A COMMENT ON THE USE OF ORDINARY LEAST SQUARES IN ESTIMATING A STRUCTURAL EQUATION

Halûk Erlat*

Using Ordinary Least Squares (OLS) to estimate the coefficients of a structural equation will, in general, give inconsistent results due to the fact that there exists current endogeneous variables among the explanatory variables which are correlated with the equation's disturbance term (see e.g. Dhrymes (1970: 174-176). However, there are certain situations where OLS estimators are consistent. A recursive model with a diagonal disturbance variance-covariance matrix is a case in point (see (Fisher (1966: 92-126)) An examination of such models will show that the use of OLS estimators is closely linked with the **identifiability restriction** of non-correlation between the disturbances of different structural equations.

In this note we shall take up a different model and argue that this intimate link between identifiability and the use of OLS estimation is also valid for this case. The reason for taking up this point is the discussion of such a model as given in Rao and Miller (1971: 185-199) which is apt to be misleading.

Our case is the simple supply and demand model. Let Q be the equilibrium quantity demanded and supplied, and P be the equilibrium price which clears the market. Assuming the functions to be linear in both variables and coefficients, we have

(1) $Q_t = a + bP_t + u_{t1}$ (demand)

(2) $Q_t = c + dP_t + u_{t2}$ (supply)

where the u_{ti} indicate the relevant residuals. We assume that

(3) $E(u_{t1}) = E(u_{s2}) = 0$

$Cov(u_{t1}\ u_{s2}) = 0$ for $t \neq s$

$= s_{ij}$ for $t = s$, $i, j = 1, 2$

The reduced forms of (1) and (2) may be written as

* Assistant Professor of Economics, Middle East Technical University, Ankara.

(4) $P_t = (c-a) / (b-d) + (u_{t2}-u_{t1}) / (b-d)$

$Q_t = (bc-ad) / (b-d) + (bu_{t2}-du_{t1}) / (b-d)$

Given only this prior information on the model, application of OLS will yield estimates of either (1) or (2) and there would be no way of knowing which one is being estimated. Assuming that we have estimated the slope of the demand equation, b, we would have

(5) $\hat{b} = \sum_t p_t q_t / \sum_t p_t^2 = b + \sum_t p_t v_{t1} / \sum_t p_t^2$

where $p_t = P_t - \bar{P}$, $q_t = Q_t - \bar{Q}$ and $v_{t1} = u_{t1} - \bar{u}_1$. To find the inconsistency involved we evaluate,

(6) plim $(\hat{b} - b)$ = plim $T^{-1} \sum_t p_t v_{t1}$ / plim $T^{-1} \sum_t p_t^2$

But using the first expression in (3) we have,

(7) plim $T^{-1} \sum_t p_t v_{t1} = (b-d)^{-1}$ (plim $T^{-1} \sum_t v_{t2} v_{t1}$ — plim $T^{-1} \sum_t v_{t1}^2$)

$= (b-d)^{-1} (s_{12} - s_{11})$

and

(8) plim $T^{-1} \sum_t p_t^2 = (b-d)^{-2}$ (plim $T^{-1} \sum_t (v_{t2}^2 - 2v_{t1}v_{t2} + v_{t1}^2)$

$= (b-d)^{-2} (s_{22} - 2s_{12} + s_{11})$

where $v_{t2} = \bar{u}_2$. Hence.

(9) plim $(\hat{b} - b) = (b-d) (s_{12} - s_{11}) / (s_{22} - 2s_{12} + s_{11})$

A similar result may be obtained for the OLS estimator of d;

(10) plim $(\hat{d} - d) = (d-b) (s_{22} - s_{12}) / (s_{11} - 2 s_{12} + s_{22})$

Thus, in both cases there is no way of knowing which equation is being estimated, i.e., there is an identification problem, and OLS applied to this unidentified equation yields inconsistent results. However, if we now impose the following restrictions on the s_{ij}'s; namely, (1) $s_{12} = 0$, and (2) either one of the s_{ii}'s is very close to zero, then one of the equations becomes identified with respect to the other because the first restriction implies that the two functions do not vary together while the second restriction states that one of the functions varies more than the other, thereby approximately tracing the one with the smaller variance. But we immediately note, from (9) and (10), that these restricitions reduce the inconsistency of

OLS estimators of the equation for which the identification restrictions have been specified. In other words, if in the model specified by (1) and (2), either one of the equations is «nearly» identified vis-a-vis the other one, then OLS estimation of the identified equation will be «opproximately» consistent.

In the light of this discussion we may take up Rao and Miller's argument. First of all they interpret the inconsistency expressions as given by (9) and (10) as **biases** and, with reference to a result in Sawa (1969), they suggest that this bias may approximate the finite sample bias of the OLS estimators. Subsequently, they argue that if $s_{12} = 0$ and if the relevant s_{ii} is small the bias will diminish. Depending upon which s_{ii} is small one would then be estimating either the demand or the supply equation. This fact, though, they argue, is independent of whether the equation in question is identified and may exist even if the equation has previously been identified using, say, the exclusion restrictions.

Taking up these arguments one by one, it is first important to establish whether or not a particular estimator has finite first order moments for the model under consideration. That all moments upto the degree T-1 exist for such models is, in fact, established by Sawa (1969: 931). However, there appears to be no reason to interpret the expressions in (9) and (10) as biases since they have not been derived from the finite sample distribution of these estimators Even if this were the case, it does not follow from the result from Sawa that the **mean** of the OLS estimator does not change as T tends to infinity; what that result indicates is that the **mode** of the finite sample distribution of OLS does not change and it appears that, although the distribution of OLS is symmetric in most cases, there are situations of asymmetry, especially if the somple size is very small. Thus, **even if** (9) and (10) are regarded as bias expressions, it would be rather forcing the issue to use them as approximate expressions for any sample size.

As for their argument that the smallness in the s_{ii} only reduce the bias (or, more aptly, the inconsistency) in the OLS estimators independent of identification considerations, our discussion above shows that this is not true at all. In fact, one may approach the problem in a different manner by noting that an identification problem in our supply-demand model arises because observational information alone cannot distinguish the true demand and supply curves from another pair constructed by taking two independent linear

combinations of the true equations. In our model given by (1) and (2), taking A times the supply equation and adding the result to the demand equation gives

(11) $Q_t = (1 + A)^{-1} ((Ac + a) + (Ad + b) P_t + Au_{t2} + u_{t1})$

while multiplying the demand equation by B and adding to the supply equation gives

(12) $Q_t = (1 + B)^{-1} ((Ac + a) + (Ad + b) P_t + Bu_{t1} + u_{t2})$

Now, one of our apriori restrictions is that

(13) $s_{12} = \text{plim } T^{-1} \sum_t u_{t1} u_{t2} = 0$

Let $w_{t1} = (Au_{t2} + u_{t1}) / (1 + A)$ and $w_{t2} = (Bu_{t1} + u_{t2}) / (1 + B)$. If we are to be able to distinguish the false equations from the true ones by using (13), these disturbances must also be uncorrelated with each other in the probability limit. Thus, we must have

(14) $\text{plim } T^{-1} \sum_t w_{t1} w_{t2} = 0$

or

(15) $\text{plim } T^{-1} \sum_t ((1 + B)(1 + A))^{-1} (Au_{t2}^2 + Bu_{t1}^2 + (1 + AB) u_{t1} u_{t2})$

$= ((1 + B) (1 + A))^{-1} (As_{22} + {}_{11})$

$= As_{22} + Bs_{11} = 0$

so that

(16) $A = -B (s_{11} / s_{22})$ and $B = -A (s_{22} / s_{11})$

The only admissible value of A, in the first equality of (16), which will make the equation obtained indistinguishable from the true supply equation is 0. A will approach this values as s_{11} approaches zero, and B will attain it as s_{22} approaches zero. In either case **one** of the equations could be «nearly» identified[1] and OLS would be, in this case, «approximately» consistent.

Finally, suppose the supply equation is identified by including Y, income, into the demand equation as an exogeneous explanatory variable, i.e.,

---

1. See Fisher (1966: 85-91) for a rigorous treatment of this point.

(17) $Q_t = a + bP_t + gY_t + u_{t1}$

It may easily be shown that in this case, with (13) holding true,

(18) $\text{plim } (\hat{b} - b) = (b\text{-}d)\ (s_{11} / (s_{11} + s_{22}))$

and

(19) $\text{plim } (\hat{d} - d) = (d\text{-}b)\ (s_{22} / (s_{11} + s_{22} + f^2 m_y))$

where

(20) $m_y = \text{plim } T^{-1} \sum_t y_t^2 < \infty$, $y_t = Y_t - \bar{Y}$, $f = g/(d\text{-}b)$

Thus, from (18) we conclude that, whereas the supply function is identified because the demand function includes an exogeneous variable which shifts the demand function relative to the supply function, the demand function may be «nearly» identified and the OLS estimator would be «approximately» consistent if s_{11} is «small».

On the other hand, from (19) we note that OLS would also be «approximately» consistent for this **identified** equation if s_{22} was small relative to s_{11}, which would imply that OLS would be utilised at the cost of leaving the demand equation unidentified, and therefore, inestimable. But other consistent estimators of d exist such as Two-Stage Least Squares (2SLS) and they may be utilised, the restriction of a small s_{11} to be used to help identify and, consequently, consistently estimate b.

Summing up, the fact that the variances s_{ii} contribute to the inconsistency of more elaborate models in which identification is provided by other means, does not imply that the problem is the same in both contexts, but that given the **appropriate identification restrictions** for the first demand and supply model considered here (equations (1) and (2)) OLS is the **appropriate estimator** while for the model specified by (2) and (17) the appropriate identification restrictions imply that OLS is **not** the appropriate estimator for the supply equation while **it is** for the demand equation.[2]

2. The discussion so far should not be taken to mean that we are belittling the importance of small sample bias in OLS and 2SLS estimators of a structural equation. In fact, both estimators are highly biased for small samples, But the important thing to note here is that this bias in OLS is very much depenDen on the difference between, say, b and the ratio, say h, of the covariance of the reduced form disturbances to the variance of the disturbance of the reduced form equation of the dependent variable. As b-h increases the bias of OLS becomes more noticeable than that of 2SLS (Sawa (1969)). Thus, in practical situations it may be dangerous to estimate a structural equation using OLS since we have no information about the covariance structure of the endogeneous variables. A more fruitful approach to reducing the bias in small samples, compared to that of Rao and Miller, may be to combine OLS estimates in the manner suggested by Sawa (1973).

REFERENCES

Dhyrmes, P.J., **Econometrics: Statistical Foundations and Applications.** New York: Harper and Row, 1970.

Fisher, F.M., **The Identification Problem in Econometrics.** New York: McGraw-Hill, 1966.

Rao, P. and R.L. Miller: **Applied Econometrics.** Belmont, California: Wadsworth, 1971.

Sawa, T., «The Exact Sampling Distribution of Ordinary Least Squares and Two - Stage Least Squares Estimators», **Journal of the American Statistical Association,** vol. 64, 1969, pp. 923-937.

Sawa T., «Almost Unibaised Estimators in Simultaneous Equations System», **International Economic Review,** vol. 14, 1973, pp. 97-106.