COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1005

ESTIMATES OF THE BIAS OF LAGGED DEPENDENT VARIABLE COEFFICIENT ESTIMATES IN MACROECONOMIC EQUATIONS

Ray C. Fair

February 1992

February 1992

ESTIMATES OF THE BIAS OF LAGGED DEPENDENT VARIABLE COEFFICIENT ESTIMATES IN MACROECONOMIC EQUATIONS

by

Ray C. Fair[1]

I. Introduction

It has been known since the work of Orcutt (1948) and Hurwicz (1950) that least squares estimates of lagged dependent variable (LDV) coefficients are biased. Macroeconometric model builders have generally ignored this problem, perhaps because they feel that the bias is likely to be small for the typical number of observations that are used. Hurwicz's estimates of the bias in an equation with only the LDV as an explanatory variable were small after about 100 observations. For example, for 100 observations the ratio of the expected value of the LDV coefficient estimate to the true value was .9804 (for small values of the coefficient). However, the results in Orcutt and Winokur (1969, Table IV) for 10, 20, and 40 observations show biases larger than those of Hurowitz for the case in which there is a constant term in the equation, suggesting that the bias in this case is also larger for, say, 100 observations. Furthermore, Andrews (1991) has recently shown that the bias is further increased when a time trend is added to the equation. For example, for 100 observations and a true coefficient of .8, the ratio of the median of the LDV coefficient estimate to the true value is .9388 in the equation with the constant term and time trend added.

---

[1]I am indebted to Don Andrews and Peter Phillips for helpful discussions, although I assume responsibility for all errors.

Typical macroeconomic equations are more complicated than the equations just discussed. They have more explanatory variables; some of the explanatory variables are likely to be endogenous; the error terms are sometimes serially correlated; and the equations may be nonlinear in both variables and coefficients. It is important to know how the size of the biases for these types of equations compare to those estimated for simpler equations. The first part of this paper provides estimates of the bias for various macroeconomic equations. A stochastic simulation procedure, based on Andrews' idea of computing median unbiased estimates, is used to estimate the bias.

The second part of the paper examines the sensitivity of the predictive accuracy and properties of a macroeconometric model to the use of median unbiased estimates rather than standard estimates. It attempts to gauge whether as a practical matter model builders should be concerned with the possible bias of LDV coefficient estimates.

II. The Procedure

The model considered in this paper can be dynamic, nonlinear, and simultaneous and can have autoregressive errors of any order. Write the model as:

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

where y_t is an n-dimensional vector of endogenous variables, x_t is a vector of predetermined variables (both exogenous and lagged endogenous), α_i is a vector of unknown coefficients, and u_{it} is an error term. It is assumed that the first m equations are stochastic, with the remaining u_{it} ($i = m+1,\ldots,n$) identically zero for all t.

Each equation in (1) is assumed to have been transformed to eliminate any autoregressive properties of its error term. If the error term in the untransformed version, say v_{it} in equation i, follows a rth order autoregressive process, $v_{it} = \rho_{1i}v_{it-1} + \ldots + \rho_{ri}v_{it-r} + u_{it}$, where u_{it} is iid, then equation i is assumed to have been transformed into one with u_{it} on the right hand side. The autoregressive coefficients $\rho_{1i}, \ldots, \rho_{ri}$ are incorporated into the α_i coefficient vector, and the additional lagged values that are involved in the transformation are incorporated into the x_t vector. This transformation makes the equation nonlinear in coefficients if it were not otherwise, but this adds no further complications to the model because it is already allowed to be nonlinear. It does result in the "loss" of the first r observations, but this has no effect on the asymptotic properties of the estimators. u_{it} in (1) can thus be assumed to be iid even though the original error term may follow an autoregressive process.

Let u_t be the m-dimensional vector $(u_{1t},\ldots,u_{mt})'$. It is assumed for the stochastic simulations below that u_t is distributed as multivariate normal $N(0,\Sigma)$, where Σ is m x m. Although the normality assumption is commonly made, the general procedure discussed in this paper does not depend on it. If another distributional assumption were used, this would simply change the way the error terms were drawn for the stochastic simulations.

Given estimates of α_i, denoted $\hat{\alpha}_i$, $(i = 1,\ldots,m)$, estimates of u_{it}, denoted $\hat{u}_{it}$, can be computed as $f_i(y_t,x_t,\hat{\alpha}_i)$. The covariance matrix Σ can then be estimated as $\hat{\Sigma} = (1/T)\hat{U}\hat{U}'$, where $\hat{U}$ is the m x T matrix of values of $\hat{u}_{it}$.

A vector of instruments Z_{it} is assumed to be available for the estimation of each equation i, where Z_{it} is correlated with the endogenous variables on the right hand side of equation i but uncorrelated with u_{it}. This allows the estimation of equation i by two stage least squares (2SLS), which under standard assumptions provides a consistent estimate of α_i.

The following procedure requires that one coefficient per stochastic equation be singled out for special treatment. The interest here, of course, is on the coefficient of the lagged dependent variable, but in future work other coefficients could be considered. Let α_{1i} denote the coefficient of interest in equation i.

The procedure for obtaining median unbiased estimates of the α_{1i} (i = 1,...,m) using the 2SLS estimator is as follows:

1. Estimate each equation i by 2SLS. Let $\hat{\alpha}_{1i}$ denote the 2SLS estimate of α_{1i}.
2. Guess the bias of $\hat{\alpha}_{1i}$, denoted b_{1i}. Add b_{1i} to $\hat{\alpha}_{1i}$ to obtain a first estimate of the true value of α_{1i}. Let α^*_{1i} denote this estimate: $\alpha^*_{1i} = \hat{\alpha}_{1i} + b_{1i}$. Constrain α_{1i} to be equal to α^*_{1i} and reestimate the other elements of α_i by 2SLS. Let α^*_i denote this estimate of α_i (i = 1,...,m). Use the estimated residuals from these constrained regressions to estimate the covariance matrix Σ, as discussed above. Let Σ^* denote this estimate of Σ.
3. Draw T values of the vector u^*_t, t = 1,...,T, from the distribution $N(0,\Sigma^*)$. Use these values and the values α^*_i (i = 1,...,m) to solve the model dynamically for t = 1,...,T. This is a dynamic simulation of the model over the entire estimation period using the drawn values of the error terms and the coefficient values α^*_i. The lagged endogenous variable values in x_t in (1) are updated in the solution process. After this solution, update Z_{it} to incorporate the new lagged endogenous variable values if lagged endogenous variable values are part of Z_{it}. Let Z^*_{it}, t = 1,...,T, denote this update.

Given the new data (i.e., the solution values of the endogenous and lagged endogenous variables), estimate each equation by 2SLS, and record the estimate of α_{1i} as $\alpha_{1i}^{(1)}$ ($i = 1,\ldots,m$). This is one repetition. Do a second repetition by drawing another T values of u_t^*, using these values and the values α_i^* to solve the model, using the new data to estimate each equation by 2SLS, and recording the estimate of α_{1i} as $\alpha_{1i}^{(2)}$ ($i = 1,\ldots,m$). Do this J times, and then find the median α_{1i}^m of the J values of $\alpha_{1i}^{(j)}$, $j = 1,\ldots,J$, ($i = 1,\ldots,m$).

4. If for each i α_{1i}^m is within a prescribed tolerance level of $\hat{\alpha}_{1i}$, go to step 6. If this condition is met, it means that for the particular coefficient values used to generate the data (the α_i^*'s), the median 2SLS estimates are within a prescribed tolerance level of the original estimates based on the historical data. If this condition is not met, take the new value of α_{1i}^* to be the previous value plus $\hat{\alpha}_{1i} - \alpha_{1i}^m$ for each i. Then constrain α_{1i} to be equal to this new value of α_{1i}^* and reestimate the other elements of α_i by 2SLS using the historical data. Let α_i^* denote this estimate of α_i ($i = 1,\ldots,m$). Again, use the estimated residuals from these constrained regressions to estimate the covariance matrix Σ. Let Σ^* denote this estimate of Σ. Now repeat step 3 for these new values.

5. Keep doing steps 3 and 4 until convergence is reached and one branches to step 6.

6. Take the median unbiased estimate of α_{1i} to be α_{1i}^*, and take the other coefficient estimates to be those in α_i^* ($i = 1,\ldots,m$). α_{1i}^* is the median unbiased estimate in that it is the value of α_{1i} that generates data that leads to the median 2SLS estimate equaling (within a prescribed tolerance level) the 2SLS estimate based on the historical

data. The estimated bias of $\hat{\alpha}_{1i}$ is $\hat{\alpha}_{1i} - \alpha^*_{1i}$.

Confidence intervals for α^m_{1i} can be computed from the final set of values of $\alpha^{(j)}_{1i}$, j = 1,...,J. For a 90 percent confidence interval, for example, 5 percent of the smallest values and 5 percent of the largest values would be excluded.

As noted above, this procedure does not require the normality assumption. Other distributions could be used to draw the u^*_t values. Also, the basic estimator need not be the 2SLS estimator. Other estimators could be used. The model in (1) can also consist of just one equation. In this case Σ is a scalar, the "solution" of the model simply consists of solving the particular equation (dynamically) over the sample period, and ordinary least squares may be the appropriate estimator to analyze.

The procedure does, however, have two limitations. First, as noted above, it focuses on just one coefficient per equation. No other coefficient estimate in an equation necessarily has the property that its median value in the final set of values is equal to the original estimate. The focus, of course, need not be on the coefficient of the LDV, but it must be on one particular coefficient per equation.

Second, there is no guarantee that the procedure will converge. Remember that overall convergence requires that convergence be reached for each equation, and achieving this much convergence could be a problem. For the results in this study, however, as will be seen, convergence was never a problem.

III. The Estimates

The procedure was used on the model in Fair (1984). The version of the model used here consists of 28 stochastic equations and 98 identities.[2] The basic estimation technique that is used for the model is 2SLS. For the present exercise, median unbiased estimates of the LDV coefficient were obtained for 13 of the 28 stochastic equations. The other 15 equations were estimated in the standard way by 2SLS. For all but 2 of the 13 equations the standard 2SLS estimate of the LDV coefficient is greater than .7. The variables that are explained by the 13 equations are presented in Table 1 along with the estimation results. The estimation period was 1954 I - 1991 II, for a total of 150 observations. In two of the equations the error term is first order autoregressive, and in one of the equations the error term is second order autoregressive.

Two sets of results are presented in Table 1. For the first set, each of the 13 equations was simply treated as an individual equation, with the basic estimation technique being ordinary least squares (OLS). Remember that

[2]The version of the model dated July 27, 1991, was used. The list of equations of the model is available from the author upon request. Two minor stochastic equations were dropped from the model for the current exercise because they have poor long-run dynamic properties. These are the equation explaining the interest payments of the firm sector (equation 19) and the equation explaining the interest payments of the federal government (equation 29). The stochastic simulation work requires that the model be solved dynamically over the entire sample period for each repetition, and the two interest payment equations were dropped to lessen the chances of the model not solving or of extreme solutions occurring on particular repetitions. Four other changes were also made to the model for the same reason. 1) The equation explaining unemployment benefits (equation 28) was taken to be in linear form rather than in log form. 2) The demand pressure variable in the price equation (equation 10) was taken to be in linear form rather than in log form. 3) The equation explaining housing investment (equation 4) was taken to have a first-order rather than a third-order autoregressive error. 4) The equation explaining production (equation 11) was taken to have a second-order rather than third-order autoregessive error. There were no solution errors in any of the computational work.

TABLE 1

Estimates of The Lagged Dependent Variable Coefficient
From Thirteen Macroeconomic Equations

Eq. in Model	OLS Est.	OLS Est. Bias	2SLS Est.	2SLS Est. Bias	Andrews Bias	No. of RHS vars.	ρ
1. CS	.784	-.041	.756	-.025	-.030	9	0
2. CN	.690	-.036	.724	-.056	-.029	7	0
3. CD	.874	-.028	.894	-.006	-.036	9	1
4. IHH	.720	-.045	.720	.024	-.029	7	1
11. Y	.361	-.007	.395	-.004	-.022	7	2
27. IM	.733	-.043	.728	-.039	-.030	14	0
5. L1	.458	-.030	.454	-.019	-.023	5	0
6. L2	.967	-.035	.960	-.019	-.043	6	0
7. L3	.821	-.033	.830	-.029	-.032	6	0
8. LM	.709	-.015	.725	-.022	-.029	4	0
30. RS	.894	-.008	.904	-.010	-.036	7	0
23. RB	.881	-.005	.882	-.005	-.035	5	0
24. RM	.837	-.009	.838	-.007	-.032	5	0
Average		-.026		-.017	-.031		

Notes:

Equations and data are discussed in Fair (1984). Version of the model dated July 27, 1991, used. See footnote 2.

Estimation period: 1954 I - 1991 II (150 observations).

Andrews Bias = Exact bias for an equation with a constant term, time trend, and lagged dependent variable and with the LDV coefficient equal to the 2SLS estimate.

ρ = order of the autoregressive process of the error term.

TABLE 1 (continued)

Notation:

CS	Consumption of services
CN	Consumption of nondurables
CD	Consumption of durables
IHH	Residential investment of the household sector
Y	Output of the firm sector
IM	Imports
L1	Labor force of males 25-54
L2	Labor force of females 25-54
L3	Labor force of all others 16 and over
LM	Number of people holding two jobs
RS	Three month Treasury bill rate
RB	AAA corporate bond rate
RM	Mortgage rate

when the above procedure is carried out for only one equation, the solution of the model is simply a dynamic solution of one equation. For these results the number of repetitions for each equation for each step 3 was 250. Convergence within a tolerance of .0001 was always achieved within five iterations (i.e., five steps 3 and 4). The OLS estimates are, of course, inconsistent if the equations have endogenous explanatory variables, and this first set of results is just meant for comparison purposes.

For the second set of results, the entire model was solved for each repetition. All 13 of the equations were estimated by 2SLS per repetition. The number of repetitions was 300 for each step 3. After 4 iterations (at which point the procedure was stopped), convergence within a tolerance of .0001 was achieved except for equation 4, where the difference between the median estimates on the third and fourth iterations was .0011. Convergence was thus not a problem for the present results.

The "Andrews bias" is also presented in Table 1 for comparison purposes. This is the exact bias for an equation with a constant term, time trend, and lagged dependent variable and with the LDV coefficient equal to the 2SLS coefficient estimated presented in Table 1.

Comparing the OLS bias with the Andrews bias, the average of the Andrews biases across the 13 equations is -.031, which compares to -.026 for the OLS bias. Six of the 13 OLS biases are greater in absolute value than the Andrews biases. These two sets of biases are thus fairly close, with the OLS biases being slightly smaller.

Turning now to the 2SLS biases, one on the biases is positive (for equation 4). The average of the 2SLS biases is -.017 if the positive bias is counted and -.020 if the positive bias is excluded. Only 2 of the 13 2SLS biases are greater in absolute value than the

Andrews biases. The results thus suggest that the 2SLS biases are generally smaller than the Andrews biases, although they are larger than would be expected from Hurwicz's original estimates.

Table 2 compares the 90 percent confidence values that are computed using the standard 2SLS technique with those that are computed using the 300 estimates from the last iteration. For the standard 2SLS estimator for a particular equation, the difference between the upper bound of the 90% confidence interval and the coefficient estimate is simply 1.645 times the 2SLS estimate of the standard error of the LDV coefficient estimate. Similarly, the difference between the lower bound of the confidence interval and the coefficient estimate is minus 1.645 times the 2SLS estimate of the standard error. For the 300 estimates, the difference between the lower bound of the confidence interval and the median estimate is the difference between the median estimate (α_1^m) and the estimate at which five percent of the estimates are below it. Similarly, the difference between the upper bound of the confidence interval and the median estimate is the difference (in absolute value) between the median estimate and the estimate at which five percent of the estimates are above it.

The results in Table 2 show that the median interval is not symmetric around the median estimate. In all but one of the 13 cases, the left tail of the distribution is thicker than the right tail. The average of the lower values across the 13 equations is -.076, and the average of the upper values is .058. The average of the lower values for the standard 2SLS estimates is -.078 (with the average of the upper values, of course, .078). The confidence intervals are thus on average slightly larger for the standard 2SLS estimates.

TABLE 2

90 Percent Confidence Values

Eq. in Model	2SLS		Median	
1. CS	-.083	.083	-.082	.073
2. CN	-.104	.104	-.115	.071
3. CS	-.074	.074	-.040	.044
4. IHH	-.189	.189	-.133	.097
11. Y	-.076	.076	-.076	.060
27. IM	-.077	.077	-.099	.068
5. L1	-.130	.130	-.122	.106
6. L2	-.044	.044	-.057	.028
7. L3	-.063	.063	-.059	.048
8. LM	-.067	.067	-.076	.062
30. RS	-.036	.036	-.045	.034
23. RB	-.030	.030	-.031	.027
24. RM	-.045	.045	-.051	.037
Average	-.078	.078	-.076	.058

Notes:

The first number for 2SLS is minus 1.645 times the 2SLS estimate of the standard error of the LDV coefficient estimate. The second number for 2SLS is the absolute value of the first number.

The first number for Median is the difference between the median estimate and the estimate at which five percent of the estimates are below it. The second number for Median is the difference (in absolute value) between the median estimate and the estimate at which five percent of the estimates are above it.

IV. Sensitivity of a Model to LDV Estimation Bias

It is hard to judge simply from the results in Table 1 whether the biases are quantitatively important in a practical sense. The biases in absolute value are generally less than a third of the 90 percent confidence range, which is some evidence that they may not be too important. This question is considered further in this section by examining the sensitivity of the accuracy and properties of the model in Fair (1984) to the estimated biases.

Consider first the predictive accuracy of the model. How is this accuracy affected by the use of the median unbiased estimates over the standard (biased) 2SLS estimates? Results that help answer this question are presented in Table 3. Root mean squared errors are presented for one-, four-, and eight-quarter-ahead predictions for the period 1954 I - 1991 II. The results labeled "2SLS" are for the regular estimates of the model, and the results labeled "median" are for the model in which the 2SLS estimates of the 13 equations listed in Table 1 are replaced by the median unbiased estimates. The other equations of the model are the same for both sets of estimates.

The results in Table 3 are easy to summarize: the RMSEs are very similar for the two estimates. There is clearly no discrimination possible between the two estimates regarding the fit of any variable in the model.

This result that the predictive accuracy of the model is little changed by the use unbiased over biased estimates is consistent with the properties of the simple equation with only the lagged dependent variable as an explanatory variable, say $y_t = \alpha y_{t1} + \epsilon_t$. Malinvaud (1970, p. 554) shows for this equation that the expected value of the prediction error is zero when the distribution if ϵ_t is symmetric even if the estimate of α that is used to

TABLE 3

Predictive Accuracy of Two Estimates of the Model
Prediction Period: 1954 I - 1991 II

Variable	Root Mean Squared Errors 1-q-ahead (150 obs.)	4-q-ahead (147 obs.)	8-q-ahead (143 obs.)
GNPR			
2SLS	0.69	1.50	1.95
Median	0.69	1.49	1.96
PF			
2SLS	0.41	0.70	0.88
Median	0.41	0.70	0.88
UR			
2SLS	0.28	0.74	0.99
Median	0.28	0.73	0.99
CS			
2SLS	0.42	0.67	0.88
Median	0.42	0.67	0.87
CN			
2SLS	0.68	1.33	1.68
Median	0.68	1.34	1.71
CD			
2SLS	2.86	5.17	7.11
Median	2.87	5.20	7.23
IHH			
2SLS	3.86	9.77	12.04
Median	3.86	9.79	12.08
IVF			
2SLS	3.18	3.78	3.95
Median	3.18	3.77	3.96
IM			
2SLS	2.95	4.69	5.18
Median	2.96	4.73	5.24
L1			
2SLS	0.18	0.18	0.19
Median	0.18	0.18	0.19

TABLE 3 (continued)

Root Mean Squared Errors

Variable	1-q-ahead (150 obs.)	4-q-ahead (147 obs.)	8-q-ahead (143 obs.)
L2			
2SLS	0.61	0.92	1.07
Median	0.61	0.92	1.07
L3			
2SLS	0.52	0.96	1.33
Median	0.52	0.94	1.31
LM			
2SLS	5.42	8.93	11.44
Median	5.43	8.90	11.42
RS			
2SLS	0.63	1.20	1.42
Median	0.63	1.20	1.41
RB			
2SLS	0.32	0.69	0.86
Median	0.32	0.68	0.85
RM			
2SLS	0.38	0.77	1.01
Median	0.38	0.76	1.00

Notes:

Errors are percentage errors in percentage points except for variables UR, RS, RB, and RM, whose errors are in units of the variables (which are percentage points), and for variable IVF, whose errors are in billions of 1982 dollars.

Prediction period: 1954 I - 1991 II. 150 observations for 1-q-ahead, 147 observations for 4-q-ahead, and 143 observations for 8-q-ahead.

Notation not in Table 1:

GNPR	Real GNP
PF	Private Nonfarm Price Deflator
UR	Unemployement rate
IVF	Inventory Investment of the Firm Sector

make the prediction is biased. The present results show that even for much more complicated models, prediction errors seem to be little affected by coefficient estimation bias.

Although the predictive accuracy of the two estimates of the model is essentially the same, it may be that the multiplier properties are different. To consider this question, the properties of the model to an export shock were examined using both sets of estimates. The results are presented in Table 4. The real value of exports was permanently increased by one percent of GNP beginning in 1988 I, and the model was solved for this change for both sets of estimates. Standard errors of the multipliers were also computed for the 2SLS estimates using the stochastic simulation method discussed in Fair (1980). The number of repetitions for these calculations was 500. The estimated standard errors are presented in parentheses in the table. Again, the results in Table 4 are easy to summarize: there is very little difference in the multiplier results. The differences are either zero or are quite small relative to the estimated standard errors. For all practical purposes, the two estimates of the model have the same properties. (To conserve space, only the results for three variables are presented in Table 4; the differences for all the other variables were also quite small.)

V. Conclusion

This paper has shown that it is possible to obtain median unbiased estimates of the LDV coefficient in macroeconomic equations using stochastic simulation. The estimated biases for 13 equations are on average somewhat smaller in absolute value than would be expected from Andrews' exact results for an equation with only a constant term, time trend, and LDV, although they are larger than would be expected from Hurwicz's original

TABLE 4

Properties of Two Estimates of the Model

Percentage change in the variable as a result of an exogenous increase in exports of one percent of real GNP

Variable	1-q-ahead	4-q-ahead	8-q-ahead	12-q-ahead
GNPR				
2SLS	0.81 (0.07)	1.33 (0.09)	0.74 (0.12)	0.55 (0.11)
Median	0.81	1.33	0.73	0.54
PF				
2SLS	0.00 (0.00)	0.18 (0.04)	0.35 (0.08)	0.37 (0.08)
Median	0.00	0.18	0.35	0.37
UR				
2SLS	-0.20 (0.03)	-0.58 (0.09)	-0.56 (0.13)	-0.49 (0.13)
Median	-0.20	-0.59	-0.54	-0.46

Notes: For UR the values are the change in the variable, not the percentage change.
The initial increase in exports was in 1988 I.
The numbers in parentheses are estimated standard errors of the changes for the 2SLS estimates. They were obtained using the method in Fair (1980).

estimates. In a practical sense the estimated biases are not very large because they have little effect on the overall predictive accuracy of the model and on its multiplier properties.

One might thus say in conclusion that macroeconometric model builders have not missed much by ignoring the Orcutt and Hurwicz warnings 40 years ago, although work with other models should be done to see if the results in this paper hold up. With hindsight, I guess I am not surprised by the present results. What they basically say is that if one changes a LDV coefficient estimate by about half of its estimated standard error and then reestimate the other coefficients in the equation to reflect this change, the fit and properties of the equation do not change very much. This is something that most model builders probably know from experience.

REFERENCES

Andrews, Donald W.K., "Exactly Unbiased Estimation of First Order Autoregresssive/Unit Root Models," Cowles Foundation Discussion Paper No. 975, April 1991.

Fair, Ray, C., "Estimating the Uncertainty of Policy Effects in Nonlinear Models," Econometrica, 1980, 1381-1391.

______, Specification, Estimation, and Analysis of Macroeconometric Models, Harvard University Press, 1984.

Hurwicz, Leonid, "Least-Squares Bias in Time Series," Chapter XV in Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph 10, T.C. Koopmans, ed., New York, John Wiley & Sons, Inc., 1950, 365-383.

Malinvaud, E., Statistical Methods of Econometrics, North-Holland Publishing Co., 1970.

Orcutt, Guy H., "A Study of the Autoregressive Nature of the Time Series Used for Tinbergen's Model of the Economic System of the United States, 1919-1932, Journal of the Royal Statistical Society, Series B, 1948, 1-45.

Orcutt, Guy H., and Herbert S. Winokur, Jr., "First Order Autoregression: Inference, Estimation, and Prediction," Econometrica, January 1969, 1-14.