WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

INSTRUMENT VARIABLE ESTIMATION OF MISSPECIFIED MODELS

by

Julio J. Rotemberg

WP# 1508-83 December 1983

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

INSTRUMENT VARIABLE ESTIMATION OF
MISSPECIFIED MODELS

by

Julio J. Rotemberg

WP# 1508-83 December 1983

Instrumental Variable Estimation of
Misspecified Models

Julio J. Rotemberg*
October 1983
(Revised November 1983)

*Sloan School of Management, MIT. I wish to thank Lars Hansen, Jerry Hausman, James Poterba, James Powell and Thomas Stoker for helpful discussions while retaining sole responsibility for any remaining errors.

Abstract

This paper studies the estimation of models in which the set of instruments is not, in fact, orthogonal to the residuals. I first show that, in overidentified models of this type, one can generally obtain arbitrary estimates by varying the weights given to different instruments. I then weaken the assumptions of instrumental variable estimation by allowing for nondegenerate price distributions over the product of instruments and residuals. If the variance covariance matrix of this distribution is diagonal, the estimates which minimize the impact of misspecification are shown to lie inside the polyhedron of estimates from the exactly identified submodels.

Introduction

Consider the single equation model:

$$Y = X\beta + \varepsilon \tag{1}$$

where Y is a T x 1 vector, X a T x k matrix, ß a k x 1 vector of parameters of interest and ε at T x 1 vector of disturbances. Often economic reasoning predicts that ε_t is uncorrelated with a series of variables Z_{it} (which may include X's). It is then natural to estimate the vector ß by the method of instrumental variables proposed by Reiersol (1945), discussed in detail in Sargan (1958) and generalized by Hansen (1982). This method considers the sample inner products of the instruments and residuals $Z_i'(Y-X\beta)$ where Z_i is the vector of observations on instrument i. It then sets k linear combinations of these products equal to zero so that

$$WZ'(Y-X\beta) = 0 \tag{2}$$

where Z is a T x m matrix of instruments, $m \geq k$ and W is a k x m weighting matrix of rank k.

The hypotheses that the expected value of $Z_{it}\varepsilon_t$ is exactly zero is probably false for most economic models. This explains in part why, in empirical papers this hypothesis is often rejected by Hausman (1978) tests and other specifications tests. In particular, such rejections are reported by: Hansen and Singleton (1983) Mankiw, Rotemberg and Summers (1982), Pindyck and Rotemberg (1983).[1] After all, the models are only an approximation to reality. The lack of concern expressed over these rejections must mean that the authors imagine on *a priori* grounds that the inconsistency of the

resulting estimates must be small. This belief may be based on Fischer's (1961) "proximity theorem," which states that for a fixed W as the mean of $\varepsilon_t Z_{it}$ goes to zero the inconsistency of ß disappears in a continuous fashion. This paper argues that this optimism may be unfounded. I show that when overidentified models (i.e. models where $m \geq k$) are misspecified even slightly, the estimated ß's may be extremely far from the true ß's. This result does not contradict Fischer's result directly. This is so because I keep the mean $\varepsilon_t Z_{it}$ fixed and I consider changes in the weighting matrix W. If the means of $\varepsilon_t Z_{it}$ differ sufficiently across instruments, one can obtain essentially arbitrary $\hat{\beta}$'s by varying W.

Methods have been proposed for selecting weighting matrices that minimize the asymptotic covariance matrix of the ß's under the assumption that the model is correctly specified. In particular if the ε_t''s are i.i.d. then the "optimal" W is $X'Z(Z'Z)^{-1}$ and the resulting estimator is obtained by two stage least squares. Here I propose a different estimation procedure. This procedure is designed to minimize the impact of misspecification. I assume that $Z'_i\varepsilon/T$ converges to V_i as T goes to infinity. However, instead of assuming V_i is zero, I treat V_i as an unknown random variable from the point of view of the econometrician. I assume that V_i has mean zero and variance σ_i^2 (so that, on average the estimates are consistent). Also the expected value of $V_i\ V_j$ is zero so that the asymptotic biases from the different instruments are uncorrelated. Under these circumstances I discuss the instrumental variables estimator which minimizes the asymptotic covariance matrix of $\hat{\beta}$. I show that this optimal $\hat{\beta}$ is strictly inside the polyhedron whose vertices are obtained from estimating the $\left|\begin{matrix} m \\ n \end{matrix}\right|$ exactly

identified submodels. I also show that the estimates obtained from two stage least squares are not necessarily inside this polyhedron. The paper proceeds as follows. Section II shows the arbitrariness of $\hat{\beta}$ when the model is misspecified, while in Section III my solution to this arbitariness is based on priors over V_i. Section IV concludes.

I. The Arbitrariness of the Estimated Parameters

Let $\hat{\beta}$ be the value of β which satisfies (2). Then:

$$\hat{\beta} = (W Z' X)^{-1} W Z' Y \tag{3}$$

Let $\beta^{j1,j2 \ldots jk}$ for $j \leq 1 < j2 \ldots < jk \leq m$ be the estimate of β obtained from using the instruments $Z_{ji} \ldots Z_{jk}$. This estimate is given by

$$\beta^{j1 \ldots jk} = \begin{bmatrix} Z_{j1} & X_1 \ldots X_k \\ Z_{j2} & \\ \vdots & \\ Z_{jk} & \end{bmatrix}^{-1} \begin{bmatrix} Z'_{ji}Y \\ \vdots \\ Z'_{jk}Y \end{bmatrix} \tag{4}$$

where X_j is the jth column of X.

Proposition 1

$$\hat{\beta} = \sum_{1 \leq j_1 < \ldots < jk \leq m} \alpha_{j1 \ldots jk} \beta^{j1 \ldots jk} \tag{5}$$

where the α's sum to one.

Proof

Consider the i^{th} element of $\hat{\beta}$. It is given by A_i/B. A_i is the inner product of the vector whose elements are the cofactors of the i^{th} column of $(W Z' X)$ with $W Z' Y$ while B is the determinant of $WZ' X$. Thus A_i is the determinant of a matrix formed by deleting the i^{th} column of $(W Z' X)$ and replacing it by

$W Z'Y$. Using the Cauchy-Binet identity as in Gantmacher (1959) the determinant of $W Z' X$ can be written as

$$|WZ'X| = \sum_{i \leq j1 < j2 \ldots jk \leq m} \begin{vmatrix} W_{1,j1} & \ldots & W_{1,jk} \\ W_{k,j1} & \ldots & W_{h,jk} \end{vmatrix} \begin{vmatrix} Z'_{j1}X_1 & \ldots & Z'_{j1}X_k \\ Z'_{jk}X_1 & \ldots & Z'_{jk}X_k \end{vmatrix} \qquad (6)$$

where W_{ij} is the typical element of W. That is, the determinant can be written as the sum of the products of determinants obtained from selecting k columns of W and the corresponding k rows of Z'X. Similarly A_i is given by:

$$A_i = \sum_{1\ j1 \ldots jk\ m} \begin{vmatrix} W_{1,j1} & \ldots & W_{1,jk} \\ W_{k,j1} & \ldots & W_{k,jk} \end{vmatrix} \begin{vmatrix} Z_{j1}X_1 & \ldots & Z'_{j1}X_{i-1} & Z'_{j1}Y & \ldots & Z'_{j1}X_k \\ Z_{jk}X_1 & \ldots & Z'_{jk}X_{i-1} & Z'_{jk}Y & \ldots & Z'_{jk}X_k \end{vmatrix}$$

Hence, using (11)

$$\hat{\beta}_i = \sum_{1 \leq j1 \ldots jk \leq m} \alpha_{j1 \ldots jk} \; \beta_i^{j1 \ldots jk}$$

where

$$\alpha_{ji \ldots jk} = \frac{\begin{vmatrix} W_{1,j1} & \ldots & W_{1,jk} \\ W_{k,j1} & \ldots & W_{k,jk} \end{vmatrix} \begin{vmatrix} Z'_{j1}X_1 & \ldots & Z'_{j1}X_k \\ Z'_{jk}x_1 & \ldots & Z'_{jk}x_k \end{vmatrix}}{\sum_{1 \leq j1 < j2 \ldots jk \leq m} \begin{vmatrix} W_{1,j1} & \ldots & W_{1,jk} \\ \vdots & & \vdots \\ W_{k,j1} & \ldots & W_{k,jk} \end{vmatrix} \begin{vmatrix} Z'_{j1}X_1 & \ldots & Z'_{j1}X_k \\ \vdots & & \vdots \\ Z'_{jk}X_1 & \ldots & Z'_{jk}X_k \end{vmatrix}} \qquad (7)$$

Q.E.D.

If the model is correctly specified, Hansen (1982) shows that any W of full rank leads to consistent estimates. Thus even when the matrices of second cross moments between $Z_{ji} \ldots Z_{jk}$ and X are positive definite, the submatrices of W can have negative determinants leading to negative α's in (7).

If the model is correctly specified, that is, if $Z_1'\varepsilon/T$ goes to zero with probability one, then every $\beta^{j1 \ldots jk}$ goes to β with probability one. In this case the signs of the α's in (7) have no importance, as long as the α's sum to one $\hat{\beta}$ is asymptotically equal to β with probability one.

On the other hand, as mentioned in the introduction, many economic models appear to be misspecified. The Hausmann (1978) test detects misspecification, i.e. differences between $Z'\varepsilon/T$ and zero when the $\beta^{j1 \ldots jk}$'s are significantly different from each other. Newey (1983) shows that the test proposed by Hansen (1982) is actually equivalent to the Hausman test under certain circumstances. So failures of these tests mean that the estimates obtained from exactly identified submodels (i.e. from using only $(Z_{j1} \ldots Z_{jk})$ as instruments) differ. Proposition 2 establishes that if these estimates differ enough asymptotically then the value of $\hat{\beta}$ is arbitrary.

Proposition 2. Suppose the $\beta^{j1 \ldots jk}$ converge asymptotically to constants. Moreover, there exist k+1 instruments such that the matrix whose columns are the estimates of the (k+1) exactly identified submodels which use these instruments is of full rank. Then, for every kx1 vector γ, one can construct a matrix W which makes $\hat{\beta}$ equal to γ.

Proof

First let W have nonzero elements only in the columns which correspond to the (k+1) instruments with the desired property. The (k+1) determinants obtained by selecting k of the nonzero columns of W are clearly arbitrary. For instance, by multiplying the first k columns with this property by λ and the last column by $1/\lambda^{k-1}$ one multiplies the first subdeteriminant by λ^k and leaves the others unchanged. Moreover, if k is even, multiplying the first k nonzero columns of W by (-1) changes the sign of all determinants except for the one of the first submatrix. Finally, multiplying a row by (-1) changes the sign of all the determinants. So, these last two operations allow

one to change just the sign of the first determinant even when k is even. A similar argument allows one to change at will any of the other determinants. Since Z'X is fixed, one can thus change at will the numerators of the α's in (7). In particular one can pick the last subdeterminant of W in such a way that the denominator of the α's in (7) is equal to one. Then $\hat{\beta}$ can be rewritten as

$$\hat{\beta} = \beta^L + CS$$

Here β^L is the estimate of β obtained using only the instruments which correspond to the last k nonzero columns of W. C is a matrix whose columns are given by the difference between the k β's obtained from the other exactly identified submodels and β^L. Finally S is a kx1 vector consisting of the k arbitary α's. If the (k+1) instruments have the desired property C is of rank k. Then one can obtain $\hat{\beta}$ equal to γ by setting S equal to $C^{-1}(\gamma-\beta^L)$.

Proposition 1 basically shows that $\hat{\beta}$ is a linear combination of β's obtained from the exactly identified submodels. Moreover, since each one of these β's is consistent, the sum of the weights on these β's must be one to ensure that $\hat{\beta}$ is consistent. However, the individual weights are arbitrary except for their need to sum to one. So one weight can be large and positive as long as another is is large and negative. As soon as any two β_i's from exactly identified submodels differ one can thus obtain an arbitary value for $\hat{\beta}_i$ by varying the weights on the two exactly identified β_i's.

This arbitariness of $\hat{\beta}$ is disturbing for a number of reasons. First, a number of different W's have been proposed for their "optimal" properties when the model is correctly specified. This, unfortunately, gives econometricians quite a bit of latitude in reporting estimates of models which fail Hausmann-type tests. In particular, under the assumption of conditional homoskedascity the optimal W is $X'Z(Z'Z)^{-1}$ which gives the two stage least squares estimator. Instead under conditional heteroskedasicity Hansen (1982) shows that the

optimal W is obtained in two stages. In the first stage any W of full rank can be used. Then the residuals from the first stage estimation are used to construct the optimal W. Proposition 2 makes it clear that this second stage W will, if the model is misspecified, vary depending on the first stage W that is chosen. When there are simultaneous equations being estimated then W will also be different depending on whether three-stage least squares or iterative three stage least squares are selected.

The second reason the arbitariness of $\hat{\beta}$ is disturbing is that it suggests nothing can be learned about β even when the model is only slightly misspecified. This is intuitively inplausible. My discussion of the situations in which something can be learned is relegated to the next section. In the rest of this section I consider whether the arbitariness of $\hat{\beta}$ disappears when instead of using the generalized method of moments one minimizes $(Y-X\beta)'Z\,W\,Z'(Y-X\beta)$ where W is a mxm positive definite weighing matrix. I show that this isnt' so by focusing on an example. In this example k is equal to one while m is two.

Thus there are two exactly identified submodels. One uses only Z_1 as an instrument while the other uses only Z_2. The instrumental variable estimates from the two submodels are given by

$$\beta^1 = \frac{Z_1'Y}{Z_1'X} = + \frac{(Z_1'\varepsilon)/T}{(Z_1'\varepsilon)/T} \qquad (8)$$

$$\beta^2 = \frac{Z_2'Y}{Z_2'X} = \beta + \frac{(Z_2'\varepsilon)/T}{(Z_2'X)/T}$$

The estimates β^1 and β^2 become good approximations to the true β as T becomes large if $Z_1'\varepsilon/Z_1'x$ and $Z_2'\varepsilon/Z_2'x$ converge respectively to $\tilde{Z}_1$ and $\tilde{Z}_2$ which are small relative to β. On the other hand consider the estimate $\tilde{\beta}$ which

minimizes

$$(Y-X\beta)'[Z_1 Z_2] \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} Z_1' \\ Z_2' \end{bmatrix} (Y-X\beta) \tag{9}$$

where $\tilde{W}$ has been chosen without loss of generality to be symmetric. $\tilde{\beta}$ is given by:

$$\tilde{\beta} = \frac{(aX'Z_1 + bX'Z_2)Z_1'Y + (bX'Z_1 + cX'Z_2)Z_2'Y}{(aX'Z_1 + bX'Z_2)Z_1'X + (bX'Z_1 + cX'Z_2)Z_2'X}$$

$$= \phi\, \beta^1 + (1-\phi)\beta^2$$

where

$$\phi = \frac{(aX'Z_1 + bX'Z_2)Z_1'X}{(aX'Z_1 + bX'Z_2)Z_1'X + (bX'Z_1 + cX'Z_2)Z_2'X} \tag{10}$$

so $\hat{\beta}$ is a weighted sum of β^1 and β^2 where the weights add to one. So, asymptotically,

$$\hat{\beta} = \beta + \tilde{Z}_1 + (\tilde{Z}_2 - \tilde{Z}_1)(1-\phi) \tag{11}$$

Unfortunately ϕ can be any real number so that if $\tilde{Z}_2$ is different from $\tilde{Z}_1$, $\tilde{\beta}$ is arbitrary. This can be seen as follows: by normalizing the Z's one can make both $Z_i'X/T$ converge to one. Then ϕ is equal to $(a+b)/(a+2b+d)$. Let a equal one, d equal $(1+\mu)$ where μ is bigger than $-.5$ and b equal to $\nu - \sqrt{1+\mu}$. As long as ν is small and positive the resulting weighting matrix is positive definite. Then:

$$\phi = \frac{1+\nu - \sqrt{1+\mu}}{2+\mu - 2\sqrt{1+\mu} + 2\nu} \tag{12}$$

and

$$\lim_{\substack{\nu \to 0 \\ \mu \to 0}} \phi = \frac{1+\nu - (1-\mu/2)}{2+\mu - 2(1+\mu/2)+2\nu} = \frac{\nu - \mu/2}{2\nu} \tag{13}$$

So, for μ positive ϕ can be induced to be in the open interval $(-\infty, 1/2)$ by varying ν. Similarly for μ negative ϕ can be in the open interval $(1/2,\infty)$. On the other hand making d equal a and choosing b equal zero ϕ becomes 1/2.

III. The Study of Misspecified Models

The previous section showed that if the model is misspecified, one can choose weighting matrices to obtain arbitrary parameters. This is true even if the model is only slightly misspecified in the sense that the $\beta^{j1 \cdots jk}$'s are close to β. As long as they are slightly different from each other, proposition 2 holds. However, if the $\beta^{j1 \cdots jk}$ are in the economic sense very similar to each other, the statistical rejection of their equality should not be viewed as a major problem. The question remains however which W to use even in this case.

One possibility is to view the failure of specification tests as a failure of a specific set of m-k instruments under the maintained assumption that the other k instruments have the untestable property that $\lim_{T\to\infty} Z_i'\varepsilon = 0$. Then, the best weighting matrix has nonzero elements only in the columns corresponding to the "valid" instruments. While this procedure may be appropriate in certain contexts, it is not so in general. In macroeconomics the Z_i''s are typically lagged values of various variables. Which of these lags is most appropriate is generally hard to decide. In panel data the instruments are usually individual characteristics like age, schooling and the wage of the working spouse. It might be thought that the last characteristic is a worse instrument in a labor supply equation for instance. However, it would seem that even the first two characteristics are probably correlated with the taste for working.

So, in the usual context it is difficult to assert that one is sure k of the instruments are indeed appropriate ones. Here, I propose a different mode of analysis of models which fail specification tests. In particular I propose that the polyhedron composed of the ß's from the exactly identified submodels be studied. If this polyhedron is large in that the various ß's have very different economic implications, the misspecification makes it hard to draw behavioral conclusions from the data. On the other hand, if the polyhedron is small, the statistical significance of misspecification doesn't stand in the way of drawing behavioral implications. The focus on the polyhedron is motivated by the fact that under assumptions strictly weaker than that the $\lim_{T\to\infty} Z'_i\,\varepsilon/T = 0$ the estimator which minimizes misspecification is indeed inside this polyhedron.

Suppose that $\lim_{T\to\infty} Z'_i\varepsilon/T$ is equal to V_i where V_i is a constant. This convergence of $Z'_i\varepsilon/T$ is also required to make the $\beta^{ji\,\cdots\,jk}$'s converge. Instrumental variables are inherently underidentified assymptotically since one cannot learn the ß's and the m V_i's. The usual "identifying" assumption is that the V_i are zero. This cannot of course be true of all the V_i if the model fails a specification test. Here I assume that econometricians do not know the values of V_i. Instead there are willing to entertain a prior distribution over V_i. Since it is felt that the instruments are reasonably close to being valid, the mean of this prior is zero. On the other hand the prior variance of V_i is nonzero and this is the weakening of the standard assumption.

This randomness of V_i can be interpreted as follows. As long as the random variable $Z_{it}\varepsilon_t$ is stationary, V_i converges almost surely to the expected value of $Z_{it}\varepsilon_t$ conditional on an invariant set of J. If, in addition, the variables $Z_{it}\varepsilon_t$ are ergodic, the invariant sets have either probability zero or one. In this case V_i converges almost surely to the unconditional mean of $Z_{it}\varepsilon_t$. On the other hand, suppose $Z_{it}\varepsilon_t$ is not erdogic. Then there are nontrivial invari-

ant subsets of the set Ω of underlying states of the world. These subsets have the property that, once the economy starts in one of these subsets, it never reaches outside the subset. Hence V_i depends explicitly on which subset the economy starts in. Then, even if the unconditional mean of $Z_{it}\varepsilon_t$ is zero, the asymptotic value of V_i can be treated as a random variable whose realization depends on the actual invariant subset in which the economy is stuck. The probability of this realization depends on the prior probability of this particular invariant subset.

I do not, however, consider completely general priors. Instead, I assume that the prior covariance matrix of V_i,Σ is diagonal. This assumption has the advantage of parsimony. If, before encountering a rejection with a specification test, an econometrician considered that a set of instruments were strictly valid, it is hard to imagine that he/she knows after the rejection how the misspecification due to one instrument depends on the misspecification caused by another. This suggests as a natural starting point the assumption that the misspecifications are uncorrelated. If, in a particular application economic theory predicts the off-diagonal terms of Σ, it should obviously be applied. On the other hand I am unaware of theories which make this type of prediction. Such theories would have to deal explictly with the invariant sets of Ω.

The standard errors in variables case considered for instance by Leamer [1978] has a residual which can be decomposed in two additive parts. The first part (the structural one) is only correlated with the dependent variable Y while the second part (the measurement error) is correlated only with X. Then treating Y and X as instruments the convariance of $Y_t\varepsilon_t$ with $X_t\varepsilon_t$ is zero and a fortiori in the iid case, so is the covariance between $Y'\varepsilon/T$ and $X'\varepsilon/T$. So this example satisfies my diagonal covariance assumption. However, in general, it isn't required for $Z_{it}\varepsilon_t$ to be uncorrelated with $Z'_{jt}\varepsilon_t$, for $Z_i'\varepsilon/T$ to be uncorrelated with $Z_j'\varepsilon/T$.

I now establish that under these assumptions about the V_i, the asymptotic variance covariance matrix of $\hat{\beta}$ gets minimized by picking an estimate strictly inside the polyhedron of estimates from exactly identified submodels.[2] This variance covariance matrix is given by the limit as T goes to infinity of $(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$. When the model is correctly specified this is simply zero and we focus on the "first order" variance covariance matrix given by the expected value of $T(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$. Here, however, since the V_i are random from the point of view of the econometrician $\hat{\beta}$ is a random variable and the $E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$ is well defined. Instead, in the presence of this type of misspecification $E\,T(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$ blows up almost surely as T goes to infinity.

Proposition 3

If $\lim_{T\to\infty} Z'\varepsilon/T$ has mean zero and a diagonal variance covariance matrix Σ, the instrumental variable estimator which minimizes the expectation of $(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$ is given by (7) with all the α's between zero and one.

Proof

(from (3))

$$\hat{\beta}-\beta = \left(\frac{W\,Z'X}{T}\right)^{-1} W\,V$$

where the typical element of V, V_i is given by $\lim_{T\to\infty} Z'\varepsilon/T_i$. Then, the asymptotic variance covariance matrix of $\hat{\beta}$ is

$$E\,(\hat{\beta}-\beta)(\hat{\beta}-\beta)' = E\left(\frac{W\,Z'X}{T}\right)^{-1} W\,V\,V'\,W'\left(\frac{X'Z}{T}W'\right)^{-1}$$

which is clearly minimized for

$$W = \frac{X'Z}{T}\Sigma^{-1} \qquad (14)$$

Thus the matrix composed of the columns j1 ... jk of W is given by:

$$\begin{vmatrix} W_{1,j1} & \cdots & W_{1,jk} \\ W_{k,j1} & \cdots & W_{k,jk} \end{vmatrix} = \frac{1}{T}\begin{vmatrix} X_1'Z_{j1} & \cdots & X_1'Z_{jk} \\ X_k'Z_{j1} & \cdots & X_k'Z_{jk} \end{vmatrix}\begin{vmatrix} 1/\sigma_{j1}^2 & \cdots & 0 \\ 0 & \cdots & 1/\sigma_{jk}^2 \end{vmatrix}$$

Q.E.D.

where σ_{ji}^2 is the expected value of V_{ji}^2. Thus the numerator as well as each of the elements of the denominator of (14) are positive. Hence all α's are positive and less than one.

Note that the two stage least squares estimator become optimal if E(V V') is proportional to (Z'Z)/T.

If Σ were know up to a multiplicative constant, the optimal estimator of ß would use the weighing matrix given by (14) with the population moments replaced by the sample moments. For instance, it might be thought that once the Z_i are normalized to have the same mean, the σ_i^2 are all equal. Then the optimal estimation of $\hat{\beta}$ is simply $(X'Z\ Z'X)^{-1}(X'Z\ Z'Y)$. If, on the other hand, information on the σ_i^2 is unavailable, then it is better to analyze only the bounds given by the polyhedron of exactly identified submodels.

It might be thought that two stage least squares which is optimal when the model is correctly specified and the ε's are iid produces estimates which are at least inside this polyhedron. The following example based on the setup of (8) shows that this isn't necessarily true.

Suppose that:

$$\lim_{T\to\infty} \frac{X'X}{T} = \lim_{T\to\infty} \frac{Z_1'X}{T} = \lim_{T\to\infty} \frac{Z_2'X}{T} = 1 \tag{15}$$

$$\lim_{T\to\infty} Z_1'Z_1 = 4 \qquad \lim_{T\to\infty} Z_2'Z_2 = 2 \qquad \lim_{T\to\infty} Z_1'Z_2 = 2.7 \tag{16}$$

where (15) is obtained from normalization. This example naturally has a positive definite second moment matrix. The weighting matrix defined in (9) becomes:

$$\begin{vmatrix} a & b \\ b & c \end{vmatrix} = \frac{1}{0.71} \begin{vmatrix} 2 & -2.7 \\ -2.7 & 4 \end{vmatrix}$$

and ϕ, given by (10), is -1.17. As the correlation between Z_1 and Z_2 goes up with fixed variances, ϕ continues to fall. The correlation in this example is slightly above .95 which isn't unusual for macroeconomic time series.

Conclusions

This paper has shown that, if one believes that the biases introduced by the correlation of the instruments with the errors are independent, one should concentrate on the polyhedron composed of the estimates from the exactly identified submodels. The "best" estimator of $\hat{\beta}$ is inside this polyhedron. Moreover, the size of the polyhedron gives an idea of the economic importance of the misspecification. On the other hand, if one is unwilling to impose any a priori structure on the covariance matrix of V, it becomes essentially impossible to learn about the β's when the model fails a test of its overidentifying restrictions. This weakness of inference must be contrasted with the optimistic results of White (1982). He shows that in the maximum likelihood content the parameters converge asymptotically to a unique vector even when the model is misspecified. Moreover, in the iid case standard inference itself remains unperturbed under misspecification. Similar results are presented for least squares in White [1980 a,b]. Maximum likelihood and least squares have the advantage of being well specified optimization problems which they tend to have well behaved solutions. On the other hand, instrument variables procedures are not well specified optimization problems until weighting matrixes have been selected. Unfortunately, standard weighting matrices like those of two stage least squares appear to have desirable properties only when the model is correctly specified.

It might be thought that weighted least squares which is considered by White [1980 a,b] is also not a well specified optimization problem in this

sense. Indeed if the model is sufficiently misspecified, arbitrary parameter values can probably be obtained by varying the weighting matrix. However, at least for prediction purposes, White [1980 a,b] shows that weighted least squares is always dominated by unweighted least squares. So this limitation of weighted least squares appears to be much less severe than the limitation of intrumental variables discussed here.

FOOTNOTES

1 Rejections are also reported in Diamond and Hausman (1983) and Dubin and McFadden (1983). However, these authors' favored estimates are not subject to specification tests.

2 This is akin to Leamer's (1978) observation that in his errors in variables case the best estimator of β lies between the estimate obtained by regressing Y on X and the inverse of the coefficient obtained by regressing X on Y.

REFERENCES

Diamond, P and S. Hausman, "Individual Retirement and Savings Behavior," (1983) forthcoming in the Journal of Public Economics.

Dubin, J. and D. McFadden, "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," (1983), forthcoming in Econometrica.

Fisher, Franklin M., "On the Cost of Approximate Specification in Simultaneous Equation Estimation," Econometrica 29 (April 1961).

Gantmacher, F.R., Matrix Theory, Chelsea Publishing Co., New York, 1954.

Hansen, L., "Large Sample Properties of Generalized Methods of Moments Estimators," Econometrica 50 (1982), pp. 1024-1054.

Hansen, L., and K. Singleton, "Generalized Instrumental Variable Estimation of Nonlinear Rational Expectations Models," Econometrica 50 (1982), 1269-86.

Hausman, J.A., "Specification Tests in Econometrics," Econometrica 46 (1978), 1251-1272.

Leamer, L., Specification Searches, NY: Wiley, 1978.

Mankiw, G., J. Rotemberg and L. Summers, "Intertemporal Substitution in Macroeconomics," mimeo 1983.

Newey, W.K., "Generalized Method of Moments Specification Testing," Mimeo, November 1982.

Pindyck, R. and J. Rotemberg, "Dynamic Factor Demands and the Effect of Energy Price Shocks," (1983) forthcoming in the American Economic Review.

Reiersol, O., "Confluence Analysis by Means of Instrumental Sets of Variables," Arkiv for Mathematik, Astronomy och Fisik, 32 (1945).

Sargan, J.D. "The Estimation of Economic Relationships Using Instrumental Variables," Econometrica 26 (1958), pp. 343-415.

White, M., "Maximum Likelihood Estimation of Misspecified Models," Econometrica 50 (1982), pp. 1-26.

White, M., "Using Least Squares to Approximate Unknown Regression Functions," International Economic Review 21 (1980a), 149-70.

White, M., "Nonlinear Regression on Cross Section Data," Econometrica, 48 (1980b), 721-46.

Date Due

Lib-26-67

Bar Code
is on last
page