DISCUSSION PAPER NO. 450

EFFICIENT ESTIMATION OF TIME SERIES MODELS WITH PREDETERMINED, BUT NOT EXOGENOUS, INSTRUMENTS

by

Fumio Hayashi and Christopher Sims*/

December 1980

---

*The authors are grateful to John Huizinga and Robert Cumby for advice and criticism during the preparation of this paper. Sims's work on the paper was completed while he was in residence at the National Bureau of Economic Research and was also supported in part by NSF grant SOC7818042 to the University of Minnesota.

Abstract

Particularly in rational expectations models, it can occur that a model has serially correlated residuals, implies that those residuals are uncorrelated with contemporaneous and lagged values of a predetermined instrument, but does not imply that the instruments are strictly exogenous. This paper proposes a method for transforming such a model to one without serial correlation, while keeping the instrument predetermined. Standard theory of instrumental variables estimation then applies. If enough lagged values of the instrument are used, standard instrumental variables estimators applied to the transformed model approach the same bound on asymptotic efficiency which applies to another class of estimators for this problem which have been proposed by others.

EFFICIENT ESTIMATION OF TIME SERIES MODELS WITH PREDETERMINED, BUT NOT EXOGENOUS, INSTRUMENTS

by Fumio Hayashi and Christopher Sims July 1980

In a number of recently developed macroeconomic models behavioral equations arise in which error terms can be asserted on the basis of economic arguments to be orthogonal to some set of instrumental variables at a certain set of dates, but not to be orthogonal to the instruments at all dates. Examples of such models are in work by Kennan (1979), and Brown and Maital (1979), to name a few. Sometimes a theory of this type will assert that the error term in question is itself serially uncorrelated and is orthogonal to all current and past values of the instruments: this corresponds to the assumption that the instruments are predetermined, which is a standard textbook assumption, so standard textbook estimation methods apply. However, often the relevant theory does not predict the absence of serial correlation, and serial correlation turns out to be substantial in fact. This situation is not a standard textbook situation, and researchers have not always avoided (at least in their first drafts) the pitfall of applying textbook methods to these situations, where the methods are not justified.

Hansen (1980) and Cumby, Huizinga and Obstfeld (1980) derived consistent estimators for models like these without a separate, explicit "correction for serial correlation" in the estimation process. Hansen (1980) derives the asymptotic distribution for these estimators and shows that they are efficient within a certain class. We will call these estimators "finite-order efficient" (FOE). Here we describe a method for making a serial correlation correction in such models which makes standard instrumental variables estimators apply to them. The class of estimates so generated is different from the FOE's class and may be either more or less asymptotically efficient.*

* The forward-filtered instrumental variables estimator was invented by Hayashi and described in his thesis (1980).

Hansen and Sargent (1980) in a footnote pointed out that ideally one would use all the available orthogonality conditions in such situations. In particular, when all lagged values of the instruments are uncorrelated with the current residual, all these zero correlations should be exploited in estimation. Recently Hansen and Sargent in unpublished work have derived in the frequency domain the form of the optimal estimator exploiting all these conditions. Applying an insight described also in Cumby, Huizinga, and Obstfeld (1980) this paper shows a connection between this paper's "forward-filtered", estimates and the FOE estimates. The optimal estimator can be approximated arbitrarily well by estimators of either class. Besides providing an interpretation of the optimal estimator which may be more transparent than that in Hansen and Sargent's unpublished work, this paper's result suggests an estimation method which is easily implemented with standard econometric statistical packages and may in some instances be the most attractive computational approach when existing statistical packages are not a constraint.

An example

To see concretely why these issues are important and the nature of the solutions proposed, consider the case of testing the rationality of expectations. We assume we have data on "expected Y", $\bar{Y}_t$ and on actual Y , Y_t . The rationality hypothesis asserts that $E(Y_t \mid \bar{Y}_s$, all $s \leq t) = \bar{Y}_t$. Thus, we can estimate the linear regression

1) $$Y_t = b\bar{Y}_t + u_t$$

and expect to find b=1 . Orthogonality of u_t to $\bar{Y}_t$ follows from the theory. Orthogonality of u_t to lagged values of u_t follows only if the information on which $\bar{Y}_t$ is based includes Y_{t-1} . It can easily happen that, say, the

forecasts are prepared at 6-monthly intervals for a 12-month time horizon, so that it is clear Y_{t-1} is not available when $\bar{Y}_t$ is formed.

One way to proceed would be to use OLS on (1), but use an asymptotically justified estimate of the covariance matrix of the estimates. This procedure, which uses $\bar{Y}_t$ as an "instrument" for itself, is in the class of estimates which we label "finite order efficient". This paper proposes, following Hayashi (1980), that we filter (1) to produce serially uncorrelated residuals, but use a forward filter -- i.e., replace Y_t, $\bar{Y}_t$, and u_t by the corresponding series filtered through $B(L^{-1})$, a polynomial in non-positive powers of the lag operator L. If B is chosen properly, the filtered u will be serially uncorrelated. Furthermore, since it has been filtered forward, orthogonality of $\bar{Y}_t$ and the filtered residual holds -- this follows from the assumption that $\bar{Y}_t$ is based on information which is still available at $t+s$ for any $s \geq 0$. Thus, $\bar{Y}_t$ remains eligible as an instrument for the transformed equation, and in fact the usual theory of instrumental variables estimation applies here -- we have transformed to a system in which $\bar{Y}_t$ is predetermined in the textbook sense. However, because $\bar{Y}_t$ is less well correlated with the filtered $\bar{Y}_t$ than $\bar{Y}_t$ is with itself, it is not clear that this "forward-filtered IV" procedure using $\bar{Y}_t$ as an instrument is in fact any more efficient than OLS.

The general model and the general estimator

We consider an equation

$$Y_t = X_t b + u_t \quad , \tag{4}$$

with the identifying assumption that $Z'_{t-s}u_t$ has zero expectation for all $s \geq 0$. We assume all the variables jointly stationary for convenience of exposition, though as Hansen (1980) has pointed out the assumption could be relaxed. If we use the finite list of qk instrumental variables $Z_t,\ldots Z_{t-k}$, then, as is shown in Hansen (1980), the best estimator of b based on this list of instruments is given by:

$$5) \qquad b = (X'ZM^{-1}Z'X)^{-1}X'ZM^{-1}Z'Y \ ,$$

where Z is the $Tx[(1+k)q)]$ matrix of instrumental variables, X is the Tx_p matrix of right-hand-side variables, Y is the Tx1 matrix of dependent variables, and $M = \lim T^{-1}E(Z'uu'Z)$. Clearly the more instruments we use (i.e., the larger is k) the more efficient our estimator will be asymptotically, in general.

The reader is referred to Hansen (1980) for a careful discussion of the regularity conditions needed to justify asymptotic normality and consistency for this type of estimator; they are also discussed in the Appendix.

In practice, the list of instruments used will always be finite, but it seems reasonable to use a longer list in a larger sample. Undoubtedly it is possible (perhaps along the lines suggested by Geweke (1979)) to give an explicit rule for choosing k which under fairly broad conditions guarantees consistency and results in an estimator which is asymptotically better than any based on a fixed finite k.

It may also be possible in some applications to do better than choosing k "large, but not too large" by taking explicit account of knowledge or data relating to the determination of X and Z. We might have a model of the way Z determines X, for example, which would tell us that lags in Z beyond some order are necessarily orthogonal to X. In a nonlinear system we might discover

nonlinear functions of Z which are better correlated with X than any linear combination of Z's and which retain the property of orthogonality to the u's. Thus, the usual considerations relevant to choosing between single-equation and multiple-equation methods and to choosing between instrumental variables and maximum likelihood apply to this type of problem as well.

Ways of estimating M

Under a stationarity assumption, M can be estimated by estimating the autocovariance function for the stochastic process $Z_t' u_t$. Of course, this autocovariance function involves infinitely many unknown parameters, in general, so that estimating M in a way that preserves consistency of the estimator is a non-trivial problem. Hansen (1980) provides some guidelines for the general case.

In some applications, including the type of forecasting example discussed at the beginning of the paper, a plausible argument can be made that we should expect $E[u_t u_{t-s} \mid Z_v, \text{ all } v \leq t]$ not to depend on any of the Z's. This assumption is justified, e.g., if the Z's and u's are jointly normal. All of this paper's results on asymptotic behavior of estimators depend on this assumption.

Let $V = E[uu']$ and let $WW' = V$ be a factorization of V. Under the condition given above, $M = \text{plim } T^{-1}Z'VZ$. Now let $Z^* = W'Z$, $X^* = W^{-1}X$, and $Y^* = W^{-1}Y$. If Y^* and X^* replace Y and X in (4) the transformed equation clearly has serially uncorrelated residuals. Furthermore, $Z^{*\prime}u^* = Z'u$, so the transformed instruments are orthogonal to the transformed residuals. Finally, it is easily verified that $\hat{b}$ from (5) is just two-stage least squares using the transformed equation and transformed instruments, i.e.

6) $$\hat{b} = (X^{*\prime}Z^{*}(Z^{*\prime}Z^{*})^{-1} Z^{*\prime}X^{*})^{-1} X^{*\prime}Z^{*}(Z^{*\prime}Z^{*})^{-1} Z^{*\prime}Y^{*} .$$ */

---

* This characterization of the finite-order efficient estimator was proposed by Cumby, Huizinga and Obstfeld (1980).

The factorization $V=W'W$ has an element of arbitrariness. One natural way to normalize W is to require that it be upper triangular. If X_1 is at the top of X and X_T at the bottom, as in the conventional notation, u_t^* will then be a combination of u_{t+s} for $s \geq 0$, while Z_t^* will be a combination of Z_{t+s} for $s \leq 0$. Thus we can obtain the estimator in (6) as follows: First estimate V, e.g., by a preliminary instrumental variables procedure which ignores serial correlation, then filter the equation forward to eliminate serial correlation and filter the instruments backward by the inverse of the same filter; then apply two-stage least squares.

The limiting covariance matrix for the estimator in (6) is

7) $$S = (X^{*\prime}Z^{*}(Z^{*\prime}Z^{*})^{-1} Z^{*\prime}X^{*})^{-1}$$

As is familiar from textbook discussions of two-stage least squares, S can be characterized as the limiting covariance matrix of $\hat{X}_t^*$, where $\hat{X}_t^*$ is the projection of X_t^* on the instruments. But now by construction the filter applied to the Z's to yield Z^* is invertible and one-sided. It is relatively easy to verify that the space spanned by Z_s^*, all $s \leq t$, is the same space as that spanned by Z_s, all $s \leq t$, when Z_s^* is obtained from Z_s by a one-sided (backward in time) invertible filter. Clearly the lower bound to (7) is the variance-covariance matrix of the projection of X_t^* on the whole space spanned by Z_s^*, $s \leq t$. But this is the same as the variance-covariance matrix of the projection of X_t^* on all current and past Z_t's. Thus, though estimates based

on using k lagged values of Z^* as instruments will always be different from estimates based on k lagged values of Z itself, for large k the difference must become negligibly small. These assertions are discussed more carefully in the Appendix.

Thus, for this case we can somewhat simplify the prescribed procedure: The step of transforming the Z's by the inverse of the filter applied to the equation will be of little marginal benefit if k is chosen fairly large. It is largly a matter of convenience whether we use (5) with $Z'VZ$ replacing M or instead filter the system forward and then apply 2SLS using a long list of current and lagged Z's as instruments.

Choosing among estimators in practice.

Though it is not hard to see that there might be other specific implementations of these estimators, we consider just two,"forward-filtered instrumental variables" $\hat{b}_{FF}$ and "finite-order efficient instrumental variables" $\hat{b}_{FOE}$. $\hat{b}_{FF}$ is characterized as follows: Estimate the equation by some consistent method; estimate a model for the serial correlation of the residuals from the estimated residuals; filter the equation variables forward in time by a filter which is estimated to eliminate serial correlation in the residuals; apply instrumental variables, using k current and lagged values of the original instrument vector. $\hat{b}_{FOE}$ is characterized this way: Estimate the equation by some consistent method; estimate a model for the serial correlation of the residuals from the estimated residuals; use the estimated model of the residual to form an estimate $\hat{V}$ of the covariance matrix V of the residuals and thus of M, using k lagged values of the original instrument vector Z in Z^k, as $M=T^{-1}\ Z^{k\prime}\hat{V}Z^k$.

Finding $\hat{b}_{FF}$ involves inversion of the V matrix in principle, which is of the order of sample size. But as is made explicit in the above characterization, this inversion is easily approximated as filtering the data, which is much easier than a general T-order inversion. Of course, the filtering operation will either require dropping some observations or computation of correct weights for observations at the end of the sample. Finding $\hat{b}_{FOE}$ requires forming M and inverting it. The inversion is of order k, thus not terribly burdensome. Computational considerations thus leave the two estimators closely comparable, except that $\hat{b}_{FF}$ can be implemented with a package that does two-stage-least-squares and data transformations, while $\hat{b}_{FOE}$ requires separate matrix-manipulation facilities.

Since Hansen's proofs already demonstrate that $\hat{b}_{FOE}$ is the most efficient estimator exploiting the orthogonality condition $EZ^{k'}u=0$ for a given k, it might seem that $\hat{b}_{FOE}$ should generally be better than $\hat{b}_{FF}$ for a given k. But this is not so, because for a given k $\hat{b}_{FF}$ in general exploits all the orthogonality conditions and thus is not in the class with respect to which $\hat{b}_{FOE}$ is best.

To be specific, suppose that, as is likely in the forecasting example in the beginning of the paper, the residuals in equation (1) are a low-order moving average -- say first order. Let the parameter of the moving average be α, so that $u_t=(1+\alpha L)e_t$. Then the relative efficiency of $\hat{b}_{FF}$ and $\hat{b}_{FOE}$ is determined by which gives the higher R^2 in determining $(1+\alpha L^{-1})^{-1}\bar{Y}_t$: k lagged values of $\bar{Y}_t$ or k lagged values of $(1+\alpha L)\bar{Y}_t$. It is easy to see that, say, in the case where $\bar{Y}_t$ is itself a moving average of first order with parameter α, and k is one, the $\hat{b}_{FF}$ estimator has asymptotic variance $(1+\alpha^2)/(1+2\alpha^2/(1+\alpha^2)^2)$ times that of $\hat{b}_{FOE}$. This is better by 33% than $\hat{b}_{FOE}$ for α near one, but worse by 6.7% than $\hat{b}_{FOE}$ for $\alpha=.4$.

When $\tilde{Y}_t$ is a first-order autoregression and u_t a moving average, $\hat{b}_{FF}$ is always asymptotically more efficient than $\hat{b}_{FOE}$, by as much as a factor of 2 when $\tilde{Y}_t$ is serially uncorrelated and α approaches one.

An intuitive explanation for the tendency of $\hat{b}_{FF}$ to be more efficient in these examples runs as follows. Both estimators can be thought of as instrumental variables applied to the forward-filtered equation. $\hat{b}_{FF}$ uses Z_t as instrument, while $\hat{b}_{FOE}$ uses $a^{-1*}Z_t$. Now if Z_t and X_t are very closely related and u_t is positively serially correlated, the forward-filtering of X_t will reduce its serial correlation and thereby weaken its relation to Z_t. If Z_t is backward-filtered through a, this <u>increases</u> Z_t's serial correlation, further weakening its relation to X_t.

That this heuristic argument is not exact is clear from the example in which we found $\hat{b}_{FOE}$ better for some a's and serial correlation patterns in Z. Nonetheless, it is probably more than chance that plausible examples with $\hat{b}_{FOE}$ much more efficient than $\hat{b}_{FF}$ are hard to construct.

APPENDIX

We take as given the vector stochastic processes $Z(t)$, $X(t)$, and $u(t)$ with $Y(t) = X(t)b+u(t)$.

The stochastic process Z is our candidate for an instrumental variable, and the "forward-filtered" instrumental variables estimator is given by

A1) $$\hat{b}_{FF} = (\bar{X}'Z^k(Z^{k\prime}Z^k)^{-1}\bar{X}'Z^k(Z^{k\prime}Z^k)^{-1}Z^{k\prime}\bar{y} \quad .$$

Each matrix in the formula (A1) is of length $T-S(T)$. $S(T)$ is the number of observations lost from the end of the sample in filtering. The typical row of $\bar{X}$ is $\bar{X}(t) = \hat{a}'*X(t)$ and of $\bar{y}$ is $\bar{y}(t) = \hat{a}'*y(t)$, where the "*" indicates convolution and the "'" when applied as here to a function of time indicates reversal of the time index, so $\hat{a}(s) = \hat{a}'(-s)$. The rationale for the estimator suggests that $\hat{a}$ should be an estimate of the autoregressive operator a for u , where $a*u(t)=e(t)$, $e(t)$ uncorrelated with $u(s)$, all $s<t$, and $a(0)=1$, $a(s)=0$ for $s<0$. The typical row of Z^k is $Z^k(t) = [Z(t), Z(t-1),\ldots,Z(t-k)]$.

The "finite-order efficient" estimator based on the same set of $k+1$ lags of the instrument process is given by

A2) $$\hat{b}_{FOE} = (X'Z^k(Z^{k\prime}\hat{V}Z^k)^{-1}Z^{k\prime}X)^{-1}X'Z^k(Z^{k\prime}\hat{V}Z^k)^{-1}Z^{k\prime}y \quad ,$$

Where $\hat{V}$ is an estimate of V , the variance-covariance matrix of the length-T vector u of residuals. In (A2) all matrices are of length T .

We now list a set of assumptions which allow rigorous proof of the claims in the text. We follow Hansen in assuming stationarity. While relaxing the stationarity assumption is possible, doing so seems to complicate the proofs a great deal and to require increased stringency in assumptions about the distribution of errors. In these assumptions g is some monotone function on the positive integers such that $\sum_{s=0}^{\infty} g(s)^{-1} < \infty$.

i) $y(t)$, $X(t)$ and $Z(t)$ are jointly stationary, ergodic, a linear process of maximal rank, and linearly regular;[/]

[/] For definitions of ergodicity, maximal rank, and linear regularity see Rozanov (1966).

ii) $E[u(t) \mid Z(t-s), \text{ all } s \leq 0] = 0$;

(iii) $E[u(t)u(t-s) \mid Z(v), \text{ all } v \leq t] = E[u(t)u(t-s)]$, for any $s \geq 0$;

(iv) a has an inverse under convolution a^{-1} satisfying $|a^{-1}(s)| = 0$ for $s<0$ and $\sum |g(s)a^{-1}(s)|$ bounded;

(v) $S(T)/T \to 0$ as $T \to \infty$;

(vi) $\sqrt{T} \sum_{s=0}^{\infty} |\hat{a}(s)-a(s)|\, g(s)$ is bounded in probability;

(vii) $\max_s |\hat{a}^{-1}(s) - a^{-1}(s)|\, g(s)$ converges in probability to zero as $T \to \infty$.

viii) letting $\underline{\underline{X}}(t) = a^{*}X(t)$, $E[\underline{\underline{X}}(t)'Z^{k}(t)] = \Sigma_{\underline{\underline{XZ}}}$ is of rank p , where p is the length of the $X(t)$ vector ;

ix) $E[Z^{k}(t)'Z^{k}(t)] = \Sigma_{ZZ}$ is of full rank and $E[X(t)'X(t)] < \infty$;

x) the process $W(t) = \begin{matrix} Z(t) \\ u(t) \end{matrix}$ has a moving average representation $W = A*n$ with $g(s)$ $A(s)$ bounded.

Some of these assumptions could be given simpler form if we restricted the generality of the assumptions. For example, if the residuals u fit a finite-order ARMA model of known order, (iv) requires only that we rule out unit roots in the numerator of the moving average operator; (x) is redundant if W is a finite-order ARMA process. (iii) is redundant given the assumption of linearity in (i); it is retained for convenience in the proof and because it would be required separately in a variant of the theorem which could be obtained by dropping the linearity assumption and strengthening (x). Note that (v), (vi) and (vii) interact. If

our estimate of the filter is to be truncated at lag $S(T)$, this might affect whether it satisfies (vi) or (vii). Showing that one can choose an estimator for a which satisfies both (v), (vi) and (vii) might be technically difficult though it seems clear that, e.g., fitting a finite-order AR of order $S(T) = \sqrt{T}$ to the residuals from an initial consistent estimate of the equation would yield the desired properties. It is easy to check that if, e.g., u follows a finite-order ARMA process whose parameters are consistently estimated with convergence at rate $T^{-1/2}$, (iv), (vi) and (vii) are all satisfied if the ARMA operator has no unit roots.

First we display the asymptotic distribution for $\hat{b}_{FF}$.

Theorem 1: Under assumptions (i) through (x), except (iv) and (vii), we have

$$T^{\frac{1}{2}}(\hat{b}_{FF}-b) \text{ converges in distribution to } N(0,(\Sigma_{\bar{\bar{X}}Z}\ \Sigma_{ZZ}^{-1}\ \Sigma_{\bar{\bar{X}}Z}')^{-1}\ \sigma_e^2).$$

Proof: If a were known exactly and if there were no truncation so that $S(T)$ were always 0, then our assumptions would lead immediately to asymptotic normality by application of Hansen's Theorem 4.1. To be specific, if we define $\hat{\hat{b}}_{FF}$ by the same formula as used for $\hat{b}_{FF}$, but with $\hat{a} = a$ and $S(T)=0$, all T, then $\hat{\hat{b}}_{FF}$ is exactly of the form considered by Hansen, with Hansen's function f given by $Z^k(t)'(\bar{\bar{y}}(t)-\bar{\bar{X}}(t)b)$ and the weighting matrix which Hansen calls a_N (N being sample size) given by $\bar{\bar{X}}'Z^k(Z^{k\prime}Z^k)^{-1}$. Note also that because with a known the filtered residuals are exactly serially uncorrelated, we are in the standard case of instrumental variables and many derivations in the existing literature of an asymptotic distribution for two-stage least squares apply to $\hat{\hat{b}}_{FF}$. To be explicit about how our assumptions match Hansen's, his (i) follows from our (i); his (ii) from our (i) and (ix); his (iii) from our (i); his (iv) from our (i) and (x); his (v) immediately from the form of our version of the function f; his (vi) from our (viii); his (vii), consistency, from our (i), (ii), (viii), and (ix); and his (viii) from our (viii), (ix) and (i). It is easy to check that Hansen's expression for the asymptotic

The derivation of Hansen's (iv) from our (i) and (x) is the only one of these implications which requires a nontrivial argument. Hansen' (iv) puts a direct restriction on $Z(t)'u(t)$ as a stochastic process, a restriction which cannot in general be checked by examining the second-order properties of u and Z as a joint process. Our theorem, by assuming linearity, generates conditions which can be checked entirely in terms of covariance properties. To translate Hansen's condition (iv) into our notation, let $Q(t) = Z(t)'u(t)$. Let $V_s = E[Q(0) \mid Q(t), t \leq s] - E[Q(0) \mid Q(t), t < s]$. Hansen requires that $\sum_{s=0}^{\infty} || E[V_s V_s'] ||^{1/2}$ be finite. Now clearly $\sum_{s=0}^{\infty} E[V_s V_s'] = E[Q(s)Q(s)']$, which is finite by (iii) and (ix). Let $S_Q(t)$ be the sigma-field generated by values of $Q(s)$ for $s \leq t$ and $S_{Zu}(t)$ be that generated by values of $Z(s)$ and $u(s)$ for $s \leq t$. Clearly $S_{Zu}(t)$ contains $S_Q(t)$. But this means that $\sum_{s=m}^{\infty} E[V_s V_s']$, being the second moment of $E[Q(0) \mid S_Q(m)]$, is smaller than $E[E[Q(0) \mid S_{zu}(m)]E[Q(0) \mid S_{Zu}(m)]']$. But now $E[Q(0) \mid S_{Zu}(m)]$ can be written, by (x) and the linearity assumption, as $\sum_{s=m}^{\infty} a_Z(s)n(-s) \sum_{t=m}^{\infty} a_u(s)n(-s)$, where a_u and a_Z are the rows of A from (x) corresponding to Z and u, respectively. From this point on, using the fact that linearity means that the n process is serially independent, it is a straightforward calculation to show that the bound on $|| g(s)a_u(s) ||$ and on $|| g(s)a_Z(s)||$ generates a bound on $|| g(s)E[V_s V'_s] ||^{1/2}$, and thus Hansen's (iv).

It is easy to check that Hansen's expression for the asymptotic

covariance matrix matches our asserted covariance matrix for $\hat{b}_{FF}$. What remains is for us to show that $T^{\frac{1}{2}}(\hat{b}_{FF} - \hat{\hat{b}}_{FF}) \to 0$ in probability as $T \to \infty$.

First note that $T^{-1}Z^{k\prime}Z^{k}$ has the same probability limit whether Z^{k} has length T or $T-S(T)$ so long as (v) is satisfied. Second note that $T^{-1}\bar{\bar{X}}'Z^{k}$ has the same probability limit as $T^{-1}\bar{X}'Z^{k}$. This follows from the regularity conditions we have imposed on $\hat{a}$ and a , as follows.

Note that $E[|X(t)'Z^{k}(t+s)|]$ is bounded above, say by H , according to assumption (ix) , with the bound independent of s . But then we have

$$\left| \sum_{s=0}^{\infty} (\hat{a}(s)-a(s))T^{-1} \sum_{t=1}^{T-S(T)} X(t)'Z^{k}(t+s) \right|$$

$$\leq \left[\max_{s} |\hat{a}(s)-a(s)| \; g(s) \right] \sum_{s=0}^{\infty} T^{-1} \sum_{t=1}^{T-S(T)} g(s)^{-1} |X(t)'Z^{k}(t+s)| .$$

The left term in this latter expression converges in probability to zero by (vi) while the right term is bounded in probability because its expectation is less than $\sum g(s)^{-1} H$.

Now our task will be complete if we can show that

$$T^{-1/2} \sum_{t=1}^{T-S(T)} Z^{k}(t)'(\hat{a}'-a')*u(t)$$

converges in probability to zero. (As with the other terms, it is easy to show here that the extra $S(T)-1$ terms in the part of this expression corresponding to $\hat{b}_{FOE}$ are asymptotically negligible.) We proceed by showing that this term can be split into two pieces, the part summed over $t \leq N$, and the part summed from N onward. The latter can be made as small as desired with probability as high as desired by choosing N large enough, while the former converges in probability to zero, which will complete the proof.

Consider first

$$T^{-1/2} \sum_{t=N+1}^{\infty} Z^k(t)'(\hat{a}'-a')*u(t)$$

$$\leq T^{1/2}\left[\max_s \left| a(s)-\hat{a}(s)\ g(s) \right| \right] \sum_{s=0}^{\infty} T^{-1} \sum_{t=N+1}^{\infty} \left| g(s)^{-1} Z^k(t)'u(t+s) \right|$$

As in the last preceding argument, we have a bound on $E\left|Z^k(t)'u(t+s)\right|$ which is uniform in s, and since $\sum_{N+1}^{\infty} g(s)^{-1}$ can be made as small as desired by picking N large enough, we can make the right-hand term in the product as small as desired with probability as high as desired by picking N large enough. The left-hand term is bounded in probability by assumption (vii).

Since for any fixed $s \geq 0$ $T^{-1} \sum_{t=1}^{T-S(T)} Z^k(t)'u(t+s)$ converges almost surely to 0 by ergodicity, it follows directly that $T^{-1/2} \sum_{s=0}^{N} (\hat{a}(s)-a(s)) \sum_{t=1}^{T-S(T)} Z^k(t)'u(t+s)$ converges to zero in probability as $T \to \infty$. This completes the proof for this term, and thereby the proof of the theorem.

To discuss the asymptotic distribution for $\hat{b}_{FOE}$ we modify assumption (viii) as follows:

viii') $E[X(t)'Z^k(t)] = \Sigma_{XZ}$ is of full row rank p.

viii') $E[X(t)'Z^k(t)] = \Sigma_{XZ}$ is of full row rank p.

We also define $\hat{V}$ to be the matrix with typical element in row i, column j $R_u(i-j) = E[u(t)u(t-i+j)]$ and V to be the matrix with typical element $\hat{R}_u(i-j) = \hat{a}^{-1}*(\hat{a}^{-1})'(i-j)$. Now we can state

Theorem 2: Under assumptions (i)-(ix), with (viii') replacing (viii) and excluding (vi), $T^{-1}Z^{k\prime}\hat{V}Z^k$ converges in probability to $M = \sum_{s=0}^{\infty} R_u(s)E[Z^k(t)'Z^k(t-s)]$, which is full rank, and $T^{\frac{1}{2}}(\hat{b}_{FOE}-b)$ converges in distribution to $N(0,(\Sigma_{XZ}M^{-1}\Sigma_{XZ}')^{-1})$.

Proof:

That M is non-singular is guaranteed by the assumption in (i) that Z is of maximal rank. Note that if we define

$$M_N = \sum_{s=-N}^{N} R_u(s)E[Z^k(t)'Z^k(t-s)] \quad ,$$

then M_N can be made arbitrarily close to M by choosing N large enough. But clearly

$$\hat{M}_N = \sum_{s=0}^{N} T^{-1} \sum_{t=1+s}^{T} Z^k(t)'Z^k(t-s)\hat{R}_u(s) + \sum_{s=-N}^{1} T^{-1} \sum_{t=1}^{T+s} Z^k(t)'Z^k(t-s)\hat{R}_u(s)$$

converges in probability to M_N, by ergodicity and the consistency of $\hat{R}_u$. So the proof is complete if we can show that

$$\left| \sum_{s=N+1}^{\infty} T^{-1} \sum_{t=1+s}^{T} Z^k(t)Z^k(t-s)\hat{R}_u(s) \right|$$

can be made arbitrarily small with arbitrarily high probability for large T if N is chosen large enough. But (iv) and (vii) guarantee that $|\hat{R}_u(s)|\, g(s)$ is bounded in probability uniformly in s. Then since

$\sum_{s=N+1}^{\infty} T^{-1} \sum_{t=1+s}^{T} |\, Z^k(t)'Z^k(t-s)\,|\, g(s)^{-1}$ goes to zero in probability uniformly in T as N goes to infinity, the argument is complete.

From this point on, the proof consists only of verifying Hansen's Theorem 4.1 assumptions, almost exactly as in Theorem 1. In this case Hansen's function f is $Z^k(t)'u(t)$ and the weighting matrix a_N is $X'Z^k(Z^k,\hat{V}Z^k)^{-1}$.

Finally, we consider the limiting behavior of the asymptotic covariance matrices of the two estimators when k, the number of lagged values of the instruments used, is allowed to increase without limit. It should be noted that Hansen has (in unpublished work reported to us by letter) already displayed the limiting form as the covariance matrix of an ideal estimator. Our contribution here is only to show that both of our two proposed methods approach the limit arbitrarily closely.

Theorem 3:

If Σ_{FF} is the asymptotic covariance matrix of $\hat{b}_{FF}$ and Σ_{FOE} is the asymptotic covariance matrix of $\hat{b}_{FOE}$ as given in Theorems 1 and 2 respectively, then Σ_{FF} and Σ_{FOE} converge to the same limit as k goes to infinity.

Proof: For two vector stochastic processes x and y we will use $R_{xy}(t)$ to mean $E[X(s)'y(t-s)]$. (The processes are treated as row vectors.) Then we can write

$$\Sigma_{FOE}^{-1} = \sum_{v=0}^{k} \sum_{w=0}^{k} R_{XZ}(v)\, Q^{vw}\, [R_{XZ}(w)]' ,$$

where Q^{vw} is the v,w'th block of Q^{-1} and Q has v,w'th block $Q_{vw} = \sum_{=-\infty}^{\infty} R_u(s) R_{ZZ}(s-v+w)$. Now let $\bar{X} = a'^{*}X$ and $\bar{Z} = a^{-1*}Z$. Clearly, $R_u = a^{-1*}(a^{-1})'\sigma_e^2$ and $R_{XZ} = R_{\bar{X}\bar{Z}}$. Now we can write

$$\Sigma_{FOE}^{-1} = \sigma_e^2 \sum_{v=0}^{k} \sum_{w=0}^{k} R_{\bar{X}\bar{Z}}(v)\, \bar{Q}^{vw}\, R_{\bar{X}\bar{Z}}(w)' .$$

where $\bar{Q}_{vw} = R_{\bar{Z}\bar{Z}}(w-v)$. But this last expression is recognizable to anyone familiar with least-squares formulas as σ_e^2 times the covariance matrix of the projection of $\bar{X}(t)$ on $\bar{Z}(t-s)$, $s=0, \ldots,k$ in the covariance inner product.

Turning our attention now to $\hat{b}_{FF}$, we can write

$$\Sigma_{FF}^{-1} = \sigma_e^2 \sum_{v=0}^{k} \sum_{w=0}^{k} R_{\bar{X}Z}(v) \, \bar{\bar{Q}}^{vw} \, R_{\bar{X}Z}(w)' \quad ,$$

where $\bar{\bar{Q}}$ is a matrix with v,w'th block given by

$$\bar{\bar{Q}}_{vw} = R_{ZZ}(w-v) \ .$$

This formula is clearly in the same form as the one we found for $\hat{b}_{FOE}$, again being σ_e^2 times the covariance matrix of a projection of $\bar{X}(t)$, but this time on the space spanned by $Z(t-s)$, $s=0,\ldots,k$. The two matrices will thus increase to the same limit if $H_Z^-(t)$, completion of the space spanned by $Z(s)$, all $s \leq t$, is the same as $H_{\bar{Z}}^-(t)$, the completion of the space spanned by $\bar{Z}(s)$, all $s \leq t$. That this is so is immediately apparent from the fact that for each s $\bar{Z}(s) = a^{-1} * Z(s)$ lies in $H_Z^-(s)$ and $Z(s) = a * \bar{Z}(s)$ lies in $H_{\bar{Z}}^-(s)$. Thus the proof is complete.

REFERENCES

Brown, Bryan W. and Shlomo Maital (1979), "What do Economists Know?: An Empirical Study of Experts' Expectations", to appear, Econometrica.

Cumby, Robert E., John Huizinga, and Maurice Obstfeld (1980), "Two-Step Two-Stage Least Squares Estimation in Models with Rational Expectations", processed.

Geweke, John and Richard Meese (1979), "Estimating Regression Models of Finite but Unknown Order", processed, University of Wisconsin.

Hansen, Lars (1980), "Large Sample Properties of Generalized Method of Moments Estimators", processed.

Hansen, Lars and Thomas J. Sargent (1980), "Formulating and Estimating Dynamic Rational Expectations Models", Journal of Economic Dynamics and Control, 2, 7-46.

Hayashi, Fumio (1980), Utility Maximization, Rational Expectations, and Aggregate Consumption , Ph.D. thesis, Department of Economics, Harvard University.

Kennan, John (1979), "The Estimation of Partial Adjustment Models with Rational Expectations", Econometrica, 47, 144-56.

Rozanov, Yu. A. (1966), Stationary Random Processes , Holden-Day.