COMPUTING MEDIAN UNBIASED ESTIMATES IN MACROECONOMETRIC MODELS

RAY C. FAIR

Cowles Foundation, Box 208281, Yale Station, New Haven, CT 06520-8281, USA
E-mail: fair@econ.yale.edu

SUMMARY

A stochastic simulation procedure is proposed in this paper for obtaining median unbiased (MU) estimates in macroeconometric models. MU estimates are computed for lagged dependent variable (LDV) coefficients in 18 equations of a macroeconometric model. The 2SLS bias for a coefficient, defined as the difference between the 2SLS estimate and the MU estimate, is on average smaller in absolute value than would be expected from Andrews' exact results for an equation with only a constant term, time trend, and LDV. The results also show that 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.

1. 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. This bias has been extensively examined for autoregressive equations, both with and without constant terms and time trends.[1] The purpose of this paper is to examine the bias of LDV coefficient estimates in typical equations in macroeconometric models. These equations are more complicated than the equations examined in the literature. At least some of the explanatory variables are usually endogenous; the error terms are sometimes serially correlated; and the equations may be non-linear in both variables and coefficients. Consistent estimation of these equations requires methods other than least squares, such as two-stage least squares (2SLS), three-stage least squares, and maximum likelihood. A stochastic simulation procedure is proposed in this paper for obtaining median unbiased (MU) estimates in macroeconometric models. From these estimates the 2SLS bias for a coefficient, defined as the difference between the 2SLS estimate and the MU estimate, can be computed.

2. A PROCEDURE FOR COMPUTING MU ESTIMATES IN MACROECONOMETRIC MODELS

The model considered here can be dynamic, non-linear, and simultaneous and can have autoregressive errors of any order. Write the model as

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

---

[1] See, for example, Orcutt and Winokur (1969), Grubb and Symons (1987), Stine and Shaman (1989), and Andrews (1993). Grubb and Symons (1987) also considered the case of exogenous variables other than the time trend in the equation.

where y_t is an n-dimensional vector of endogenous variables, x_t is a vector of predetermined variables (including lagged endogenous variables), α_i is a vector of unknown coefficients, and u_{it} is the error term for equation i for observation t. It will be assumed that the first m equations are stochastic, with the remaining u_{it} $(i = m + 1, ..., n)$ identically zero for all t. Each equation in (1) is assumed to have been transformed to eliminate any autoregressive properties of its error term, where the autoregressive parameters are incorporated into α_i.

Let u_t, be the m-dimensional vector $(u_{1t}, ..., u_{mt})$. It is assumed for the stochastic simulations below that u_t is distributed as multivariate normal $N(0, \Sigma)$, where Σ is $m \times m$. Although the normality assumption is commonly made, the procedures discussed in this paper do 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 $\hat{\alpha}_i$, consistent 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 $(1/T)\hat{U}\hat{U}'$, where $\hat{U}$ is the $m \times T$ matrix of the 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 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 is on the coefficient of the LDV, but other coefficients could be considered. Let α_{1i} denote the coefficient of interest in equation i.[2]

The procedure for obtaining MU estimates of the α_{1i} coefficients $(i = 1, ..., m)$ using the 2SLS estimator is as follows:[3]

(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 re-estimate 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. 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 equation (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 $\alpha_{1i}^{(1)}$ $(i = 1, ..., m)$. 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

---

[2] Rudebusch (1992, pp. 675–76) also presents a stochastic-simulation method for deriving 'median-unbiased' estimates (in the context of a single-equation, OLS framework). His method does not single out one coefficient per equation for special treatment, but, as discussed in footnote 4, it does so at a possible cost.

[3] This procedure is an extension of Andrews' (1993) method of computing *exact* MU estimates in an equation with a constant term, time trend, and LDV, which itself is based on a method discussed in Lehmann (1959). Andrews and Chen (1994) use essentially the present stochastic-simulation procedure in their estimation of pth-order autoregressive equations with time trends. Andrews and Chen's paper was written after the first version of this paper (Fair, 1992), and Andrews and Chen cite this version.

recording the estimate of α_{1i} as $\alpha_{1i}^{(2)}$ $(i = 1, ..., m)$. Do this J times, and then find the median α_{1i}^m of the J values of $\alpha_{1i}^{(j)}$ $(j = 1, ..., J)$, $(i = 1, ..., m)$.

(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 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 re-estimate the other elements of α_i by 2SLS using the historical data. Let α_i^* denote this estimate of α_i $(i = 1, ..., m)$. 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 MU estimate of α_{1i} to be α_{1i}^*, and take the other coefficient estimates to be those in $\alpha_i^*(i = 1, ..., m)$. α_{1i}^* is the MU estimate in that it is the value of α_{1i} that generates data that lead to the median 2SLS estimate equalling (within a prescribed tolerance level) the 2SLS estimate based on the historical data. The estimated bias of $\hat{\alpha}_{1i}$ is $\hat{\alpha}_{1i} - \hat{\alpha}_{1i}^*$.[4]

Confidence intervals for α_{1i}^m can be computed from the final set of values of $\alpha_{1i}^{(j)}(j = 1, ..., J)$. For a 90% confidence interval, for example, 5% of the smallest values and 5% 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 and the 'solution' of the model simply consists of solving the particular equation (dynamically) over the sample period. 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 reported in this paper, however, convergence was never a problem.

3. RESULTS FROM A MACROECONOMETRIC MODEL

The above procedure for obtaining MU estimates was used on the model in Fair (1994). The version of the model used here consists of 30 stochastic equations and 101 identities. The basic estimation technique that is used for the model is 2SLS, and these estimates were used as starting values.[5] Starting from these values, MU estimates of the LDV coefficients were obtained for 18 of the 30 stochastic equations. The estimates for the other 12 equations were fixed at their 2SLS values. The estimation period was 1954:1–1993:2, for a total of 158

---

[4] Rudebusch's method in the present context would be the following. In step (2) the bias of all the coefficients per equation would be guessed to form α_i^*. No re-estimation would be done to get the other elements of α_i^* once α_{1i}^* was chosen. In step (3) the median of all the coefficients per equation would be found. Convergence would be reached when for each coefficient per equation its median was within a prescribed tolerance level of its original estimate. The possible problem with this method is the following. Assume for sake of simplicity that there is only one equation, and consider the J regressions (repetitions) of the equation in step (3). It will not in general be the case that the medians for all the coefficients in the equation come from the same regression. For example, the value of coefficient 1 may come from regression 432, the value of coefficient 2 from regression 789, and so on, and this may not be desirable.

[5] Other estimates of the model aside from 2SLS estimates are also presented in Fair (1994). These include estimates obtained by three-stage least squares, full information maximum likelihood, and two-stage least absolute deviations. Some MU estimates are also presented, using the procedure discussed in this paper.

observations. The number of repetitions per iteration (i.e. the value of J in step (3) above) was 500. After three iterations (i.e. after steps (3) and (4) were done three times), the largest difference between the successive estimates of any LDV coefficient was less than 0·001 in absolute value. Convergence thus occurred very quickly.[6]

The bias for each of the 18 LDV coefficients is defined as the difference between its 2SLS estimate and its MU estimate. It is interesting to compare this bias to what will be called the 'AR bias', which is the exact bias for an equation with a constant term, time trend, and LDV and with the LDV coefficient equal to its 2SLS coefficient estimate. These biases are interpolated from Table III in Andrews (1993).

The results[7] show that the estimated biases are essentially zero for two of the 18 coefficients and negative for the rest. The average bias across the 18 estimates is −0·018. The average AR bias is −0·033, and so the results suggest that the bias of a typical macroeconometric equation is on average less than the bias of an equation that includes only a constant term, time trend, and LDV. For only four of the 18 coefficients is the AR bias smaller in absolute value.[8]

An interesting question is whether the biases just reported are quantitatively important regarding the properties of the model. For example, is the predictive accuracy of the model sensitive to the use of the MU estimates over the 2SLS estimates? To examine this question, root mean squared errors (RMSEs) were computed using the two sets of estimates. The prediction period was the same as the estimation period, namely 1954:1–1993:2, and one through eight-quarter-ahead RMSEs were computed. The results show that the RMSEs are very similar for the two sets of estimates. There is clearly no discrimination possible between the two sets on predictive accuracy grounds.

This result—that the predictive accuracy of the model is little changed by the use of unbiased over biased estimates—is consistent with the properties of a simple equation with only the LDV as an explanatory variable, say $y_t = \alpha y_{t-1} + \varepsilon_t$. Malinvaud (1970, p. 554) shows for this equation that the expected value of the prediction error is zero when the distribution of ε_t is symmetric even if the estimate of α that is used to 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 model is essentially the same using the MU and 2SLS estimates, it may be that the multiplier properties are different. To examine this question, multipliers were computed using the two sets of estimates. The multipliers are for a sustained increase in government spending of one percent of real GDP beginning in 1989:3. The results show very little difference, and for all practical purposes the two sets of estimates of the model have the same properties.

4. CONCLUSION

This paper has shown that it is possible to obtain MU estimates of coefficients in

---

[6] To lessen stochastic simulation error, the same draws of the error terms were used for each iteration. The number of errors drawn per iteration is 2,370,000 = (500 repetitions) × (30 stochastic equations) × (158 observations). The model is solved dynamically over the estimation period for each repetition, and each of the 18 equations is estimated for each repetition.

[7] Detailed results are available from the author upon request, including the confidence-interval, predictive-accuracy, and multiplier results reported below.

[8] Confidence intervals were also computed using the 500 repetitions from the last (third) iteration. These intervals were similar to the 2SLS intervals, although the left tail of the distribution of the MU estimates was slightly thicker than the right tail.

macroeconomic equations using stochastic simulation.[9] The estimated biases of the LDV coefficient estimates for 18 equations are, on average, smaller in absolute value than would be expected from Andrews' exact results for an equation with only a constant term, time trend, and LDV. 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.

ACKNOWLEDGEMENTS

REFERENCES

Andrews, D. W. K. (1993), 'Exactly median-unbiased estimation of first order autoregressive/unit root models', *Econometrica*, **61**, 139–65.

Andrews, D. W. K. and Hong–Yuan Chen (1994), 'Approximately median-unbiased estimation of autoregressive models', *Journal of Business & Economic Statistics*, **12**, 187–204.

Fair, R. C. (1992), 'Estimates of the bias of lagged dependent variable coefficient estimates in macroeconomic equations', Discussion Paper 1005, Yale University, Cowles Foundation.

Fair, R. C. (1994), *Testing Macroeconometric Models*, Harvard University Press, Cambridge, MA.

Grubb, D. and J. Symons (1987), 'Bias in regressions with a lagged dependent variable', *Econometric Theory*, **3**, 371–386.

Hurwicz, L. (1950), 'Least-squares bias in time series', in T. C. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph No. 10, John Wiley, New York, 365–383.

Lehmann, E. L. (1959), *Testing Statistical Hypotheses*, John Wiley, New York.

Malinvaud, E. (1970), *Statistical Methods of Econometrics*, North-Holland, Amsterdam.

Orcutt, G. H. (1948), '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, 1–45.

Orcutt, G. H. and H. S. Winokur, Jr. (1969), 'First order autoregression: inference, estimation and prediction', *Econometrica*, **37**, 1–14.

Rudebusch, G. D. (1992), 'Trends and random walks in macroeconomic time series: a re-examination', *International Economic Review*, **33**, 661–80.

Stine, R. A. and P. Shaman (1989), 'A fixed point characterization for bias of autoregressive estimators', *Annals of Statistics*, **17**, 1275–84.

---

[9] A very similar stochastic simulation procedure can be used to examine the asymptotic distribution accuracy of estimators of macroeconometric models by estimating 'exact' distributions and comparing these to the asymptotic distributions. The results, which are presented in Fair (1994), suggest that the two distributions are fairly close.