# ON THE BEHAVIOR OF INCONSISTENT INSTRUMENTAL VARIABLE ESTIMATORS*

## Esfandiar MAASOUMI

*University of Iowa, Iowa City, IA 52242, USA*

## Peter C.B. PHILLIPS

*Yale University, New Haven, CT 06520, USA*

Some theoretical results published recently in Hendry (1979) for the limiting distribution of inconsistent instrumental variable estimators in misspecified dynamic systems are incorrect. In particular, the derivations there involve the use of an inappropriate control variate and lead to an expression for the covariance matrix of the limiting distribution which, in general, omits many additional terms. Correct formulae are given in the present paper. Further, the accuracy of the asymptotic distribution in finite samples is investigated in a simple case using the known exact small sample distribution. On the basis of our exact results and in view of other strong theoretical and practical considerations, we argue for caution in the use of response surface regressions of the type recommended by Hendry in Monte Carlo experiments; and we emphasize the need for qualifying statements concerning the parameter environments in which the adequacy of these regressions has been substantiated.

## 1. Introduction

The recent article by Hendry (1979) in this Journal contains some theoretical results and makes certain methodological recommendations

concerning the use of control variables (CV's). It is argued that CV's can be used as an explicit analytic tool to extract limiting distributions and to derive expressions for their second moments. It is further argued, as in the earlier article by Hendry and Harrison (1974), that CV's serve as a numerical lever in experimental studies to improve the efficiency of the simulations.

Hendry (1979) is concerned with the behavior of inconsistent instrumental variables (IV) estimators in a general linear dynamic model. CV's are used to find the asymptotic distributions of the estimators, to assist in determining by simulation the accuracy of the asymptotic (second) moments in finite power and to test the predictive accuracy of asymptotic theory in finite samples. It was concluded that:

> The theoretical analysis establishes, and the simulation findings confirm, that the asymptotic results provide an extremely good guide to the finite sample behaviour of the second moments of the estimators investigated.   [Hendry (1979, pp. 308–309)]

Unfortunately, the general formula obtained in that article for the covariance matrix of the limiting distribution of the IV estimator and employed in its response surface regressions and associated predictions is incorrect. The error arises because the convergence theorem of Cramér, cited in the paper, does not apply, leading to a control variate which does not, in general, have the same limiting distribution as the econometric estimator under consideration. Since the simulation results in Hendry (1979) employ the incorrect variance formula, they do not provide valid general evidence to be employed in the general context intended in that paper. The simulation estimates of the sampling variances in these experiments do appear to be close to the variance of a generally inappropriate CV but the value and interpretation of this information is open to question. Moreover, as we will discuss, the asymptotic formulae in Hendry (1979) do turn out to be correct in certain specialised cases. We emphasize this point both for its relevance in such special cases in practice and because such cases influence the overall and summary analysis of simulation experiments. However, the precise parameter values used in the reported simulation experiments are not stated in Hendry (1979) and are, in fact, based on random draws from a finite population of arbitrary, selected values that were given in the closely related article by Hendry and Harrison (1974, p. 166). Lacking such information, it is difficult for a reader to determine the share of special cases that occur in these experiments and this accentuates the risks associated with extrapolations from specialised simulation evidence.

In section 2 we consider the source of the error in Hendry's derivation and in section 3 give a correct general formula for the covariance matrix of the limiting distribution of the IV estimator. As usual, the derivations leading to the limiting distribution suggest a control variate which does have the same

limiting distribution and which could be used for finite sample approximations or in numerical simulations if these were considered to be of value. Section 4 details some numerical results which compare the asymptotic and exact small sample distributions of an inconsistent IV estimator in a simple case. These results endorse the strong evidence which has appeared elsewhere in the literature [for example, in Phillips (1977b) and in Evans and Savin (1980)] that the discrepancies between asymptotic and finite sample behavior are parameter dependent and do not permit such general conclusions as those in the above quotation. Section 5 is concerned with what appear to us to be important limitations in the practical use of response surface regressions of the type discussed in Hendry (1979) and Hendry and Harrison (1974). We demonstrate that more care is needed in specification and testing of such regression devices and that there are limitations to their usefulness as inferential tools.

Summary and conclusions appear in section 6. The same section points out that there remain parts of Hendry (1979) which are valuable for further research. In particular, we applaud a general goal that has been pursued by Hendry in the past. This goal is to improve upon crude simulation experimentation by means of analytic tools and to improve upon the analyses of experiments through the greater use of theoretical results and better design of experiments. One of the points that will become evident in the present paper is that the 'capital intensive' approach, generally favoured by Hendry and strongly advocated by him in seemingly analytically intractable situations, may not offer quite the level of 'labour saving' which is desired. Certainly, the reader of Hendry (1979) could go away expecting ·much from a capital intensive approach supplemented with a little input of labour by way of analytic derivations. It is clear, at least to us, that *good* capital intensive research along the lines suggested in Hendry (1979) is a great deal more labour intensive than many an investigator may be led to expect. More so and somewhat ironically, when the capital intensive approach is seemingly the most called for. That is, when analytical results are hard to obtain.

From Hendry (1981) it appears that some legitimate differences of opinion remain between us. However, we do agree on the necessity for correcting the error in Hendry (1979) and on the necessity for the cautions we have called for here.

## 2. The model and derivations

We will work with the same model and notation as Hendry (1979), using $(\ )_H$ to indicate equation numbers in that article when we need to reference them. The observable variables, collected into the $(2n+m) \times 1$ vector $f_t$ and

representing $n$ current period endogenous, $n$ lagged endogenous and $m$ exogenous variables are generated by the system [see $(5)_H$]

$$f_t = Df_{t-1} + w_t, \qquad t = 1, \ldots, T, \tag{1}$$

which is assumed to be stationary and where the $w_t$ are independent $N(0, \Sigma_w)$. $(1)_H$–$(3)_H$ detail the structural system giving rise to (1). The particular equation under study is written [see $(6)_H$] as

$$y = X\beta + u, \tag{2}$$

where the vector $u$ is multivariate $N(0, \Delta)$ and the $T \times 1$ vector $y$ and $T \times k$ matrix $X$ are composed of observations on a subset of the variables $f_t$ of (1). An instrument matrix $Z$ of $T$ observations on $\kappa$ instrumental variables is introduced, which are also taken from the set of variables in (1). Inconsistent estimators result since at least some of the instruments are correlated with $u$.

Hendry considers the following IV estimator of the parameter vector $\beta$ in eq. (2)

$$\tilde{\beta} = (X'NX)^{-1}X'Ny, \tag{3}$$

[see $(8)_H$] where $N = Z(Z'Z)^{-1}Z'$. In view of the stationarity and distributional assumptions of (1), we have

$$A = \plim_{T \to \infty} \left\{ \left(\frac{X'Z}{T}\right)\left(\frac{Z'Z}{T}\right)^{-1} \right\} = E\left(\frac{X'Z}{T}\right)\left(E\left(\frac{Z'Z}{T}\right)\right)^{-1}, \tag{4}$$

$$G = \plim_{T \to \infty} \left(\frac{Z'Z}{T}\right) = E\left(\frac{Z'Z}{T}\right), \tag{5}$$

$$K = \plim_{T \to \infty} \left(\frac{X'Z}{T}\right)\left(\frac{Z'Z}{T}\right)^{-1}\left(\frac{Z'X}{T}\right) = AGA', \tag{6}$$

$$\alpha = \plim_{T \to \infty} \left(\frac{Z'u}{T}\right) = E\left(\frac{Z'u}{T}\right), \tag{7}$$

[see $(9)_H$, $(12)_H$ and following paragraph]. We deduce

$$\plim_{T \to \infty} \tilde{\beta} = \beta + K^{-1}A\alpha = \beta + p = \beta_l, \quad \text{say}, \tag{8}$$

[see $(13)_H$ but we are omitting the tilde on $\beta_l$], so that $p$ is the vector of inconsistencies in the estimator $\tilde{\beta}$.

To find the limiting distribution of an appropriately centred statistic based on the estimator $\tilde{\beta}$, Hendry follows the usual procedure of writing

$$\tilde{\beta} - \beta_I = \tilde{p} - p = (X'NX)^{-1}X'Ne, \tag{9}$$

where $\tilde{p}$ is the pseudo-estimator of $p$ given by

$$\tilde{p} = (X'NX)^{-1}X'Nu, \tag{10}$$

and where $e = u - Xp = y - X\beta_I$. It follows from the definition of $p$ that

$$\plim_{T \to \infty} T^{-1}X'Ne = A \plim_{T \to \infty}\left(\frac{Z'e}{T}\right) = A\,\mathrm{E}\left(\frac{Z'e}{T}\right) = 0. \tag{11}$$

The argument now hinges on the use of the simplified statistic [see $(16)_\mathrm{H}$],

$$p^* = p + T^{-1}K^{-1}AZ'e, \tag{12}$$

as a control variate for $\tilde{\beta}$. In fact, this is an inappropriate choice of control variate and, contrary to the assertion on page 301, the limiting distribution of $\sqrt{T}(\tilde{\beta} - \beta_I)$ cannot in general be derived by obtaining that of $\sqrt{T}(p^* - p)$. We will, in fact, derive an appropriate control variate for $\tilde{\beta}$ below.

The difficulty in the above treatment can be seen explicitly in the final paragraph of page 300. In this paragraph, the convergence theorem of Cramér (1946, p. 254) is applied to derive the limiting distribution of $\sqrt{T}(\tilde{p} - p)$ and, hence, $\sqrt{T}(\tilde{\beta} - B_I)$. Unfortunately, the conditions under which Cramér's theorem is valid do not generally hold in this particular application and the limiting distribution that is consequently obtained for $\sqrt{T}(\tilde{\beta} - \beta_I)$ is incorrect. Specifically, the error occurs in the last four lines of page 300 where it is argued that

... since

$$\plim_{T \to \infty}\left(\frac{X'NX}{T}\right)^{-1}\left(\frac{X'Z}{T}\right)\left(\frac{Z'Z}{T}\right)^{-1} = K^{-1}A, \tag{13}$$

then $\tilde{p}$ and $p^*$ have the same limiting distributions in that $\tilde{p} = p^* + \mathrm{O}(T^{-1})$ and $\plim T^{1/2}(\tilde{p} - p^*) = 0$. [Hendry (1979, p. 300)]

In fact, it is not generally true that $\plim_{T \to \infty} T^{1/2}(\tilde{p} - p^*) = 0$ as is claimed.

To see the error in this reasoning we write

$$\sqrt{T}(\tilde{p}-p)=K^{-1}A\frac{Z'e}{\sqrt{T}}+\left[\left(\frac{X'NX}{T}\right)^{-1}\left(\frac{X'Z}{T}\right)\left(\frac{Z'Z}{T}\right)^{-1}\right.$$

$$\left.-K^{-1}A\right]\frac{Z'e}{\sqrt{T}}$$

$$=\sqrt{T}(p^{*}-p)+\left[\left(\frac{X'NX}{T}\right)^{-1}\left(\frac{X'Z}{T}\right)\left(\frac{Z'Z}{T}\right)^{-1}\right.$$

$$\left.-K^{-1}A\right]\frac{Z'e}{\sqrt{T}}.\tag{14}$$

The argument would hold, in view of (13), if $T^{-1/2}Z'e$ had a limiting distribution. For, in this case, the second term of (14) tends in probability to zero [see, for instance, Proposition 4 on p. 370 of Malinvaud (1970)] so that $\sqrt{T}(\tilde{p}-p)$ and $\sqrt{T}(p^{*}-p)$ have the same limiting distribution. But, as is clear from the definition of $A$, $G$, and $P$,

$$\mathrm{E}(Z'e)=\mathrm{E}\{Z'(u-Xp)\}=T\alpha-TGA'p=T(I-GA'K^{-1}A)\alpha,\tag{15}$$

[see (18)$_{\mathrm{H}}$] so that $T^{-1/2}Z'e$ has, in general, a non-zero mean which tends to infinity with $T$. Therefore, we cannot, in general, assert that $T^{-1/2}Z'e$ has a limiting distribution as $T\to\infty$. For this reason, the second term of (14) will not generally vanish as $T\to\infty$ and it will not generally follow that $\sqrt{T}(\tilde{p}-p)$ and $\sqrt{T}(p^{*}-p)$ have the same limiting distribution.

Only in certain special cases will the argument and the stated results be valid. Some examples where this is so are: (i) when the set of instruments includes only truly exogenous variables; (ii) when the number of instruments is identical to the number of coefficients in the equation to be estimated; and (iii) when the set of instruments includes all regressors in the equation to be estimated together with a set of truly exogenous variables that are uncorrelated with these regressors and that are stationary with zero means. In the first case, of course, $\mathrm{E}(Z'u)=0$, the IV estimator is consistent, $p=0$ and $\mathrm{E}(Z'e)=0$ (an inadmissable special case in the study of inconsistent IV estimators); while in the second case, the fact that $\mathrm{E}(AZ'e)=0$ will ensure that $\mathrm{E}(Z'e)=0$ provided $\mathrm{E}(X'Z)$ (and, hence, the matrix $A$) is of full rank (this is already an implicit assumption in Hendry's article for the existence of the IV estimator); in the third case, we may partition the instrument matrix as $Z=[X\vdots Z_{b}]$ and it now follows from the definition of $e$ that $\mathrm{E}(X'e)=0$; and

$E(Z_b'e) = 0$ because the variates in $Z_b$ are truly exogenous and have zero means. It is worth noting in this last case that the zero mean assumption implicit in the general set up [see $(1)_H$–$(5)_H$] of the model (1) is an important simplification. For, this assumption will sometimes reduce $E(Z'e)$ to zero in cases where this would not normally be so, as indeed in the special case just considered. In such special cases, $\sqrt{T}(\tilde{p}-p)$ will have the same limiting distribution as $\sqrt{T}(p^*-p)$.

In general, however, when the number of instruments is greater than the number of regressors we will find that the expression for $E(Z'e)$ in (15) is non-zero and the limiting distribution of $\sqrt{T}(\tilde{p}-p)$ will depend on terms involving the elements of $\sqrt{T}\{T^{-1}X'Z - E(T^{-1}X'Z)\}$ and $\sqrt{T}\{T^{-1}Z'Z - E(T^{-1}Z'Z)\}$ as well as the term $\sqrt{T}(p^*-p)$. Specifically, we need to decompose (14) further as[1]

$$\sqrt{T}(\tilde{p}-p) = \sqrt{T}(p^*-p)$$

$$+\sqrt{T}\left\{\left(\frac{X'NX}{T}\right)^{-1} - K^{-1}\right\}\left\{\left(\frac{X'Z}{T}\right)\left(\frac{Z'Z}{T}\right)^{-1}\left(\frac{Z'e}{T}\right)\right\}$$

$$+K^{-1}\sqrt{T}\left\{\left(\frac{X'Z}{T}\right) - \plim_{T\to\infty}\left(\frac{X'Z}{T}\right)\right\}\left(\frac{Z'Z}{T}\right)^{-1}\left(\frac{Z'e}{T}\right)$$

$$+K^{-1}\plim_{T\to\infty}\left(\frac{X'Z}{T}\right)\sqrt{T}\left\{\left(\frac{Z'Z}{T}\right)^{-1}\right.$$

$$\left. - \plim_{T\to\infty}\left(\frac{Z'Z}{T}\right)^{-1}\right\}\left(\frac{Z'e}{T}\right). \tag{16}$$

Now the second term on the right side of (16) tends in probability to zero, since the first factor has limiting distribution, while the second factor has a limit in probability of zero because

$$\plim_{T\to\infty}\{(T^{-1}X'Z)(T^{-1}Z'Z)^{-1}(T^{-1}Z'e)\} = A\plim_{T\to\infty}(T^{-1}Z'e)$$

$$= A\,E(T^{-1}Z'e) = 0.$$

On the other hand, since $T^{-1}Z'e \to \alpha - GA'p$ in probability, which is in general a non-zero vector, there is no reason why the third and fourth terms

---

[1]This decomposition follows the usual lines in extracting the limiting distribution of inconsistent estimators. See, for example, the derivations and discussions in Phillips and Wickens (1978, problem 6.10).

on the right side of (16) should vanish as $T \to \infty$. These terms will then also contribute to the limiting distribution of $\sqrt{T}(\tilde{p} - p)$ as an examination of a simple example will show.[2] Thus, the formulae derived and stated in Hendry (1979) for the covariance matrix of the limiting distribution of $\sqrt{T}(\tilde{p} - p)$ and, hence, $\sqrt{T}(\tilde{\beta} - \beta_l)$, are generally invalid.

## 3. The limiting distribution of $\sqrt{T}(\tilde{\beta} - \beta_l)$

It follows from (16) and the non-singularity of $G = \text{plim}_{T \to \infty} T'Z'Z$ that the limiting distribution of $\sqrt{T}(\tilde{p} - p)$ is equivalent to that of the vector

$$\sqrt{T}(p^* - p) + K^{-1}\sqrt{T}\left\{T^{-1}X'Z - \text{plim}_{T \to \infty}(T^{-1}X'Z)\right\}G^{-1}\alpha_e$$

$$-K^{-1}A\sqrt{T}\{T^{-1}Z'Z - G\}G^{-1}\alpha_e, \tag{17}$$

where $\alpha_e = E(T^{-1}Z'e) = (I - GA'K^{-1}A)\alpha$. We now write $Z = FS_1$, $X = FS_2$ and $e = y - X\beta_l = [y \vdots X](1, -\beta_l')' = FS_3\psi = F\phi$, say, where $S_1$, $S_2$, and $S_3$ are appropriately dimensioned selector matrices and $\phi = S_3\psi$ with $\psi' = (1, -\beta_l')$. Using this notation, we can write the vector (17) in the form

$$K^{-1}AS_1'\sqrt{T}\{T^{-1}F'F - E(T^{-1}F'F)\}\phi$$

$$+K^{-1}S_2'\sqrt{T}\{T^{-1}F'F - E(T^{-1}F'F)\}S_1 G^{-1}\alpha_e$$

$$-K^{-1}AS_1'\sqrt{T}\{T^{-1}F'F - E(T^{-1}F'F)\}S_1 G^{-1}\alpha_e$$

$$= H\sqrt{T}\text{vec}\{T^{-1}F'F - E(T^{-1}F'F)\}, \tag{18}$$

where vec denotes vectorization by rows,

$$H = K^{-1}AS_1' \otimes \phi' + K^{-1}S_2' \otimes \alpha_e' G^{-1}S_1' - K^{-1}AS_1' \otimes \alpha_e' G^{-1}S_1', \tag{19}$$

and $\otimes$ is the right-hand Kronecker product.

---

[2]We may, for instance, take the following system where the first equation is to be estimated [and, therefore, corresponds to (2) above] and the last two variables are used as instruments

$$y_{1t} = b_{12}y_{2t} + u_{1t}, \quad y_{2t} = u_{2t}, \quad y_{3t} = u_{3t}, \quad y_{4t} = u_{4t}$$

If we write $m_{ij} = E(y_{it}y_{jt})$ and to simplify matters set $m_{14} = m_{34} = 0$, we find that $p = m_{13}m_{23}m_{44}\{m_{23}^2m_{44} + m_{24}^2m_{33}\}^{-1}$, and then

$$E(T^{-1}Z'e) = \begin{bmatrix} m_{13} - m_{23}p \\ -m_{24}p \end{bmatrix}.$$

Since $T^{-1}F'F = T^{-1}\sum_t f_t f_t'$ is the sample second moment matrix of the $f_t$ as generated by the stationary system (1), it follows that the vector (18) has the limiting normal distribution $N(0, C)$ where $C = H\Psi H'$ [see, for example, Hannan (1970, theorem 14, p. 228) and Hannan (1976) who proves a stronger result for $\Sigma_w$ non-singular]. Under the normality assumption for the errors driving (1), we have

$$\Psi = 2\pi \int_{-\pi}^{\pi} f(\omega) \otimes f(-\omega) \, d\omega (I + K_l), \tag{20}$$

where

$$f(\omega) = (2\pi)^{-1} \left( \sum_{j=0}^{\infty} D^j e^{ij\omega} \right) \Sigma_w \left( \sum_{j=0}^{\infty} D'^j e^{-ij\omega} \right)$$

is the spectral density matrix of $f_t$ and $K_l$ is the commutation matrix [Magnus and Neudecker (1979)] of dimension $l^2 \times l^2$ where $l = 2n + m$. In the case where the errors are not necessarily normally distributed, but fourth moments of the errors on (1) still exist, the expression (20) for $\Psi$ involves the additional term depending on the fourth cumulants of the errors which is given in eq. (3) of Hannan (1976).

The vector $H \text{ vec}(T^{-1}F'F)$ can be regarded as a control variate for $\tilde{\beta}$. Appropriately centered and standardized by $\sqrt{T}$ it has the same limiting distribution as $\sqrt{T}(\tilde{\beta} - \beta_l)$ and can be used for finite sample approximations or as a device to improve the precision of simulation estimates if these were required. The moments of such a control variate are best regarded as pseudo-moments of $\tilde{\beta}$ or, as has become more customary terminology in the econometrics literature, moments of an approximating distribution. It is also worth noting that the vector $H \text{ vec}(T^{-1}F'F)$ is only a special control variate in a whole class of such variates constructed as stochastic approximations to $\tilde{\beta}$ and based on simple polynomial approximations to $\tilde{\beta}$ in terms of the elements of the sample moment matrix $T^{-1}F'F$. The distribution of such control variates is then, to an appropriate order in $T^{-1/2}$, just the Edgeworth approximation. The framework developed in the papers by Sargan (1976) and Phillips (1977a) can, in fact, be used to develop a class of control variates for econometric estimators in quite general circumstances. It is also possible to use the framework in these latter articles to construct control variates in cases where the estimator under consideration is determined implicitly by a system of equations whose functions do depend explicitly on a vector of sample moments of the data. This approach then covers most known econometric estimators with the exception of some that are based on moving average error models.

## 4. Small sample and asymptotic comparisons

Using simulation evidence, Hendry also addresses the question of the relevance of asymptotic theory for inconsistent estimators in finite samples. The value of these simulations findings is now uncertain to us. For, although a careful scrutiny of the parameter environments and IV estimators chosen for the experiments will no doubt reveal for which experiments Hendry's asymptotic variance formula is correct,[3,4] these experiments will nonetheless be special cases of a general asymptotic theory in which the asymptotic covariance matrix involves many more terms than those used in the present computations. The risks associated with extrapolations from specialised simulation evidence are, therefore, more acute than usual. Moreover, since the asymptotic formulae in Hendry (1979) are, in general, wrong we would expect to find some evidence of this in terms of a systematic discrepancy between the simulation results for large sample size experiments and the stated asymptotics. From Hendry's discussion, we judge this not to be the case, although we do *not* know from the reported details of the simulations how many large $T$ experiments were actually conducted.

Some complementary evidence is available from exact small sample distribution theory in certain simple cases. We take, for instance, the following model:[5]

$$y_{1t} = b_{12} y_{2t} + u_{1t}, \tag{21a}$$

$$y_{2t} = b_{21} + u_{2t}, \qquad t = 1, \ldots, T, \tag{21b}$$

in which the $y_{it}$ are endogenous variables and the $u_{it}$ are independent identically distributed normal variates for all $t$. Ordinary least squares (OLS) applied to (21a) is generally inconsistent and comes within the framework of estimators being investigated by Hendry. We assume that the usual standardising transformation which reduces the covariance matrix of the $y_{it}$ to the identity matrix has been carried out. Then, the exact density function and first two moments of this estimator can be readily deduced from the results of Richardson and Wu (1970, 1971).[6] If we let $\hat{b}_{12}$ be the OLS

---

[3]For example, using the arguments given in section 3 in the paragraph following (15), we may deduce that ordinary least squares is such a special case.

[4]Hendry (1979) does not report the parameter values used in the experiments, only that they are based on the 40 experiment set III of Hendry and Harrison (1974), where a random selection was chosen within a finite population of possible parameter values.

[5]We recognize that this model involves variables with a non-zero mean. As we discussed earlier in section 2, the zero mean assumption can have important consequences in terms of simplifying the asymptotic covariance matrix of the IV estimator. It seems appropriate to avoid working within a framework that leads to specialized results when the objective is to extract usable general formulae.

[6]Takeuchi (1970) also gives analytic expressions for these moments. However, there appear to be errors in his expressions arising out of his formulae (2-7) and (2-8) so we have not used them here. In particular (2-7) and (2-8) confuse the even and odd order moments.

estimator of $b_{12}$ in (21a) we obtain in our notation

$$\mathrm{pdf}(\hat{b}_{12}) = \frac{\exp\left\{-\frac{\mu^2}{2}(1+b_{12}^2)\right\}}{B\left(\frac{1}{2},\frac{T}{2}\right)(1+\hat{b}_{12}^2)^{(T+1)/2}}$$

$$\times \sum_{j=0}^{\infty} \frac{\left(\frac{T+1}{2}\right)_j}{\left(\frac{T}{2}\right)_j j!}\left\{\frac{\mu^2}{2}\frac{(1+b_{12}\hat{b}_{12})^2}{1+\hat{b}_{12}^2}\right\}^j$$

$$\times {}_1F_1\left(\frac{T-1}{2},j+\frac{T}{2};\frac{T}{2}\frac{\mu^2 b_{12}^2}{2}\right), \tag{22}$$

with

$$\mu^2 = Tb_{12}^2, \qquad (a)_j = \Gamma(a+j)/\Gamma(a).$$

Then

$$\mathrm{E}(\hat{b}_{12}) = b_{12}b_{12}^2 e^{-Tb_{21}^2/2}\,{}_1F_1\left(\frac{T}{2},\frac{T+2}{2};\frac{Tb_{21}^2}{2}\right) = a_1(T,b_{12},b_{21}), \quad \text{say,}$$

$$\mathrm{var}(\hat{b}_{12}) = \frac{1}{2}e^{-Tb_{21}^2/2}\left\{\frac{2}{T-2}\,{}_1F_1\left(\frac{T}{2}-1,\frac{T}{2};\frac{Tb_{21}^2}{2}\right)\right.$$

$$+\frac{2b_{12}^2 b_{21}^2}{T-2}\,{}_1F_1\left(\frac{T}{2}-1,\frac{T}{2}+1;\frac{Tb_{21}^2}{2}\right)$$

$$\left.+\frac{2Tb_{12}^2 b_{21}^4}{T+2}\,{}_1F_1\left(\frac{T}{2},\frac{T}{2}+2;\frac{Tb_{21}^2}{2}\right)\right\} - \{a_1(T,b_{12},b_{21})\}^2$$

$$= a_2(T,b_{12},b_{21}), \quad \text{say.} \tag{23}$$

The probability limit of $\hat{b}_{12}$ and its asymptotic variance can now be found by

applying the asymptotic expansion of the confluent hypergeomctric function $_1F_1(\ ,\ ;\ )$ as $T \to \infty$. We find

$$\lim_{T \to \infty} a_1(T, b_{12}, b_{21}) = \frac{b_{12}b_{21}^2}{1 + b_{21}^2} = a_3(b_{12}, b_{21}), \quad \text{say},$$

$$\lim_{T \to \infty} Ta_2(T, b_{12}, b_{21}) = \frac{1}{1 + b_{21}^2} + \frac{b_{12}^2 b_{21}^2 (1 + b_{21}^4)}{(1 + b_{21}^2)^4}$$

$$= a_4(b_{12}, b_{21}), \quad \text{say}. \tag{24}$$

The asymptotic normal approximation to the pdf of $\hat{b}_{12}$ is then given by

$$\text{aspdf}(\hat{b}_{12}) = \left(\frac{T}{2\pi a_4}\right)^{1/2} \exp\left\{-\frac{T}{2}\frac{(\hat{b}_{12} - a_3)^2}{a_4}\right\}. \tag{25}$$
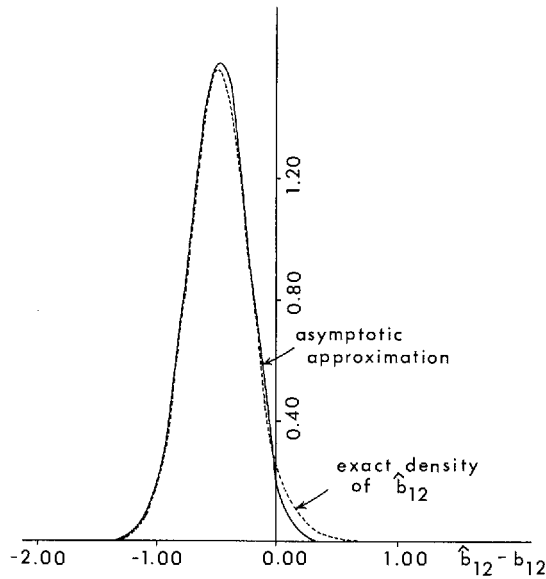


Fig. 1. $T = 10$, $b_{12} = 1.0$, $b_{21} = 1.0$.

In figs. 1 and 2 we have graphed the asymptotic density (25) against the exact density (22). These graphs show, as we might expect, that the adequacy of the asymptotic approximation is parameter dependent. In particular, the approximation (25) deteriorates as the value of $b_{12}$ increases. In fig. 2, where
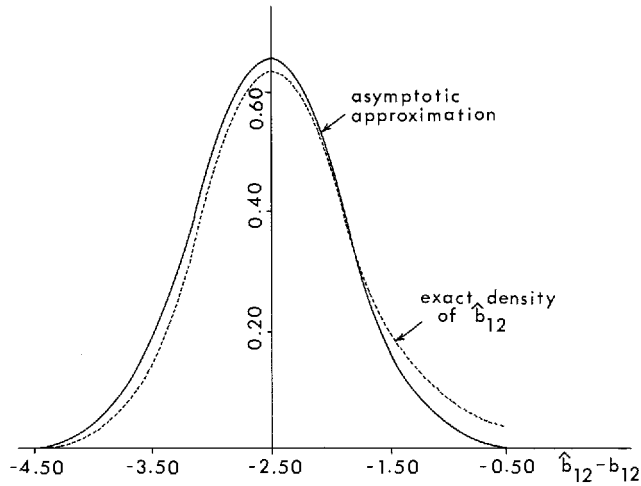
Fig. 2. $T=10$, $b_{12}=5.0$, $b_{21}=1.0$.

$T=10$, $b_{12}=5.0$ and $b_{21}=1.0$, we find for the variances that $a_2(T,b_{12},b_{21})$ $=0.5012$ and $T^{-1}a_4(b_{12},b_{21})=0.3625$; so, in this case, the exact variance is 38% larger than the asymptotic. In fig. 1, where $b_{12}=1.0$ and the other parameters are unchanged, the corresponding percentage difference between the variances is 23%.
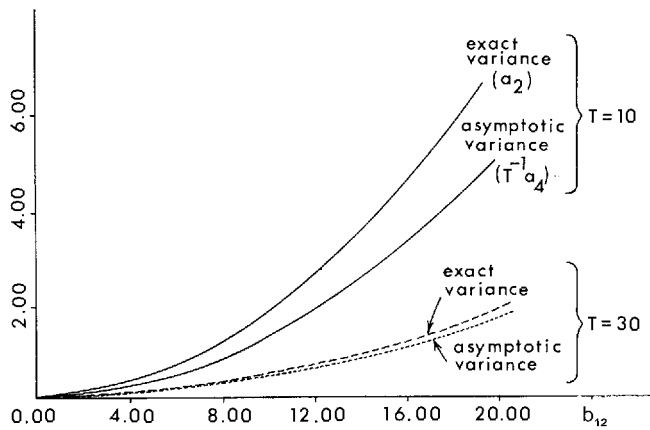


Fig. 3. Asymptotic and exact small sample variances.

Both $a_2(T,b_{12},b_{21})$ and $a_4(b_{12},b_{21})$ are quadratic in $b_{12}$, so that the divergence between the exact and asymptotic variances continues to increase with $b_{12}$. Fig. 3 shows the extent of this divergence between the two variances for $b_{12}$ over the domain $0<b_{12}<20.00$ and for $T=10,30$.

These results indicate that care should be exercized in evaluating the adequacy of asymptotic theory. Strong assertions about the accuracy and usefulness of asymptotic results in finite samples are best avoided unless they are conditioned by a clear statement concerning the parameter environment to which the results refer. As the dimension of the parameter space increases, we recognize that this becomes more difficult; but this should sharpen, not lessen, the need to express caution about the limitations of the results.

## 5. Response surface regressions relating finite sample outcomes to asymptotic results

Particular care seems to be necessary in the use of the response surface regressions reported in Hendry (1979) and the earlier article by Hendry and Harrison (1974). While these regressions do summarize a large body of simulation evidence, it is clear from the discussion in these articles that the regressions are being used for inferential purposes and to widen the scope of the simulation studies as the quotations below attest. For example, Hendry (1979) uses response surface regressions of the form

$$\ln S_j = \gamma_1 \ln V_j + \gamma_2 T^{-1} + v_j, \qquad j = 1, \ldots, 36, \tag{26}$$

[see (30)$_H$] where $S_j$ denotes the finite sample simulation estimate of a certain quantity (such as the mean value of an estimated error variance or the finite sample variance of an IV estimator) in experiment $j$ and $V_j$ is its asymptotic equivalent. It is then suggested that:

> $\gamma_1 = 1$ and $\gamma_2 = 0$ provide interesting testable hypotheses about the predictive accuracy of asymptotic theory in finite samples. [Hendry (1979, p. 305)]

and, in discussing the first set of response surface regressions it is observed that

> These estimates strongly support the claim that asymptotic theory constitutes an excellent explanation of finite sample outcomes for $\sigma_u^2$; $\hat{\gamma}_1$ is not significantly different from unity (or $\hat{\gamma}_2$ from zero), $R^2$ is very high and the fitted equations predict reasonably to randomly chosen points outside the sample space. [Hendry (1979, p. 306)]

Moreover, in the earlier article by Hendry and Harrison (1974) in which the technique is more fully discussed we find the following observation made to motivate the use of the technique:

> The two most powerful methods of analysis of experimental data are regression and variance analysis. The former seemed attractive as we sought quantitative evaluation of the determinants of any observed biases formulated in *expressions which could predict outcomes beyond the*

*range of our experiments.* [Hendry and Harrison (1974, p. 162), our italics]

They conclude that:

... by relating the small sample outcome to the asymptotic result (which is known exactly) one can derive expressions which generalise beyond the sample of experiments conducted. [Hendry and Harrison (1974, p. 171)]

Response surface regressions of the type (26) are useful in summarizing simulation outcomes for the relevant statistics and can sensibly be used to examine the extent of the correspondence between finite sample outcomes and known asymptotic results within the experiments conducted. But, in our view, extreme care should be exercized in the use of such regressions to 'predict outcomes beyond the range of the simulation experiments'. This is because it is inevitable that regressions such as (26) will be misspecified.[7] The finite sample statistics being estimated by the mean simulation outcomes $S_j$ will normally be highly complex functions of the underlying parameters, as are $a_1(T, b_{12}, b_{21})$ and $a_2(T, b_{12}, b_{21})$ in our simple example above. The asymptotic equivalents $V_j$ also normally depend on a similar set of parameters, as do $a_3(b_{12}, b_{21})$ and $a_4(b_{12}, b_{21})$ above. Just as it is quite incorrect to argue that $a_1$ and $a_3$ or $a_2$ and $T^{-1}a_4$ satisfy (26) exactly with $v_j = 0$ (this would, in fact, impose a restriction on the free parameters $T$, $b_{12}$ and $b_{21}$), so we will find that, in practice, equations of the form (26) actually define the errors $v_j$. These errors then become heavily parameter dependent. The resulting equation once estimated is likely to display substantial residual autocorrelation and heteroscedasticity. This will be associated with the way in which the true errors $v_j$ change in response to the change in the parameter environment indexed by $j$. Since the form of this dependence is, of course, unknown to the investigator running the simulation experiment, inferences and predictions based on (26) become extremely hazardous. In other words, (26) is a spurious regression in the terminology of Granger and Newbold (1974).

To illustrate the typical misspecification of (26) we use the analytically determined finite sample and asymptotic variance of the estimator $\hat{b}_{12}$ in the model (21). We set $S_j = a_2(T, b_{12}, b_{21})$ and $V_j = T^{-1}a_4(b_{12}, b_{21})$. Using values of $S_j$ and $V_j$ computed for 51 values of $b_{12}$ on an equispaced grid over the interval $0 \leq b_{12} \leq 5.0$ and fixed $T = 10$, $b_{21} = 1.0$, a regression of the form (26) produced the following results:[8]

$$\ln S_j = 1.0458 \ln V_j + 0.3535, \quad R^2 = 0.9997, \quad DW = 0.0252, \qquad (27)$$
$$\phantom{\ln S_j = } (0.0028) \phantom{\ln V_j +} (0.0044)$$

[7]We agree with the first sentence in the Introduction of Hendry (1979).

[8]The data for this regression can be readily computed from the analytic formulae given in section 4 of this article.

where coefficient standard errors are shown in parentheses. The residuals in (27) are, in fact, highly autocorrelated as implied by the low value of the Durbin–Watson statistic $(DW)$ and confirmed by a residual plot. Yet the sample fit as measured by $R^2$ is consummately high.

Should we or should we not use the regression (27) to test the predictive accuracy of asymptotic theory and extrapolate beyond the range of our computations or experiments, as in Hendry (1979) and Hendry and Harrison (1974)? In our view, the answer to each of these questions is: no, not unless the adequacy of the regression has been carefully diagnosed beforehand and not unless there are clear qualifying statements concerning the parameter environment in which the regression has been shown to be relevant. With respect to (27), there is strong evidence to suggest that this equation is badly misspecified. Thus, while the linear relation (27) may well suffice to describe the pattern of behavior in $\ln S_j$ and $\ln V_j$ within the domain of variation of the parameters considered (in this case $b_{12}$) there is danger of a serious mistake being made if the relation (27) is used outside this domain. Moreover, in the presence of this misspecification, the standard errors of the coefficients given in (27) are no longer relevant and the usual significance tests on the coefficients are invalid. Nor are these problems removed by randomizing the sample points taken in the parameter space. For, the linear expression (27) will, nevertheless, adequately represent the behavior of $\ln S_j$ and $\ln V_j$ only within the domain allowed by the sample. The misspecification in (27) will still be present since the true errors that occur in (26) are functionally dependent on the parameters; and this misspecification will be detected in the usual way by rearranging the sequence of residuals according to increasing values of the parameters before testing the random character of the errors. Nor, also, are these problems affected by the use of simulation data for $S_j$ in (26). For, in this case, the errors may be regarded as being composed of the sum of two elements, the first of which is functionally dependent on the parameters, the second of which is a random variable arising from the simulation errors.

These arguments suggest that more tests of specification and more diagnostic checks need to be conducted than are presently reported or encouraged in Hendry (1979) and Hendry and Harrison (1974), at least before response surface regressions such as (26) are set up for inferential use. As in the above example, $DW$ statistics can provide important diagnostic information if correctly applied. In more complicated cases than the above, these can be computed from regressions on suitable subsets of the data naturally ordered according to the changes in a single parameter. For this to be possible, the finite population parameter environment that underlies the simulation study should be carefully designed to ensure adequate sampling of points in each of the major directions of known parameter variability. Low $DW$ values then provide early warning signals against strong inferential

statements. Use of the *DW* statistic in this way is really only a minimum requirement and other standard diagnostic procedures, if appropriately applied, will be helpful and desirable in the evaluation of the regression. As indicated above, the overall design of the experiment will play an important role in the successful application of such test diagnostics.

For these reasons, we recommend caution if response surface regressions which relate finite sample outcomes to asymptotic results are to be used in making strong inferential assertions about the predictive accuracy (or adequacy) of asymptotic theory. We emphasize the need for clear qualifying statements concerning the parameter environments in which their relevance has been established. And we suggest that the overall design of the experiment should take into account the data needs that will arise later in the careful diagnostic checking and testing of such response surface regressions that should rightly precede their use for inferential purposes.

## 6. Summary and conclusion

The general formula derived in Hendry (1979) for the asymptotic covariance matrix of the IV estimator in models where this estimator is inconsistent is incorrect. Moreover, the procedure used there to extract the limiting distribution of the estimator and to develop a control variate is flawed and leads to a generally inappropriate control variate. We have not determined the full extent to which Hendry's simulation findings (which use both the asymptotic results and the control variate) are influenced by these errors. We have shown that in some specialized cases Hendry's formulae are correct and to the extent that these cases occur in his simulations (as, for example, they do for the OLS estimator) they are albeit more than usually open to the qualification that they are limited to a specific environment.

Our analysis of a simple inconsistent estimator for which the exact finite sample density and moments are known indicates that the asymptotic distribution does provide a good approximation to the distribution of this estimator in certain parameter environments but is less satisfactory in others. These results accord with those of other investigations of the adequacy of asymptotic theory in small samples for consistent estimators and test statistics [see, in particular, Phillips (1977b), Maasoumi (1977, ch. 5) and Evans and Savin (1980)].

It seems necessary to emphasize that care should be exercised in statements about the adequacy of asymptotic theory. In our view, strong assertions about the accuracy and usefulness of asymptotic results in finite sample situations are best avoided altogether. If they are made, they should be accompanied by a clear declaration of the parameter environment for which these assertions have been substantiated. Special care is necessary in the use of response surface regressions which relate finite sample simulation

outcomes to asymptotic results in Monte Carlo experimentation. In such cases, the experimental design should allow for the appropriate data that is necessary in the diagnostic checking and testing of such regressions that should properly precede their use for inferential purposes. The simulation findings and associated assertions in Hendry (1979) do not meet these requirements and should, therefore, be regarded with caution.

The recent paper Hendry (1981) has provided some qualifications in the light of our objections. These are welcome even though they do not go as far as we believe our arguments necessitate. Lack of a sufficient emphasis on the required qualifications and on the required labour input (by way of careful and correct analytic derivations) in improved capital intensive research will be most unfortunate. Indeed, it will be contrary to the apparent goal of Hendry and others which is to bring order and more theoretical information to bear upon experimental research. This is a goal that we wholeheartedly support.

## References

Cramér, H., 1946, Mathematical methods of statistics (Princeton University Press, Princeton, NJ).

Evans, G.B.A. and N.E. Savin, 1980, Testing for unit roots: I, Econometrica 49, 753–780.

Granger, C.W.J. and P. Newbold, 1974, Spurious regressions in econometrics, Journal of Econometrics 2, 111–120.

Hannan, E.J., 1970, Multiple time series (Wiley, New York).

Hannan, E.J., 1976, The asymptotic distribution of serial covariances, Annals of Statistics 4, 396–399.

Hendry, D.F., 1979, The behavior of inconsistent instrumental variable estimators in dynamic systems with autocorrelated errors, Journal of Econometrics 9, 295–314.

Hendry, D.F., 1980, Monte Carlo experimentation in economics, Invited talk at the Fourth World Congress of the Econometric Society, Aug./Sept.

Hendry, D.F., 1981, Monte Carlo experimentation in economics, in: Z. Griliches and M.D. Intriligator, eds., Handbook of econometrics, Vol. 2 (North-Holland, Amsterdam) forthcoming.

Hendry, D.F. and R.W. Harrison, 1974, Monte Carlo methodology and the small sample behavior of ordinary and two stage least squares, Journal of Econometrics 2, 151–174.

Maasoumi, E., 1977, A study of improved methods of estimating reduced form coefficients based on 3SLS estimators, Unpublished Ph.D. dissertation.

Magnus, J.R. and H. Neudecker, 1979, The commutation matrix: Some properties and applications, Annals of Statistics 7, 381–394.

Malinvaud, E., 1970, Statistical methods of econometrics (North-Holland, Amsterdam).

Phillips, P.C.B., 1977a, A general theorem in the theory of asymptotic expansions as approximations to the finite sample distributions of econometric estimators, Econometrica 45, 1517–1534.

Phillips, P.C.B., 1977b, Approximations to some finite sample distributions associated with a first order stochastic difference equation, Econometrica 45, 463–486.

Phillips, P.C.B. and M.R. Wickens, 1978, Exercises in econometrics (Philip Allan and Ballinger, Cambridge, MA).

Richardson, D.H., 1971, A note on the comparison of ordinary and two stage least squares, Econometrica 39, 973–982.

Richardson, D.H. and De Min Wu, 1970, Least squares and grouping method estimators in the errors in variables model, Journal of the American Statistical Association, 724–748.

Sargan, J.D., 1976, Econometric estimators and the Edgeworth approximation, Econometrica 44, 421–448.

Takeuchi, K., 1970, Exact sampling moments of ordinary least squares, instrumental variable and two stage least squares estimators, International Economic Review 11, 1–12.

## Postscript added in proof

In view of the clarification provided by Professor David Hendry in his Reply (hereafter DH) to our article, the following points may further reduce the erroneous usages of the formulae in Hendry (1979) and Hendry and Harrison (1974) (hereafter HH):

(i) Contrary to the assertions in HH (second paragraph, p. 159), and the impression created by DH (second paragraph), the ultimate formulae for the CV in HH (given by (32)$_{HH}$ in the text of HH) is *not valid* for the case of 2SLS and other IV estimators. This can be seen simply by observing that (32)$_{HH}$ requires that its constituent matrices $\Phi$ and $\bar{\Phi}$ be square and non-singular, which they patently are *not* in the general IV case;

(ii) The error that Professor Hendry modestly describes as 'bad algebra' in DH (second paragraph), arises naturally from the mistaken use of (32)$_{HH}$ in the general case. Specifically, since $\tilde{p}$ is a consistent estimator of $p$ and $\Phi\bar{\phi}$ measures inconsistency, the mistaken use of (32)$_{HH}$ involves setting $\bar{\Phi}\bar{\phi}=0$, and this leads mechanically to the inappropriate CV $\Phi\phi$. In the present case, this is just $p^*-p$, the CV we have shown to be invalid by other means in section 2 of our article;

(iii) With respect to the argument that, 'amusingly, $p^*$ is a valid, if inefficient, CV for $\tilde{\beta}$' (DH, third paragraph) the reader should note the clear counterarguments to the use of CV's in HH (p. 158, second paragraph). In fact, HH (footnote 4, p. 158) report experimental evidence which, amusingly, confirms that such CV's provide 'no real increase in accuracy'!