estimates of less than .00001 being set equal to .00001. Given the weights, L* was maximized using the same algorithms that were used to maximize L. The experience maximizing L* using the algorithms was similar to the experience maximizing L, although the problem of maximizing L* seemed slightly more difficult. Because of cost considerations, no iterations on the weights were performed. In other words, L* was only maximized once, and the new residual estimates from this solution were not used to construct new weights to be used for a second maximization, and so on. This estimator will be called FIMLWLS-I.

VI. Within-Sample Comparison of the Six Sets of Estimates

In Table 2 the six sets of estimates are presented for each of the eleven stochastic equations. The two sets of FIML estimates tend to differ more from the other four sets of estimates than the other four sets of estimates differ from each other. In particular, this is true for the coefficient estimates of the inventory equation and for the estimate of the constant term, β_{71}, in the price equation. There were no important cases of sign reversals among the different estimates of the same parameter. The only sign reversals occurred for ρ_3 and for two dummy-variable coefficients, β_{67} and β_{89}.

Table 2

The Six Sets of Coefficient Estimates of the Model

Coefficient		Method					
		OLS	FIML	WLS-I	FIMLWLS-I	WLS-II	WLS-III
1.	β_{11}	-37.66	-32.59	-36.33	-33.43	-37.03	-35.96
2.	β_{12}	.1158	.1135	.1134	.1134	.1140	.1141
3.	β_{13}	.0900	.1413	.0900	.0354	.1099	.1050
4.	β_{14}	.1437	.0564	.1502	.1682	.1345	.1251
5.	β_{15}	-2.236	-2.144	-2.359	-1.943	-2.366	-2.324
6.	β_{16}	2.459	2.302	2.308	2.827	2.384	2.441
7.	β_{17}	-6.369	-6.756	-5.869	-5.829	-6.315	-6.389
8.	β_{18}	1.068	2.543	2.045	2.101	1.345	1.186
9.	ρ_1	.5832	.3162	.5638	.5462	.5216	.5568
10.	β_{21}	.05815	.04085	.04809	.04866	.05145	.05362
11.	β_{22}	.7792	.8522	.8169	.8146	.8053	.7972
12.	β_{23}	.04802	.02771	.04608	.04635	.04515	.04504
13.	ρ_2	-.1195	-.2716	-.2379	-.2556	-.2187	-.1739
14.	β_{31}	.03584	.02802	.03708	.03504	.03579	.03727
15.	β_{32}	.8891	.9186	.8843	.8919	.8885	.8829
16.	β_{33}	-.02338	-.02074	-.02402	-.02347	-.02214	-.02241
17.	ρ_3	.2694	.1293	.0286	-.0560	.2044	.3216
18.	β_{41}	-10.32	-9.54	-8.62	-8.59	-11.21	-12.20
19.	β_{42}	.07964	.07734	.07395	.07603	.07350	.07412
20.	β_{43}	.4707	.4942	.5163	.4968	.5685	.5804
21.	β_{44}	-3.908	-3.844	-4.517	-4.322	-4.151	-3.898
22.	β_{45}	-1.947	-2.218	-2.618	-2.777	-2.292	-1.791
23.	ρ_4	.8514	.8650	.8458	.8345	.8825	.8983

Table 2 (continued)

Coefficient		OLS	FIML	WLS-I	FIMLWLS-II	WLS-II	WLS-III
24.	β_{51}	-20.71	-28.23	-15.57	-30.38	-16.32	-18.16
25.	β_{52}	.03339	.04249	.02705	.02690	.02833	.03108
26.	β_{53}	.07631	.06178	.07854	.08202	.07836	.07612
27.	β_{54}	.07416	.07379	.08242	.08271	.08208	.07856
28.	β_{55}	.03444	.02583	.03548	.03650	.03460	.03386
29.	ρ_5	.9427	.9616	.9260	.9942	.9149	.9278
30.	β_{61}	-29.40	-65.08	-30.79	-60.32	-33.15	-30.95
31.	β_{62}	.4081	.5755	.3984	.5461	.4045	.4279
32.	β_{63}	-.3139	-.4090	-.3047	-.4104	-.3069	-.3372
33.	β_{64}	.3736	.1786	.3345	.3299	.3353	.3188
34.	β_{65}	-2.435	-1.695	-2.000	-2.966	-2.182	-2.203
35.	β_{66}	4.964	3.975	4.998	5.439	4.842	4.784
36.	β_{67}	-1.825	.397	-1.712	-1.446	-1.623	-1.485
37.	β_{68}	5.219	2.998	4.705	5.559	4.853	4.912
38.	ρ_6	.9683	.9029	.9493	.9191	.9533	.9648
39.	β_{71}	1.361	1.220	1.365	1.156	1.366	1.364
40.	β_{72}	-.02508	-.02503	-.02510	-.02208	-.02484	-.02510
41.	β_{81}	-.4629	-.5425	-.4127	-.4929	-.4660	-.4605
42.	β_{82}	.00007157	.00007172	.00007650	.00007202	.00007685	.00007216
43.	β_{83}	-.1267	-.1485	-.1127	-.1351	-.1275	-.1261
44.	β_{84}	.07038	.03545	.10200	.05477	.08091	.07566
45.	β_{85}	.1751	.2605	.2064	.2312	.1664	.1729
46.	β_{86}	.00202	.00315	.00197	.00212	.00169	.00191
47.	β_{87}	.00154	.00111	.00123	.00036	.00152	.00155
48.	β_{88}	-.00290	-.00161	-.00319	-.00173	-.00341	-.00309
49.	β_{89}	.00365	-.00034	.00233	.00199	.00370	.00365
50.	ρ_8	.2648	.3089	.2073	.3684	.2552	.2495

Table 2 (continued)

Coefficient		Method					
		OLS	FIML	WLS-I	FIMLWLS-II	WLS-II	WLS-III
51.	β_{91}	-16974.	-18809.	-17808.	-18859.	-17494.	-17496
52.	β_{92}	-126.2	-142.9	-136.1	-144.2	-137.3	-135.7
53.	β_{93}	.4884	.5383	.5137	.5400	.5104	.5085
54.	ρ_{9}	.6768	.5910	.6418	.6175	.6226	.6352
55.	$\beta_{10,1}$	1.001	1.000	.999	1.000	1.000	1.003
56.	$\beta_{10,2}$	-.0004472	-.0004416	-.0004261	-.0004343	-.0004394	-.0004681
57.	ρ_{10}	.7883	.7703	.7835	.8007	.7779	.7883
58.	$\beta_{11,1}$	.2679	.2368	.2697	.2540	.2503	.2621
59.	$\beta_{11,2}$	.0008282	.0008153	.0009257	.0008304	.0009424	.0009060
60.	$\beta_{11,3}$	.2401	.2933	.2239	.2654	.2524	.2382
61.	ρ_{11}	.8642	.8467	.8371	.8778	.8462	.8597

The root mean square errors and mean absolute errors for six variables are presented in Table 3 for each of the six estimators. The six variables are GNP in current dollars (GNP_t), the private output deflator (PD_t), GNP in constant dollars ($GNPR_t$), private nonfarm employment (M_t), the difference between establishment employment data and household survey employment data (D_t), and the level of the secondary labor force (LF_{2t}). The errors for the six variables are not independent of one another in the sense that, for example, large errors in predicting GNP_t are likely to lead to large errors in predicting the other variables. GNP_t is determined in the linear, simultaneous-equations block of the model, and the other variables are determined in the nonlinear, recursive block. The five variables presented in Table 3 from the recursive block are the five most important variables in the block. The estimates of the serial correlation coefficients were used in the generation of the predictions from the model.

The results in Table 3 are fairly self-explanatory. Consider GNP_t first. OLS is obviously the worst, being last on all grounds except the one- and two-quarter-ahead predictions, where it is better than FIMLWLS-I. WLS-I is better than WLS-II and WLS-III for the three-quarter-ahead predictions and beyond, beating them on all counts, although not by much for the three-quarter-ahead prediction. For the one- and two-quarter-ahead predictions, the results are close. FIML does well for all but the simulation over the entire period, where it falls down somewhat. FIMLWLS-I is the best for the simulation over the entire period, but is not particularly good for the other predictions.

Consider PD_t next. The two FIML estimators are the worst, which is caused in large part by the different estimates of the constant term

Table 3. Prediction Errors for GNP_t, PD_t, $GNPR_t$, M_t, D_t, LF_{2t}

RMSE = Root Mean Square Errors

MAE = Mean Absolute Errors

Estimator	RMSE: Number of Quarters Ahead 1	2	3	4	5	Entire Period	MAE: Number of Quarters Ahead 1	2	3	4	5	Entire Period
	52 obs.	51 obs.	50 obs.	49 obs.	48 obs.	52 obs.	52 obs.	51 obs.	50 obs.	49 obs.	48 obs.	52 obs.
GNP_t												
OLS	3.64	6.08	7.78	9.11	10.26	14.00	2.84	4.89	6.37	7.73	8.82	11.78
FIML	3.50	5.85	7.28	8.21	9.02	13.32	2.80	4.72	5.72	6.56	7.20	10.16
WLS-I	3.59	5.88	7.27	8.21	8.86	9.63	2.82	4.58	5.73	6.46	6.93	7.73
FIMLWLS-I	3.89	6.21	7.63	8.60	9.39	9.37	3.22	4.75	5.72	6.23	6.64	7.33
WLS-II	3.56	5.83	7.30	8.35	9.20	11.99	2.74	4.62	5.84	6.92	7.66	9.67
WLS-III	3.60	5.88	7.43	8.61	9.65	13.68	2.76	4.69	6.01	7.19	8.10	11.24
PD_t												
OLS	0.29	0.45	0.56	0.67	0.77	2.99	0.22	0.35	0.43	0.50	0.58	2.57
FIML	0.32	0.52	0.70	0.87	1.05	3.19	0.26	0.44	0.61	0.77	0.95	2.54
WLS-I	0.29	0.45	0.56	0.67	0.77	2.16	0.22	0.35	0.43	0.50	0.58	1.97
FIMLWLS-I	0.33	0.53	0.70	0.88	1.08	2.75	0.25	0.43	0.60	0.76	0.96	2.21
WLS-II	0.29	0.45	0.56	0.67	0.77	2.29	0.22	0.35	0.43	0.50	0.57	2.07
WLS-III	0.29	0.45	0.56	0.66	0.77	2.60	0.22	0.35	0.43	0.50	0.57	2.28
$GNPR_t$												
OLS	3.05	5.06	6.59	7.79	8.87	20.39	2.39	4.08	5.46	6.56	7.71	17.32
FIML	2.87	4.72	6.08	7.16	8.28	20.20	2.34	3.68	4.89	5.92	6.78	15.04
WLS-I	2.94	4.78	6.01	6.84	7.50	15.03	2.36	3.79	4.62	5.31	5.98	13.24
FIMLWLS-I	2.99	4.65	5.70	6.46	7.25	18.15	2.44	3.55	4.40	5.21	6.14	13.81
WLS-II	2.98	4.83	6.16	7.16	8.02	17.01	2.37	3.88	4.99	5.86	6.72	14.84
WLS-III	3.02	4.89	6.30	7.43	8.47	18.91	2.37	3.93	5.16	6.17	7.14	16.14

Table 3 (continued)

Estimator	RMSE: Number of Quarters Ahead 1 (52 obs.)	2 (51 obs.)	3 (50 obs.)	4 (49 obs.)	5 (48 obs.)	Entire Period (52 obs.)	MAE: Number of Quarters Ahead 1 (52 obs.)	2 (51 obs.)	3 (50 obs.)	4 (49 obs.)	5 (48 obs.)	Entire Period (52 obs.)
M_t—												
OLS	166	296	416	516	602	1618	138	245	327	422	479	1423
FIML	168	291	397	483	566	1718	137	235	325	393	459	1267
WLS-I	167	290	395	475	540	1106	134	239	324	403	460	943
FIMLWLS-I	168	291	396	477	549	1707	135	239	324	402	472	1259
WLS-II	166	295	414	508	588	1374	137	242	331	422	478	1188
WLS-III	165	293	408	504	587	1490	137	241	323	415	468	1298
D_t—												
OLS	209	247	290	321	344	804	167	192	225	245	253	733
FIML	217	244	277	295	314	922	171	196	223	240	243	736
WLS-I	210	242	276	294	305	587	165	193	221	237	241	523
FIMLWLS-I	217	244	278	292	299	899	171	192	221	232	233	712
WLS-II	211	243	280	301	314	689	164	188	219	235	234	619
WLS-III	210	243	280	303	319	749	165	187	216	235	234	678
$L\ F_{2t}$—												
OLS	218	272	295	313	321	357	171	230	245	246	269	271
FIML	220	281	314	341	357	501	172	244	270	280	302	410
WLS-I	220	273	298	315	321	365	171	233	249	251	266	287
FIMLWLS-I	220	282	316	347	374	559	174	244	273	286	312	506
WLS-II	220	272	292	304	303	380	169	227	240	234	248	288
WLS-III	219	270	291	305	307	361	170	227	239	236	252	272

in the PD_t equation. The results for the other four estimators are quite close except for the simulation over the entire period, where the ranking is WLS-I, WLS-II, WLS-III, and OLS. This ranking is the same as that for GNP_t for the simulation over the entire period, which is explained by the fact that for the simulation over the entire period the perdictions of GNP_t have an important effect on the predictions of PD_t.

For $GNPR_t$, OLS is again the worst, being last on all grounds. WLS-I is better than WLS-II and WLS-III on all grounds. FIML does better than WLS-I for the one-and two-quarter-ahead predictions, even considering the poorer FIML predictions of PD_t,which are used in the computation of the predictions of $GNPR_t$, but the opposite is true for the three-quarter-ahead predictions and beyond. FIMLWLS-I is the best for the two-through five-quarter-ahead predictions, but falls down slightly for the other two.

For M_t, the results are fairly close except for the simulation over the entire period, where the RMSE ranking is WLS-I, WLS-II, WLS-III, OLS, FIMLWLS-I, and FIML, and the MAE ranking is WLS-I, WLS-II, FILMWLS-I, FIML, WLS-III, and OLS. For D_t, WLS-I does consistently well, but the results are again fairly close except for the simulation over the entire period. For LF_{2t}, the FIML estimators get worse as the period ahead lengthens. For the simulation over the entire period, OLS is best by a slight amount.

The following is a tentative list of conclusions drawn from the results in Table 3.

1. WLS-I appears better than WLS-II and WLS-III, and all three appear better than OLS. In this regard it is interesting to note that it is not just the treatment of large residuals that appears important, since WLS-II, which is a combination of OLS for small residuals and WLS-I

for large residuals, does not do as well as WLS-I. The different treatment of small residuals by WLS-I compared with OLS appears also to be of importance.

2. For the predictions of GNP_t, FIML is obviously better than OLS, which is the same conclusion reached in [6]. For the other variables, which are not determined simultaneously, FIML is not always better. In other words, more gain appears likely from using FIML over OLS when the model is simultaneous than when it is recursive.

3. Among WLS-I, FIML, and FIMLWLS-I there is no obvious winner since the rankings differ depending on the variable predicted and the number of periods ahead for which the prediction is made. Overall, however, WLS-I probably has an edge, especially if emphasis is put on the results for the variables in the recursive block, where FIML and FIMLWLS-I do not in general do particularly well, Given the success of WLS-I, it may be of interest in future work to examine the performance of the combination of two-stage least squares and WLS-I.[7]

4. For the one-quarter-ahead (static) predictions, the results are all fairly close, which means that if one is only interested in static predictions, the choice of an estimator is not too important (assuming the estimator accounts for first-order serial correlation). For dynamic predictions the choice is important, and a conclusion reached in [6]

---

[7]One obvious way to combine two-stage least squares and WLS-I is simply to run first-stage regressions in the usual way and use the fitted values of the endogenous variables from these regressions in place of the actual values of the right-hand-side endogenous variables in the present procedure of obtaining WLS-I estimates.

is also relevant here, namely that more work ought to be done on developing estimators that take into account the fact that values of the lagged endogenous variables are not known after the one-period-ahead predictions.

It should finally be noted that predictions were also generated based on WLS-I estimates obtained after the first iteration from ordinary least squares (rather than after the fourth iteration as above). The results were better than the OLS results, but not as good as the WLS-I results based on four iterations. Iterating more than once clearly improved the prediction accuracy of the estimator.

VII. Outside-Sample Comparisons of OLS and WLS-I Estimates

In order to see if the superiority of WLS-I over OLS also held up for outside-sample predictions, the model was reestimated by WLS-I and OLS only through 1968 IV. Predictions for the 1969 I - 1973 I period were then generated based on these two sets of estimates. In Table 4, error measures for the simulation over the entire prediction period (17 observations) are presented for fifteen variables. For GNP_t, WLS-I outperforms OLS. Of the six components of GNP_t, WLS-I is better for three. Of the other eight variables, which are determined in the recursive block, WLS-I is better for all but two (M_t and UR_t). Overall, WLS-I appears to outperform OLS,[8] although the superiority of WLS-I here does not appear as pronounced as it was for the within-

[8] This conclusion is consistent with the results of Meyer and Glauber [10], who found the LAR estimator to be an improvement over ordinary least squares in terms of outside-sample, single-equation prediction accuracy.

Table 4. Outside-Sample Prediction Errors for Fifteen Variables.

Estimation Period: 1960 II - 1968 IV

Prediction Period: 1969 I - 1973 I
(Error measures for the simulation over the entire prediction period only)

RMSE = Root Mean Square Errors

MAE = Mean Absolute Errors

	RMSE		MAE	
	OLS	WLS-I	OLS	WLS-I
GNP_t	13.48	9.84	10.76	8.22
CD_t	4.63	3.94	3.71	3.15
CN_t	11.24	8.27	9.55	7.10
CS_t	2.13	2.32	1.80	1.97
IP_t	2.89	3.36	2.43	2.83
IH_t	5.91	6.14	4.45	4.65
$V_t - V_{t-1}$	6.80	6.84	6.08	6.08
PD_t	0.85	0.82	0.72	0.69
$GNPR_t$	8.23	7.46	6.64	5.81
M_t	421.	468.	355.	429.
D_t	500.	376.	434.	322.
E_t	729.	696.	565.	536.
LF_{1t}	260.	240.	229.	207.
LF_{2t}	2276.	2230.	2109.	2067.
UR_t	.0163	.0164	.0149	.0150

sample comparisons. This same conclusion also emerged from examining the predictions for the 1969 I - 1973 I period in more detail (e.g., by the number of periods ahead predicted) and from examining predictions for the 1970 III - 1973 I period based on estimates through 1970 II. All of the outside-sample comparisons are, of course, based on only a small number of different periods predicted and so must be interpreted with some caution.

VIII. Conclusion

The main conclusion of this paper is that robust estimators appear quite promising for the estimation of econometric models. Of the robust estimators considered in this paper, the one based on minimizing the sum of the absolute values of the residuals performed the best. The FIML estimator and the combination of the FIML and least-absolute-residual estimators also appear promising, at least for simultaneous equations models.

The same caveats discussed in [6] regarding the methodology of that study are also relevant here. The comparisons in this paper are based only on the criterion of prediction accuracy, and the model used for the comparisons has some special features that are not characteristic of other models. Whether the conclusions reached in [6] and in this paper hold for other models is an open question, and the conclusions are merely put forth as indicating what might be the case for such models.

References

[1] Andrews, D.F., "Some Monte Carlo Results on Robust/Resistant Regression," mimeographed.

[2] Andrews, D.F. et al., Robust Estimates of Location. (Princeton: Princeton University Press, 1972).

[3] Chow, Gregory C., "On the Computation of Full-Information Maximum Likelihood Estimates for Non-Linear Equations Systems", Research Memorandum No. 142, Econometric Research Program, Princeton University, June 1972.

[4] Fair, Ray C., A Short-Run Forecasting Model of the United States Economy. (Lexington: D.C. Heath and Co., 1970).

[5] Fair, Ray C., "On the Solution of Optimal Control Problems as Maximization Problems", Annals of Economic and Social Measurement, forthcoming 1973.

[6] Fair, Ray C., "A Comparison of Alternative Estimators of Macroeconomic Models", International Economic Review, XIV (June 1973), pp.261-77.

[7] Goldfeld, Stephen M., Richard E. Quandt, and Hale F. Trotter, "Maximization by Quadratic-Hill-Climbing," Econometrica, XXXIV (July 1966), pp. 541-551.

[8] Huang, H.Y., "Unified Approach to Quadratically Convergent Algorithms for Function Minimization", Journal of Optimization Theory and Applications, V (June 1970), pp. 405-423.

[9] Hughes, Nan, "Two Empirical Studies of Some Robust Estimation Procedure in Multiple Regression", paper delivered at the Conference on Robust Regression at the National Bureau of Economic Research, Cambridge, Massachusetts, June 21, 1973.

[10] Meyer, John R., and Robert R. Glauber, Investment Decisions, Economic Forecasting, and Public Policy. (Boston: Division of Research, Graduate School of Business Administration, Harvard University, 1964).

[11] Powell, M.J.D., "An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives", Computer Journal, VII (July 1964), pp. 155-162.