Annals of Economic and Social Measurement, 3/4, 1974

ON THE ROBUST ESTIMATION OF ECONOMETRIC MODELS

BY RAY C. FAIR*

The computational aspects of obtaining robust estimates of a general nonlinear econometric model are described, and some results of estimating a particular model are presented. When robust estimators are considered as weighted-least-squares estimators, it clearly appears feasible, by a combination of solving unconstrained optimization problems and iterating, to obtain robust estimates of econometric models. In estimating the particular model, the robust estimators performed well in terms of prediction accuracy.

Most of the work that has been done on robust estimation techniques has been concerned with the estimation of a small number of parameters.[1] This paper considers the use of such techniques for the estimation of econometric models. The computational aspects of obtaining robust estimates of a general nonlinear econometric model are described, and then some results of estimating a particular model are presented. The particular model, described in Fair [4], is nonlinear in both variables and parameters, and the version used here consists of 11 stochastic equations and has 61 unknown parameters to estimate.

1. THE COMPUTATION OF ROBUST ESTIMATES OF ECONOMETRIC MODELS

Write the g-th equation of the model to be estimated as:

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

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.

It will be useful to consider first the estimation of the model by full information maximum likelihood (FIML). The FIML estimates of the unknown parameters in (1) are obtained by maximizing

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

with respect to the unknown parameters,[2] where

$$(3) \qquad S = (s_{gh}); \quad s_{gh} = \frac{1}{T}\sum_{t=1}^{T} u_{gt}u_{ht}; \quad J_t = \left(\frac{\partial \phi_g}{\partial y_{ht}}\right), \quad g, h = 1, \ldots, G.$$

If $G - M$ of the G equations are identities, then S is $M \times M$, but J_t remains $G \times G$.

* Research supported in part by National Science Foundation Grant GJ-1154X2 to the National Bureau of Economic Research, Inc. 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. Computations reported here were performed in part on the TROLL system.

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

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

The maximization of L in (2) is a computationally difficult problem, and few nonlinear models of any size have been estimated by FIML. There has, however, been recent progress in the development of algorithms for solving unconstrained optimization problems. Some of these algorithms were tested and compared in Fair [5] in the context of solving optimal control problems; the results indicate that large unconstrained optimization problems can be solved. One problem of 239 unknown parameters was solved, and problems of 100 parameters were solved routinely. Another encouraging aspect of these results is that analytic derivatives were never used. If an algorithm required first or second derivatives, the derivatives were always obtained numerically. The advantage of not having to compute analytic derivatives is the human effort saved. When numeric derivatives are used, the only human effort needed to set up the problem (other than acquiring the algorithm programs) is to write a program to compute the value of the objective function for a given vector of parameters.[3] The three main algorithms considered in [5] were 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. These are the algorithms that were used to obtain the FIML estimates for the results in Section 2.

Consider next the estimation of a single equation of (1) by the least-absolute-residual (LAR) technique, a type of robust estimator. The LAR estimates are obtained by minimizing

$$Q = \sum_{t=1}^{T} |u_{gt}| \tag{4}$$

with respect to the unknown parameters. Since in general u_{gt} is a nonlinear function of the unknown parameters, Q cannot be minimized through the solution of a linear programming problem. An attempt was first made in this study to minimize Q for the results in Section 2 by using the approach and algorithms discussed above, but this attempt failed. The algorithms were not in general successful in finding global optima. Often they converged to different answers from 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 into a weighted-least-squares problem. Rewrite Q as:

$$Q = \sum_{t=1}^{T} \frac{(u_{gt})^2}{|u_{gt}|}. \tag{5}$$

The problem of minimizing Q in (5) 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

[3] For the FIML problem, derivatives are, of course, involved in computing J_t in (2). In most cases of this type it is probably better to obtain analytic expressions for the derivatives that are involved in the direct computation of the objective function, rather than to compute these derivatives numerically as well.

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. If u_{gt} is a nonlinear function of the parameters, then a nonlinear optimization problem has to be solved to obtain the weighted-least-squares estimates for each iteration. This type of a nonlinear optimization problem is, however, usually easy to solve. In the iterative technique just described some account has to be taken of zero or near zero residual estimates.[4] The easiest way to handle this is to set residual estimates that are less than some small number ε in absolute value equal to ε. For the work in Section 2, ε was taken to be 0.00001, and the program was allowed to run for four iterations. The estimates were usually changing only slightly after the first or second iteration following the initial ordinary-least-squares estimates. Because of the ε-treatment of small residuals, the estimates obtained by the procedure just described will not be exactly LAR estimates, but for practical purposes they should be quite close. The estimates obtained by this procedure will be called WLS-I estimates.

Many other robust estimators can be considered as weighted-least-squares estimators; two of these were used for the work in Section 2. The first is a combination of ordinary-least-squares for small residuals and LAR for large residuals. For this estimator the denominator in (5) 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 $\hat{m}/0.6745$, 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 second of the other weighted-least-squares estimators weights each residual as[6]

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

and 0 otherwise, where

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

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 $\hat{m}/0.6745$, 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.

[4] In the linear case, the true optimum will, of course, correspond to k of the residual estimates being exactly zero, where k is the number of parameters estimated.

[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.

Both WLS-II and WLS-III also require that a nonlinear optimization problem be solved for each iteration if u_{gt} is a nonlinear function of the parameters; but again this type of problem is usually easy to solve.

The robust estimators considered so far are single-equation estimators and do not take into account the problems associated with estimating systems of equations. Nevertheless, when robust estimators are considered as weighted-least-squares estimators, it is easy to modify, say, the FIML estimator to be a robust estimator. 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

$$L^* = -\tfrac{1}{2}T \log |S^*| + \sum_{t=1}^{T} \log |J_t|, \tag{6}$$

where

$$S^* = (s^*_{gh}); s^*_{gh} = \frac{1}{T} \sum_{t=1}^{T} \frac{u_{gt}u_{ht}}{\sqrt{|u_{gt}|}\sqrt{|u_{ht}|}}, \qquad g, h = 1, \ldots, G, \tag{7}$$

and where J_t is the same as in (3). 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 can, of course, also be used, which merely changes the computation of s^*_{gh} in (7). One can also iterate, if desired, in the same manner as described above for the single-equation estimators. In this case, each iteration corresponds to the solution of one weighted FIML maximization problem.

The same algorithms that were used to maximize L in (2) can be used to maximize L^* in (6). The only change needed in the program that computes the objective function is to change the computation of s_{gh}. The advantage of using computational procedures that do not require the use of analytic derivatives is obvious in the present case, where it would be laborious to modify the analytic derivatives for each new weighting scheme tried. For the results in the next section 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 0.00001 being set equal to 0.00001. 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.

Any other estimators of simultaneous equations models that are based on minimizing a function of the residuals can likewise be modified to be robust estimators by weighting the residuals in different ways. One obvious way to modify the two-stage least squares estimator, for example, is to run the first-stage regressions in the usual way, replace in the usual way the actual values of the right-hand-side endogenous variables in the structural equation being estimated with the resulting fitted values, and then run a weighted-least-squares regression for the second stage. One could iterate, if desired, in the same way as described above. Again, the availability of optimization procedures that do not require analytic derivatives should greatly increase the number of modifications of a particular estimator that it is feasible to consider.

2. An Example

The model used for the results in this section is described in [4] and will not be discussed in any detail here. For present purposes, the monthly housing starts sector in the model has not been used, and housing starts have been taken to be exogenous. Imports were also taken to be exogenous. The period of estimation was 1960 II–1973 I, a total of 52 observations. Dummy variables were added to a few of the equations to adjust for the effects of two auto strikes.[7] Adjusting for strikes in this way is, of course, already a form of robust estimation in the sense that one has adjusted for large residuals that occur because of the strikes.

The model was estimated using six different estimators: ordinary least squares (OLS), FIML, WLS-I, WLS-II, WLS-III, and FIMLWLS-I. All but one of the equations were estimated under the assumption of first-order serial correlation of the error term. For each of the six estimators, 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 parameters. If, for example, the equation to be estimated is:

$$y_t = b_1 + b_2 x_t + b_3 y_{t-1} + u_t, \tag{8}$$

where

$$u_t = \rho u_{t-1} + \varepsilon_t, \tag{9}$$

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

$$y_t = \rho y_{t-1} + b_1(1 - \rho) + b_2(x_t - \rho x_{t-1}) + b_3(y_{t-1} - \rho y_{t-2}) + \varepsilon_t, \tag{10}$$

which is a standard nonlinear equation in the parameters. This is a convenient way of handling serial correlation in the present context, since the only complication it introduces is to make what might otherwise be a linear equation in parameters into a nonlinear one.

The model has the computational advantage that it decomposes into two blocks: a linear, simultaneous block and a nonlinear, recursive block. This means that J_t in (3) can be factored into two parts: one that is a function of some parameters but not of time and one that is a function of time but not of any parameters. Consequently, in the computation of the FIML and FIMLWLS-I estimates, the determinant of J only had to be computed once per evaluation of L or L^*, rather than the T times required for the more general case. In computing the FIML estimates, estimates were first obtained for the two blocks separately, using the ordinary least squares estimates as starting points, which required estimating 38 and 23 parameters, respectively. FIML estimates of all 61 parameters were then

[7] Aside from treating housing starts and imports as exogenous and adding a few dummy variables, two other small changes were made to the model in [4]. The price equation was taken to be linear with a length of lag of 20, and in equation (9.12) E_t was replaced by $M_t + MA_t + MCG_t$. See Table 11-4 in [4] for the original model. Dummy variables were not used for the work in [4], and strike observations were merely excluded from the sample period.

obtained, using the FIML estimates of the two blocks as starting points. FIMLWLS-I estimates were obtained in a similar manner. In contrast to the work in [5], no systematic attempt was made in this study to compare the various optimization algorithms, and so no comparisons of 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 and FIMLWLS-I computational problems here were not as well behaved and as robust to the use of different algorithms as were the optimal control problems in [5].

The six sets of estimates are available from the author on request. The two sets of FIML estimates tended to differ more from the other four sets of estimates than the other four sets of estimates differed from each other. There were no important cases of sign reversals among the different estimates of the same parameter.

The six different sets of estimates are compared in Table 1 in terms of within-sample prediction accuracy. Each set of estimates was used to generate static and dynamic predictions of the endogenous variables. Root mean square errors and mean absolute errors for five variables are presented in Table 1 for each set of estimates. The comparison here is similar to the comparison in Fair [6], where ten estimators were analyzed. 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 important, and for this reason all the estimators have been modified to account for serial correlation here.[8]

The five variables in Table 1 are GNP in current dollars (GNP_t), the private output deflator (PD_t), GNP in constant dollars ($GNPR_t$), private nonfarm employment (M_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 four variables presented in Table 1 from the recursive block are the four most important variables in the block. The estimates of the serial correlation parameters were used in the generation of all the predictions from the model.

The results in Table 1 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

[8] To be consistent with the notation in [6], "AUTO1" should be added to the name of each estimator in Table 1, but since all estimators considered in this section are "AUTO1" estimators, this will not be done.

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 turns out to be caused in large part by different FIML and FIMLWLS-I estimates of the constant term in the PD_t equation. The results for the other four estimators are 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 predictions 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, FIMLWLS-I, FIML, WLS-III, and OLS. 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 1.

1. WLS-I appears better than WLS-II and WLS-III, and all three appear better than OLS. 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 important.

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.

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 may not be too important (assuming the estimator accounts for first-order serial correlation).

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

TABLE 1

PREDICTION ERRORS FOR GNP_t, PD_t, $GNPR_t$, M_t, AND LF_{2t}
RMSE = Root Mean Square Errors MAE = Mean Absolute Errors

Estimator	RMSE Number of Quarters Ahead						MAE Number of Quarters Ahead					
	1 52 obs.	2 51 obs.	3 50 obs.	4 49 obs.	5 48 obs.	Entire Period 52 obs.	1 52 obs.	2 51 obs.	3 50 obs.	4 49 obs.	5 48 obs.	Entire Period 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
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
LF_{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

TABLE 2

OUTSIDE-SAMPLE PREDICTION ERRORS

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
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.
LF_{2t}	2276.	2230.	2109.	2067.

WLS-I results based on four iterations. Iterating more than once clearly improved the prediction accuracy of the estimator.

One final comparison was made here to see if the superiority of WLS-I over OLS also held 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 2, error measures for the simulation over the entire prediction period (17 observations) are presented for the same five variables presented in Table 1. For GNP_t, WLS-I outperforms OLS. Of the other four variables, which are determined in the recursive block, WLS-I is better for all but one (M_t). Overall, WLS-I appears to outperform OLS,[9] although the superiority of WLS-I here does not appear as pronounced as it was for the within-sample comparisons. This same conclusion also emerged from examining in more detail the predictions for the period 1969 I–1973 I (e.g., by the number of periods ahead predicted) and from examining predictions for the period 1970 III–1973 I based on estimates through 1970 II.

3. CONCLUSION

The purpose of this paper has been to discuss the computational aspects of robust estimates of econometric models and to present a few results of estimating a particular model. When robust estimators are considered as weighted-least-squares estimators, robust estimates can be obtained by a combination of solving unconstrained optimization problems and iterating. The unconstrained optimization problem for unweighted or weighted FIML estimates is likely to be by far the most expensive to solve for a given model, but even in this case it now appears feasible to estimate models of up to about 50 parameters. Certainly the computations involved in obtaining robust, single-equation estimates appear feasible for any model. Also, by using optimization algorithms that do not require derivatives or for which derivatives are obtained numerically, one greatly decreases the human effort involved in considering alternative estimators.

[9] 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.

The conclusions in Section 2 are clearly tentative. The comparisons among the estimators 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. Nevertheless, the robust estimators do predict well, and the results should at least provide encouragement to further work in this area.

Yale University and NBER Computer Research Center

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 Nonlinear Equations Systems," *The Review of Economics and Statistics* LV (February 1973), pp. 104–109.

[4] Fair, Ray C., *A Short-Run Forecasting Model of the United States Economy.* Lexington: D. C. Heath and Co., 1971.

[5] Fair, Ray C., "On the Solution of Optimal Control Problems as Maximization Problems," *Annals of Economic and Social Measurement* III (January 1974), pp. 135–154.

[6] Fair, Ray C., "A Comparison of Alternative Estimators of Macroeconomic Models," *International Economic Review* XIV (June 1973), pp. 261–277.

[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.