Center for Research on Economic and Social Theory
CREST Working Paper

Multiple Minima in the Estimation of Models with Autoregressive Disturbances

Howard Doran
Jan Kmenta

April, 1991
Number 91-2

DEPARTMENT OF ECONOMICS
University of Michigan
Ann Arbor, Michigan 48109-1220

MULTIPLE MINIMA IN THE ESTIMATION OF MODELS WITH AUTOREGRESSIVE DISTURBANCES

Howard Doran and Jan Kmenta[1]

Abstract

In this paper we show that the problem of multiple minima obtained by using the search procedure in the context of the Cochrane-Orcutt transformation disappears when the observation set is extended to include the first observation, as proposed by Prais-Winsten.

1. Introduction

We consider, without a loss of generality, the following simple regression model with autoregressive disturbances:

$$Y_t = \alpha + \beta X_t + \epsilon_t, \qquad t = 1,2,\ldots,n,$$

$$\epsilon_t = \rho\epsilon_{t-1} + u_t, \qquad |\rho| < 1,$$

where all the usual definitions and assumptions apply. We also assume that u_t is normally distributed.

To remove the autoregressive ϵ_t, one can apply the following transformation:

$$Y_t^* = \alpha W_t^* + \beta X_t^* + u_t$$

where, for t = 1,

$$Y_t^* = Y_t\sqrt{1-\rho^2}, \quad W_t^* = \sqrt{1-\rho^2}, \quad X_t^* = X_t\sqrt{1-\rho^2},$$

and, for t = 2,3,...,n,

$$Y_t^* = Y_t - \rho Y_{t-1}, \quad W_t^* = 1-\rho, \quad X_t^* = X_t - \rho X_{t-1}.$$

When the first observation, (Y_1^*, W_1^*, X_1^*), is dropped, the transformation is called Cochrane-Orcutt (C-O); when it is included, the transformation is known as Prais-Winsten (P-W).

[1] University of New England and University of Michigan, respectively. Kmenta's contribution was supported by the Alexander-von-Humboldt Foundation.

The transformed equation is usually estimated in one of two ways.

(a) Iterative procedure

Starting with the least squares estimates of the untransformed equation, the residuals are used to obtain an initial estimate of ρ. This estimate is used to transform the original equation and to obtain the second-stage estimates of α and β, and so on. The procedure is repeated until convergence.

From Huzurbazar (1948) and Oberhofer and Kmenta (1974) it follows that this procedure converges and the resulting estimator is unique and consistent regardless of whether the C-O or P-W transformations are used.

(b) Search procedure

Suggested originally by Hildreth and Lu (1960), the sum-of-squared-errors (SSE) is computed as a function of ρ and the chosen estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\rho}$ are those that correspond to minimum SSE. This minimum is located by searching over ρ in the range $|\rho|<1$. It is with this method that the phenomenon of multiple minima has been documented, always in the context of the C-O transformation.

2. Multiple minima

The first to raise the question of multiple minima were Hildreth and Lu (1960), who provided an artificial, five-observation example of the existence of double minima of SSE. Another example, involving a more realistic model and data, was provided by Dufour *et al.* (1980). The issue was also more extensively treated by Oxley and Roberts (1986) who used a lagged dependent variable model. (It should be pointed out, though, that in this case the iterative C-O estimator is inconsistent since the starting least squares estimator is inconsistent unless $\rho = 0$). A rigorous treatment of the problem of multiple minima in the context of a lagged dependent variable model can be found in Betancourt and Kelejian (1981).

In the example of Hildreth and Lu (1960) the authors found dual minima of SSE at $\hat{\rho} = -0.9$ and $\hat{\rho} = 0.3$, while Dufour *et al.* (1980) reported minima at $\hat{\rho} = 0.3289$ and $\hat{\rho} = 0.9318$. We have recomputed both sets of estimates using double precision and confirmed these dual minima. Thus the existence of multiple minima in small samples cannot be ruled out when the C-O transformation is used.

In this paper we examine the possibility of the existence of multiple minima when using the search procedure with the P-W transformation. To this end we reestimated the parameters of the models of Hildreth and Lu (1960) and Dufour *et al.* (1980), using the authors' respective data sets but including the first observation (Y_1^*, W_1^*, and X_1^*). The results turned out to be rather startling: in both cases the dual minima of SSE completely disappear. The unique minimum in the Hildreth and Lu case occurs at $\hat{\rho} = -0.99$, and in the Dufour *et al.* case at $\hat{\rho} = 0.3$. (The latter is shown in Figure 1.) When using the full maximum likelihood procedure that allows for the appropriate Jacobian, the results turned out to be similar. In the case of Hildreth and Lu, the likelihood function peaked at $\hat{\rho} = -0.78$ and in the case of Dufour *et al.* at $\hat{\rho} = 0.315$. These unique minima correspond to the estimates of ρ obtained by the iterative procedure. Since the importance of the first observation diminishes as the sample size increases, our results are consistent with the claim that the occurrence of multiple minimum of the SSE curve (or multiple maxima of the likelihood function) will asymptotically disappear.

3. An explanation

During our analysis of both the Hildreth and Lu (1960) and the Dufour *et al.* (1980) data sets, two features emerged. First, as emphasized above, when the transformed first observation was included, the dual minima problem disappeared. Second, when the intercept α was omitted from the model, the same thing happened even when the first observation was omitted. As dropping the first transformed observation converts the variable W_t^* into a column of constants, there is the strong suggestion that the occurrence of dual minima is associated with the presence of a constant term in the transformed model.

To investigate the cause of the occurrence of the second, 'spurious' minimum — in addition to the 'genuine' minimum that corresponds to the maximum likelihood estimate of ρ — let us consider the error sum of squares SSE* as a function of ρ. This can be written as

$$SSE^* = SST^* (1 - R^{*2}),$$

where SST* is the total sum of squares of Y*, and SSE* and R^{*2} refer to the regression of Y* on W* and X*. We focus our attention first on the C-O transformation, thus restricting ourselves to the observations $t = 2,3,\ldots,n$ and, by implication, including an intercept in the transformed model since W* is constant. Note that SSE*, SST*, and R^{*2} are all based on mean-corrected values of X* and Y*.

Let us now suppose that the values of the untransformed dependent variable Y_t in a particular sample are such that they can be adequately described by the least squares regression

$$Y_t = c + dY_{t-1} + v_t \tag{1}$$

where, by construction, $\Sigma v_t = \Sigma v_t Y_{t-1} = 0$. By 'adequately described' we mean that equation (1) gives a good fit, that Y_{t-1} accounts for the systematic movements in Y_t in the sample, and that there is no systematic variation left in the sample values of v_t. It should be emphasized that equation (1) does not represent a general statement about the process of generating Y_t, which is assumed to be given by the regression model specified at the outset. With respect to equation (1), we are interested in cases where $|d| < 1$.

Now, from the definition of Y* given as

$$Y_t^* = Y_t - \rho Y_{t-1}$$

it follows that

$$Y_t^* = c + (d - \rho)Y_{t-1} + v_t \tag{2}$$

and hence

$$SST^* = (d - \rho)^2 \Sigma y_{t-1}^2 + \Sigma v_t^2 \tag{3}$$

where y_{t-1} is the mean-corrected value of Y_{t-1}. Thus SST* is a quadratic function of ρ, having a minimum at $\rho = d$. Furthermore, the magnitude of the percentage change in SST* for a small change in ρ is determined by the size of Σv_t^2 relative to $d^2 \Sigma y_{t-1}^2$. If Σv_t^2 is relatively small (that is, the R^2 from regression (1) is high), then a small change in ρ will yield a relatively large percentage change in SST*, and vice versa.

Let us consider now the $(1 - R^{*2})$ part of SSE* as a function of ρ. In the neighborhood of $\rho = d$ we have, by (2), that

$$Y_t^* \simeq c + v_t$$

and the C-O regression of Y_t^* on W_t^* (a constant vector) and X_t^* must yield small R^{*2} regardless of the values of X. In fact, for the Dufour *et al.* (1980) data and the C-0 transformation, the values of R^{*2} corresponding to different values of ρ are:

ρ:	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
R^{*2}:	0.76	0.73	0.68	0.63	0.55	0.44	0.33	0.21	0.14	0.16	0.17

which indicates a minimum for values of ρ between 0.8 and 0.9. When $1 - R^{*2}$ is close to 1 and insensitive (in percentage terms) to a small change in ρ, then the changes in SSE*, given by the product of SST* and $(1 - R^{*2})$, is dominated by those of SST* and hence, in the neighborhood of $\rho = d$, the product will have quadratic characteristics.

To summarize, in the cases where the descriptive least-squares regression (1) yields $|d| < 1$ and the fit is reasonably good, we can expect to observe a second minimum when the C-O transformation is used, due to the dominating influence of the quadratic SST* and the minimal influence of $(1 - R^{*2})$.

When the least-squares regression (1) was applied to the Dufour *et al.* (1980) data, we obtained $d = 0.896$ and $R^2 = 0.74$. Thus the observed second minimum of SSE* at about

$\rho = 0.90$ (see Figure 1) is to be expected. Note also that R^{*2} is at minimum very near the point where $\rho = d$.

The obvious question remaining is why, when the P-W transformation is used (implying no intercept in the transformed model) or when the intercept α is dropped from the original model, a second minimum seems not to occur.

When there is no intercept in the transformed model, the preceding analysis can again be followed except that quantities are not mean-corrected and the relevant least squares regression corresponding to (1) is now

$$Y_t = d_1 Y_{t-1} + V_t \qquad t = 2,3,\ldots,n. \tag{4}$$

As d_1 is simply a weighted average of the ratios Y_t/Y_{t-1}, then if in a particular sample the series Y_t is trending upwards, we can expect $d_1 > 1$. (Note that — unlike in the case of equation (1) — it is not necessary that equation (4) gives a good fit, the only relevant point is that $d_1 > 1$.) The P-W transformation can now be written as

$$Y_t^* = \sqrt{1-\rho^2}\, Y_t \qquad t=1$$

$$= (d_1 - \rho)Y_{t-1} + V_t \qquad t = 2,3,\ldots,n$$

and so

$$SST^* = (1-\rho^2)Y_1^2 + (d_1-\rho)^2 \Sigma Y_{t-1}^2 + \Sigma V_t^2$$

which is quadratic in ρ having a minimum at

$$\rho = Kd_1, \tag{5}$$

where

$$K = \Sigma_2^n Y_{t-1}^2 / \Sigma_3^n Y_{t-1}^2 \tag{6}$$

As $K \geq 1$, it follows that

$$\rho \geq d_1 > 1. \tag{7}$$

Thus SST^*, while still quadratic in ρ, now has a minimum outside the relevant range $|\rho| < 1$. It follows that the spurious minimum which can occur when C-O is used, will not occur when P-W is used, as long as the dependent variable Y_t is trending upward, which is commonly the case with economic data.

Further examination of the Dufour *et al.* data shows that the dependent variable is trending upward, and that regression (4) yields $d_1 = 1.03$ which is outside the relevant range, as expected.

4. Summary and conclusion

We believe that the reason for the occurrence of multiple minima when the C-O search procedure is used lies in the nature of the observations on the dependent variable. Whenever a least squares regression of the form

$$Y_t = c + dY_{t-1} + v_t$$

provides an adequate description of the sample data, a second minimum is likely to occur near $\rho = d$.

When the P-W transformation is used, a second minimum may still occur, but when the dependent variable is trending upward, the second minimum will be located outside the relevant region $|\rho| < 1$, removing any ambiguity in the search results. Thus our results supersede the recommendation of Dufour *et al.* (1980, p.46) "to combine a search routine ... with the Cochrane-Orcutt procedure" by the recommendation always to replace the C-O transformation by the P-W transformation that requires the inclusion of the transformed first observation in the observation set.

REFERENCES

Betancourt, Roger, and Harry Kelejian, "Lagged Endogenous Variables and the Cochrane-Orcutt Procedure," Econometrica 49 (July 1981), 1073–1078.

Box, George E.P., and Gwilym M. Jenkins, Time Series Analysis (Holden-Day, San Francisco, 1976).

Dufour, Jean-Marie, Marc J.I. Gaudry, and Tran Cong Liem, "The Cochrane-Orcutt Procedure: Numerical Examples of Multiple Admissible Minima," Economics Letters 6 (1980), 43–48.

Hildreth, Clifford, and John Y. Lu, "Demand Relations with Autocorrelated Disturbances," Technical Bulletin 276, Department of agricultural economics (Michigan State University, East Lansing, MI, 1960).

Huzurbazar, V.S., "The Likelihood Equation, Consistency and the Maxima of the Likelihood Function," Annals of Eugenics 14 (1948), 185–200.

Oberhofer, Walter and Jan Kmenta, "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models," Econometrica (1974), 579–590.

Oxley, Leslie T., and Colin J. Roberts, "Multiple Minima and the Cochrane-Orcutt Technique: Some Initial Monte-Carlo Results," Economics Letters (1986), 247–250.

Recent Crest Working Papers

91-01: Mark Bagnoli, Stephen W. Salant, Joseph E. Swierzbinski, "Price Discrimination and Intertemporal Self-Selection," June, 1990.

91-02: Howard Doran, Jan Kmenta, "Multiple Minima in the Estimation of Models with Autoregressive Disturbances," April 1991.

91-03: Ted Bergstrom, Mark Bagnoli, "Courtship as a Waiting Game," April 3, 1991.

91-04: Ted Bergstrom, Jeffrey K. MacKie-Mason, "Peak-Load Pricing-With and Without Constrained Rate of Return," April 2, 1991.

91-05: Laura E. Kodres, Daniel P. O'Brien, "The Existence of Pareto Superior Price Limits and Trading Halts," February 1991.

91-06: Theodore Bergstrom, David Lam, "The Effects of Cohort Size on Marriage Markets in Twentieth Century Sweden," November 1989.

91-07: Theodore Bergstrom, David Lam, "The Two-Sex Problem and the Marriage Squeeze in an Equilibrium Model of Marriage Markets," April 1991.

91-08: Greg Shaffer, "Capturing Strategic Rent: Full-line Forcing, Maximum Resale Price Maintenance, Brand Discounts and Aggregate Rebates," March 1991.

91-09: Jeffrey K. MacKie-Mason, Roger H. Gordon, " Taxes and the Choice of Organizational Form," April 3, 1991.

Doran, Howard and
Kmenta, Jan
"Multiple minima in
the estimation of
models with
autoregressive
disturbances"

DATE	ISSUED TO