Review of Economic Studies (1990) 57, 99-125
0034-6527/90/00060099$02.00


# Statistical Inference in Instrumental Variables Regression with I(1) Processes

PETER C. B. PHILLIPS and BRUCE E. HANSEN
*Cowles Foundation for Research in Economics, Yale University*

*First version received April* 1988; *final version accepted April* 1989 (*Eds.*)

This paper studies the asymptotic properties of instrumental variable (IV) estimates of multivariate cointegrating regressions and allows for deterministic and stochastic regressors as well as quite general deterministic processes in the data-generating mechanism. It is found that IV regressions are consistent even when the instruments are stochastically independent of the regressors. This phenomenon, which contrasts with traditional theory for stationary time series, is a beneficial artifact of spurious regression theory whereby stochastic trends in the instruments ensure their relevance asymptotically. Problems of inference are also addressed and some promising new theoretical results are reported. These involve a class of Wald tests which are modified by semiparametric corrections for serial correlation and for endogeneity. The resulting test statistics which we term fully-modified Wald tests have limiting $\chi^2$ distributions, thereby removing the obstacles to inference in cointegrated systems that were presented by the nuisance parameter dependencies in earlier work.

Some simulation results are reported which seek to explore the sampling behaviour of our suggested procedures. These simulations compare our fully modified (semiparametric) methods with the parametric error-correction methodology that has been extensively used in recent empirical research and with conventional least squares regression. Both the fully-modified and error-correction methods work well in finite samples and the sampling performance of each procedure confirms the relevance of asymptotic distribution theory, as distinct from super-consistency results, in discriminating between statistical methods.

## 1. INTRODUCTION

Economic time series are widely believed to possess certain non-classical properties which invalidate the routine application of many standard statistical procedures. The first of these is the joint dependence of most aggregate time series. In dealing with this complication econometricians produced the body of statistical theory that is now known as simultaneous equations and involves methods such as instrumental variables (IV) and full-information maximum likelihood (FIML).

The second non-classical property is non-stationarity. Until recently non-stationarity has been dealt with in practice largely by trend elimination through pre-filtering and more often than not it has simply been ignored in theoretical developments. The last few years have seen major research efforts to alleviate these shortcomings. Problems of estimation and inference in regression models with autoregressive unit roots have been examined in some detail. In such models a complete theory of regression is well within reach. The approach developed in earlier work (Phillips (1986a, 1987) and Phillips and Durlauf (1986)) has proved especially fruitful. It is used in two recent papers by Park and Phillips (1988, 1989) to construct a general asymptotic framework for multivariate regressions with integrated processes of different orders allowing for drifts, trends and cointegration.

Related work has been done on the subject by other researchers, notably Stock (1987) and Sims, Stock and Watson (1987).

The present paper follows the framework of Park and Phillips (1988, 1989). Our primary objective is to extend the results in these papers to allow for IV regressions. In doing so, we allow for deterministic as well as stochastic instruments. We also permit quite general deterministic processes in the data-generating mechanism. It is found that the Park-Phillips results extend quite readily to the new models and estimators. However, some results stand out as being of particular interest.

First, we discover that an IV cointegrating regression leads to consistent estimates even when the instruments are stochastically independent of the regressors. This phenomenon may strike some as surprising since with stationary time series stochastically independent instrumental variables clearly fail to satisfy the asymptotic relevance condition for consistency. However, for integrated regressors the individual stochastic trends of a set of instruments are sufficient to ensure that the relevance condition holds even when the instruments are independent. This outcome is, of course, an artifact of spurious regression theory—see Phillips (1986*a*) for details. Indeed, the very correlation that gives rise to spurious regression also ensures the validity of the relevance condition for independent instruments in IV regressions.

Second, problems of inference in IV regressions are studied—with some promising theoretical results. Earlier work has shown up the importance of second-order asymptotic bias effects in least squares cointegrating regressions (see the asymptotic theory in Phillips and Durlauf (1986), Stock (1987) and the simulation findings in Banerjee *et al.* (1986)). Given the original objective of IV regression in the context of simultaneous equations it is of special interest to determine the extent to which suitable instruments can help to solve this problem in the present context. Our analysis shows that instruments are not themselves sufficient to eliminate the bias effects asymptotically when there is endogeneity in the regressors. Instead, we suggest an alternative semiparametric correction which does lead to asymptotically median-unbiased estimators. The correction may be employed in OLS or IV regressions. The modified estimators form the basis of what we call *fully modified Wald* tests. These are Wald statistics for testing general linear hypotheses about the coefficients in a cointegrating regression. Their asymptotic distributions are $\chi^2$ and traditional methods of inference are therefore applicable provided the correct modifications to conventional Wald tests are used. These results provide a major extension of the Park-Phillips analysis and help to solve the inference problem in cointegrating regressions.

The new results mentioned above provide an alternative to the optimal inference procedures considered recently in Phillips (1988*b*). The later are based on full maximum likelihood estimation (MLE) of the cointegrated system and require complete specification and estimation of the system, typically but not exclusively in error-correction mechanism format. Such full MLE procedures are parametric in nature. The procedures in the present paper rely on semiparametric corrections. They are of the type that were developed originally in Phillips (1987) for unit root tests. In the present IV multivariate setting they are more involved and require two levels of correction: one serial correlation correction as in Phillips (1987); and a second long-run endogeneity correction.

The paper is organized as follows. Section 2 outlines the models and discusses some background theory. Section 3 describes the estimators that are studied and develops an asymptotic theory for the estimated coefficients in the case of both deterministic and stochastic instruments. Section 4 considers Wald tests of linear hypotheses about the coefficients and gives a general asymptotic theory. Block tests are also studied and particular attention is given to characterizing the parameter dependencies in the limit distributions. Section 5 develops some new statistics called fully-modified Wald tests which are asymptotically distributed $\chi^2$ criteria. Section 6 reviews some experimental evidence with these new procedures and reports the results of a simulation study that compares our fully-modified semiparametric methods with the error-correction model (ECM) methodology that is now popular in empirical research. This section is inspired by the recent analytical investigation in Phillips (1988*c*) of the methodological prescriptions outlined in Hendry and Richard (1982, 1983) for empirical time-series research. Some conclusions and suggestions for further work are given in Section 7.

Our notation follows that of earlier papers in this sequence. We use the symbol "$\Rightarrow$" to signify weak convergence, the symbol "$\equiv$" to signify equality in distribution and the inequality "$>0$" to signify positive-definite when applied to matrices. Stochastic processes such as the Brownian motion $W(r)$ on $[0, 1]$ are frequently written as $W$ to achieve notational economy. Similarly, we write integrals with respect to Lebesgue measure such as $\int_0^1 W(s)ds$ more simply as $\int_0^1 W$. Vector Brownian motion with covariance matrix $\Omega$ is written "$BM(\Omega)$". We use $\|A\|$ to represent the Euclidean norm $\operatorname{tr}(A'A)^{1/2}$ of the matrix $A$, $O(n)$ to denote the orthogonal group of order $n$ and $I(1)$ and $I(0)$ to signify time series that are integrated of order one and zero, respectively. Finally, all limits given in the paper are as the sample size $T\to\infty$ unless otherwise stated.

## 2. MODELS AND BACKGROUND THEORY

We shall be working with an $n$-dimensional time series $\{y_t\}_0^\infty$ partitioned as

$$y_t' = (\underset{n_1}{y_{1t}'}, \underset{n_2}{y_{2t}'}, \underset{n_3}{y_{3t}'}); \qquad n = n_1 + n_2 + n_3, \qquad n_3 \geqq n_2 \tag{1}$$

and generated by the system

$$y_{1t} = Ay_{2t} + \Pi k_{1t} + u_{1t} \tag{2}$$

$$\Delta y_{2t} = u_{2t} \tag{3}$$

$$\Delta y_{3t} = u_{3t}. \tag{4}$$

The initialization of this system is at $t=0$ and $y_0$ may be any random variable. The innovation vector $u_t = (u_{1t}', u_{2t}', u_{3t}')'$ is taken to be strictly stationary and ergodic with zero mean, finite covariance matrix $\Sigma > 0$ and continuous spectral density matrix $f_{uu}(\lambda)$ with $\Omega = 2\pi f_{uu}(0)$. Unless otherwise stated we shall suppose that $\Omega > 0$. We further assume that the partial sum process constructed from $u_t$ satisfies the multivariate invariance principle

$$T^{-1/2}\Sigma_1^{[Tr]} u_j \Rightarrow B(r) \equiv BM(\Omega), \qquad 0 < r \leqq 1, \tag{5}$$

where [ ] denotes "integer part." We decompose the "long run" covariance matrix $\Omega$ as follows:

$$\Omega = \Sigma + \Lambda + \Lambda'$$

where

$$\Sigma = E(u_0 u_0'), \qquad \Lambda = \Sigma_{k=1}^\infty E(u_0 u_k');$$

and we define

$$\Delta = \Sigma + \Lambda.$$

Explicit conditions under which (5) holds are discussed in detail in earlier work—see Phillips (1988*a*) for references and a review. We partition $B$, $\Omega$, $\Sigma$, $\Lambda$ and $\Delta$ comformably with $u_t$. Thus, in the case of $\Omega$ we write

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega'_{21} & \Omega'_{31} \\ \Omega_{21} & \Omega_{22} & \Omega'_{32} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} \end{bmatrix}.$$

The vector $k_{1t}$ in (2) is a subvector of

$$k'_t = (\underset{m_1}{k'_{1t}}, \underset{m_2}{k'_{2t}}); \; m = m_1 + m_2, \; m_2 \geqq n_2 \tag{6}$$

which is a deterministic function of time. In the most common applications $k_{1t}$ will consist of a constant, a time trend or a simple polynomial trend. In such cases estimates of the matrix $A$ in (2) that are discussed in this paper are invariant to the replacement of (3) and (4) by the alternative generating mechanisms

$$\Delta y_{2t} = \Pi_2 \Delta k_{1t} + u_{2t} \tag{3'}$$

$$\Delta y_{3t} = \Pi_3 \Delta k_{1t} + u_{3t}. \tag{4'}$$

In some cases it is convenient to work with a triangular array $\{\{y_{tT}\}_{t=1}^{T}\}_{T=1}^{\infty}$ in place of $y_t$. This allows us the additional flexibility of using deterministic functions, $k_{Tt}$, which are also indexed by the sample size $T$. It is then possible to accommodate such deterministic functions as the sinusoidal trends

$$\{t^{p_i} \sin(\lambda_j t/T), \, t^{p_i} \cos(\lambda_j t/T); \, i = 1, \ldots, I; j = 1, \ldots, J\}.$$

Given some such vector $k_{Tt}$ we assume the existence of a diagonal matrix of weights $\delta_T > 0$ satisfying $\|\delta_T\| \to 0$ and a vector of functions $k(\;)$ for which

$$\lim \sup_{T\to\infty} \sup_t \sup_{(t-1)/T \leqq r < t/T} \|\delta_T k_{Tt} - k(r)\| = 0 \tag{7}$$

and

$$\int_0^1 kk' > 0. \tag{8}$$

We partition $\delta_T$ conformably with $k$ as $\delta_T = \text{diag}(\delta_{1T}, \delta_{2T})$. When we need only work with single indexed deterministic functions like $k_t$, we shall drop the additional subscript but continue to assume that (7) and (8) hold. We shall not overburden the notation when we do use $k_{Tt}$ by insisting also on triangular array notation for $y_t$ and $u_t$. The extensions to the underlying theory that are needed to accommodate this generalization are rather obvious. For example, we may conveniently replace (5) by a functional CLT for triangular arrays.

## 3. ESTIMATION THEORY

Our framework and approach is related closely to that of earlier work on OLS procedures. Two IV estimators in a regression on (2) will be considered. The first uses the vector

$$z'_t = (y'_{3t}, k'_{1t}) \tag{9}$$

as instruments; and the second uses $k_t$ (or $k_{Tt}$ as the case may be) as in (6). We shall call these, in brief, the IVZ and IVK estimators, respectively. If we rewrite (2) as

$$y_{1t} = \Gamma x_t + u_{1t}$$

where

$$\Gamma = (A, \Pi), \qquad x_t = (y_{2t}', k_{1t}')',$$

standard regression theory supplies the following formulae

$$\hat{\Gamma} = (\Sigma_1^T y_{1t} x_t')(\Sigma_1^T x_t x_t')^{-1}$$
$$\tilde{\Gamma} = (\Sigma_1^T y_{1t} \tilde{x}_t')(\Sigma_1^T \tilde{x}_t \tilde{x}_t')^{-1}$$
$$\bar{\Gamma} = (\Sigma_1^T y_{1t} \bar{x}_t')(\Sigma_1^T \bar{x}_t \bar{x}_t')^{-1}$$

where here and elsewhere in the paper we use "^", "~" and "–" affixes on the parameter matrices (including the submatrices of $B$) to signify OLS, IVZ and IVK estimators, respectively. In the formulae above

$$\tilde{x}_t = (\tilde{y}_{2t}', k_{1t}')'$$
$$\bar{x}_t = (\bar{y}_{2t}', k_{1t}')'$$
$$\tilde{y}_{2t} = (\Sigma_1^T y_{2t} z_t')(\Sigma_1^T z_t z_t')^{-1} z_t$$
$$\bar{y}_{2t} = (\Sigma_1^T y_{2t} k_t')(\Sigma_1^T k_t k_t')^{-1} k_t.$$

As in Park and Phillips (1988), the limit distributions of these estimators may be expressed rather conveniently in terms of the functional

$$f(B, M, E) = \left(\int_0^1 dBM' + E\right)\left(\int_0^1 MM'\right)^{-1} \tag{10}$$

where $B$ is a vector Brownian motion and $M$ is a stochastic process obtained from $B$ by a suitable Hilbert projection. Since the coefficient estimates will converge at different rates, we define the weight matrix

$$W_T = \begin{bmatrix} I_{n_2} T^{1/2} & 0 \\ 0 & \delta_{1T}^{-1} \end{bmatrix}.$$

**Theorem 3.1.** (a) $T^{1/2}(\hat{\Gamma} - \Gamma) W_T \Rightarrow f(B_1, J_2, (\Delta_{21}', 0))$, (b) $T^{1/2}(\tilde{\Gamma} - \Gamma) W_T \Rightarrow f(B_1, \tilde{J}_2, (\Delta_{31}' F_{23}^{*\prime}, 0))$, (c) $T^{1/2}(\bar{\Gamma} - \Gamma) W_T \Rightarrow f(B_1, \bar{J}_2, 0)$ *where*

$$J_2(r) = (B_2'(r), k_1'(r))'$$
$$\tilde{J}_2(r) = (\tilde{B}_2'(r), k_1'(r))'$$
$$\bar{J}_2(r) = (\bar{B}_2'(r), k_1'(r))'$$
$$\tilde{B}_2(r) = \int_0^1 B_2 Z' \left(\int_0^1 ZZ'\right)^{-1} Z(r)$$
$$\bar{B}_2(r) = \int_0^1 B_2 k' \left(\int_0^1 kk'\right)^{-1} k(r)$$
$$F_{23}^* = \int_0^1 B_2^* B_3^{*\prime} \left(\int_0^1 B_3^* B_3^{*\prime}\right)^{-1}$$
$$Z(r) = (B_3(r)', k_1(r)')'$$
$$B^*(r) = B(r) - \int_0^1 Bk_1' \left(\int_0^1 k_1 k_1'\right)^{-1} k_1(r).$$

*Remark* (*a*). All three estimators of $\Gamma$ are consistent. The result is of special interest in the case of the IVZ estimator $\tilde{A}$ of the submatrix $A$. This is because consistency holds irrespective of the properties of the instruments $y_{3t}$. For example, $y_{3t}$ may comprise a set of spurious instruments which are statistically independent of the regressors $y_{2t}$ in (2). At first this appears surprising because in the case of stationary time series such instruments would fail the usual relevance condition asymptotically. But here, since both processes are $I(1)$, we find

$$T^{-2}\Sigma_1^T y_{2t}y_{3t}' \Rightarrow \int_0^1 B_2B_3' \tag{11}$$

and, moreover, since $n_3 \geqq n_2$ we have

(C1) $$\operatorname{rank}\left(\int_0^1 B_2B_3'\right) = n_2 \quad \text{a.s.}$$

as shown in Lemma (A3) and the remark following Lemma (A3) in the Appendix. Result (11) arises in spurious regression theory—see Phillips (1986*a*)—and is a manifestation of thè fact that two independent $I(1)$ processes appear correlated even in the limit because they both carry stochastic trends. In effect, the consistency of the IVZ estimator $\tilde{A}$ is a beneficial artifact of spurious regression theory. It tells us that we can generate a stochastic trend from purely random numbers and still obtain consistent estimates by using the resulting $I(1)$ series as instruments for $y_{2t}$ in (2).

Note that the orthogonality condition for consistency also holds because

(C2) $$T^{-2}\Sigma_1^T u_{1t}y_{3t}' \underset{p}{\rightarrow} 0,$$

again irrespective of the properties of $I(1)$ process $y_{3t}$.

$\bar{A}$ is consistent for similar reasons. Since the regressor $y_{2t}$ has a stochastic trend we find that

$$T^{-3/2}\Sigma_1^T y_{2t}k_{2t}'\delta_{2T} \Rightarrow \int_0^1 B_2k_2'. \tag{12}$$

and by Lemma (A3) in the Appendix

(C3) $$\operatorname{rank}\left(\int_0^1 B_2k_2'\right) = n_2 \quad \text{a.s.}$$

so that the relevance condition for the instruments $k_2$ is again satisfied. Thus, any deterministic regressor retains an asymptotic correlation with a stochastic trend upon appropriate standardization.

*Remark* (*b*). The processes that appear in the limit distributions given in Theorem 3.1 bear a close relationship in form to the time series that are used in the construction of the estimates. In particular, the Euclidean projections that appear in the formulae for $\tilde{\Gamma}$ and $\bar{\Gamma}$ are replaced by Hilbert projections in the limit distributions. This relationship between finite sample regression formulae and limit theory has been noted and discussed in earlier work—see Phillips (1988*a*) for details. Here the projections are superposed because of the multiple regressor nature of (2) and the use of instruments. Thus $\tilde{B}_2$ is the projection in $L_2[0,1)^{n_2}$ of $B_2$ onto the subspace spanned by the elements of $I_{n_2} \otimes Z'$. $\bar{B}$, in turn, is the projection in $L_2[0,1)^n$ of $B$ onto the subspace spanned by the elements of $I_n \otimes k'$. These limit processes are the function space analogues of the time series $\tilde{x}_{2t}$ and $\bar{y}_t$, respectively. Our understanding of the limit distributions in Theorem 3.1 is enhanced by noting these similarities.

*Remark* (*c*). The expressions $\Delta_{21}$ and $F_{23}^*\Delta_{31}$ that appear in the limit functionals in (a) and (b) of Theorem 3.1 are second-order bias effects. We use the terminology "second-order" because the consistency of the estimates is, of course, unaffected. However, the bias does influence the centring of the limit distribution and is normally indicative of the presence of bias in finite sample which can be substantial. The bias effect arises because of the contemporaneous and serial dependence of the regressor and its instruments ($y_{2t}$ and $y_{3t}$) in the case of the estimators $\hat{A}$ and $\tilde{A}$. Note that since deterministic instruments $k_{2t}$ are used for $y_{2t}$ in $\bar{A}$ and these are asymptotically uncorrelated with the regressor error $u_{1t}$ in (2) no second-order bias effect is present in this case.

As in Park and Phillips (1988) these bias effects may be consistently estimated and eliminated. In what follows we use $\hat{\Delta}$ and $\tilde{\Delta}$ to denote consistent estimates of $\Delta$ constructed from OLS and IVZ regression residuals respectively. (The construction of such estimators is discussed in Park and Phillips.) Next we define the "bias-corrected" estimators

$$\hat{\Gamma}^* = [\Sigma_1^T y_{1t}x_t' - T(\hat{\Delta}_{21}', 0)](\Sigma_1^T x_t x_t')^{-1}$$

$$\tilde{\Gamma}^* = [\Sigma_1^T y_{1t}z_t' - T(\hat{\Delta}_{31}', 0)](\Sigma_1^T z_t z_t')^{-1}(\Sigma_1^T z_t x_t')(\Sigma_z^T \tilde{x}_t \tilde{x}_t')^{-1}.$$

The resulting limit distributions no longer involve the non-centralities.

**Theorem 3.2.** (a) $T^{1/2}(\hat{\Gamma}^* - \Gamma) W_T \Rightarrow f(B_1, J_2, 0)$, (b) $T^{1/2}(\tilde{\Gamma}^* - \Gamma) W_T \Rightarrow f(B_1, \tilde{J}_2, 0)$.

*Remark* (*d*). Results for OLS that are equivalent to part (a) of Theorem 3.1 are given by Park and Phillips (1988, Theorem 3.3). The present result applies for rather general deterministic regressors and this is reflected in the definition of the Gaussian process $B^*$. The Park–Phillips results were obtained explicitly for the case of a drift and time trend in (2).

*Remark* (*e*). Theorem 3.1 and 3.2 hold as stated when $y_{2t}$ and $y_{3t}$ are generated by (3) and (4). If the alternative generating mechanisms (3′) and (4′) apply then $\hat{\Pi}$, $\tilde{\Pi}$ and $\bar{\Pi}$ are still consistent estimators but they have different limit distributions. The differences are caused by the fact that under (3′) and (4′) $y_{2t}$ and $y_{3t}$ have elements which are in general dominated by the deterministic rather than the stochastic trends. In the case of the coefficient estimates $\hat{A}$, $\tilde{A}$ and $\bar{A}$ the deterministic trends are eliminated by projection because $k_{1t}$ is also present in (2). However, the effect of the deterministic trends on $y_{2t}$ and $y_{3t}$ must be allowed for in the estimation of $\Pi$. From (3′) we have

$$y_{2t} = S_{2t} + \Pi_2(k_{1t} - k_{10}) + y_{20}$$

where

$$S_{2t} = \Sigma_1^t u_{2j}.$$

Define the weight matrix $W_T^*$ by

$$W_T^* = \begin{bmatrix} I_{n_2}T^{-1/2} & -\Pi_2 T^{-1/2} \\ 0 & \delta_{1T} \end{bmatrix}^{-1},$$

so that

$$W_T^{*-1}\begin{pmatrix} y_{2t} \\ k_{1t} \end{pmatrix} = \begin{pmatrix} T^{-1/2}S_{2t} + o_p(1) \\ \delta_{1T}k_{1t} \end{pmatrix}.$$

Standard manipulations reveal that Theorems 3.1 and 3.2 now hold as stated if $W_T^*$ replaces $W_T$.

## 4. HYPOTHESIS TESTING

### 4.1. *General theory*

The limit theory presented in Theorems 3.1 and 3.2 is nonstandard. The distributions belong to the limiting Gaussian functional (LGF) family explored in Phillips (1989). In general, these limit distributions cause problems for statistical inference through their dependence on many nuisance parameters and their nonstandard nature. In particular, traditional methods of inference which rely on $t$- and $F$-ratios and Wald tests are not useful without modification in this context. Earlier work, commencing with Phillips' (1987) unit root tests, showed how to perform such modifications. A fairly general theory in the linear model was formulated in Park and Phillips (1988). This section shows how to extend the theory in Park-Phillips to the present IV set up. As far as possible we shall use the Park-Phillips notation to facilitate reference to that work.

We start by considering the following linear hypotheses about the coefficient matrix $\Gamma=[A, \Pi]$ in (2):

$$H_0: \quad R \operatorname{vec} \Gamma = r$$

where $R$ $(g \times n_1 \times (n_2+m_1))$ has rank $g$. To the extent that the diagonal elements of $W_T$ differ in orders of magnitude (associated with differing asymptotic behaviour in the elements of $k_{1t}$ and $y_{2t}$), we are effectively restricted in asymptotic tests to tests of separable restrictions, i.e. about $A$ alone, or individual columns of $\Pi$. Thus if we rewrite $H_0$ as

$$H_0: \quad R^+ \operatorname{vec}(\Gamma') = r$$

where $R=R^+K$ and $K$ is the commutation matrix of order $n_1(n_2+m_1)$, then $R^+$ must be block-diagonal across columns of $\Gamma$ which are of different orders. For example, if the model is

$$y_{1t} = Ay_{2t} + \pi t + u_{1t}$$

then the hypothesis

$$H_0': \quad A = A_0, \qquad \pi = \pi_0$$

may be mounted, but the hypothesis

$$H_0'': \quad A + \pi = r \qquad (n_2 = m_1 = 1)$$

cannot be tested using our asymptotic theory.

$H_0$ is frequently tested in traditional regression models by the following Wald statistics

$$G_R(\underline{\Gamma}, V) = (R \operatorname{vec} \underline{\Gamma} - r)'[R(V \otimes \underline{M})R']^{-1}(R \operatorname{vec} \underline{\Gamma} - r) \tag{13}$$

In these formulae we employ the generic notation

$$\underline{\Gamma} = \hat{\Gamma}, \tilde{\Gamma} \text{ or } \bar{\Gamma}$$

$$\underline{M} = \hat{M}, \tilde{M} \text{ or } \bar{M}$$

where

$$\hat{M} = (\Sigma_1^T x_t x_t')^{-1}, \qquad \tilde{M} = (\Sigma_1^T \tilde{x}_t \tilde{x}_t')^{-1}, \qquad \bar{M} = (\Sigma_1^T \bar{x}_t \bar{x}_t')^{-1}$$

and

$$V = \hat{\Sigma}_{11} \text{ or } \underline{\Omega}_{11} \quad \text{where } \underline{\Omega}_{11} = \hat{\Omega}_{11}, \tilde{\Omega}_{11} \text{ or } \bar{\Omega}_{11}. \tag{15}$$

Here $\hat{\Sigma}_{11}$ and $\underline{\Omega}_{11}$ are consistent estimators of $\Sigma_{11}$ and $\Omega_{11}$, respectively. When $V=\hat{\Sigma}_{11}$ the $G$ statistics are formulated in the conventional manner for linear regression. As shown in Park and Phillips (1988) the formulation with $V=\Omega_{11}$ is more useful in regressions with $I(1)$ processes since it is the long-run covariance matrix $\Omega_{11}$ upon which the asymptotic distributions depend. Consistent estimation of $\Omega_{11}$ is discussed elsewhere—see Phillips and Durlauf (1986), Newey and West (1987) and Andrews (1988). We use the notation $\hat{\Omega}_{11}$, $\tilde{\Omega}_{11}$ and $\bar{\Omega}_{11}$ in (15) to signify estimates of $\Omega_{11}$ that are based on OLS, IVZ and IVK residuals from (2), respectively.

To simplify the presentation of the asymptotic theory for the $G$-statistics we use the following functional from Park and Phillips (1988):

$$g_R(B, M, E)=\operatorname{vec}(f(B, M, E))'R'\left\{R\left(\Omega_{11}\otimes\left(\int_0^1 MM'\right)^{-1}\right)R'\right\}^{-1} R\operatorname{vec}(f(B, M, E)).$$

The non-centrality parameters which appear in the limit representations for $\hat{\Gamma}$ and $\tilde{\Gamma}$ (see Theorem 3.1) also arise in $G$-statistics constructed from these estimates, just as in the theorems of Park and Phillips. These distributions are less useful for inference than those based on estimators which have less nuisance parameter dependencies. We examine, therefore, the "bias-corrected" estimators $\hat{\Gamma}^*$ and $\tilde{\Gamma}^*$ rather than $\hat{\Gamma}$ and $\tilde{\Gamma}$. The limit theory is as follows:

**Theorem 4.1.** *Under $H_0$*
(a) $G_R(\hat{\Gamma}^*, \hat{\Omega}_{11})\Rightarrow g_R(B_1, J_2, 0)$
(b) $G_R(\tilde{\Gamma}^*, \tilde{\Omega}_{11})\Rightarrow g_R(B_1, \tilde{J}_2, 0)$
(c) $G_R(\bar{\Gamma}, \bar{\Omega}_{11})\Rightarrow g_R(B_1, \bar{J}_2, 0)$.

*Remark (a).* Theorem 4.1 extends the Park–Phillips theory in several ways. First, it allows for the presence of general deterministic regressors rather than a constant and a time trend. Second, it provides for general restrictions on the $\Pi$ coefficients rather than simple block tests. Finally, it accommodates general instrumental variable regressions as well as least squares.

*Remark (b).* Theorem 4.1 examines tests of linear hypotheses, but the results easily extend to general non-linear hypotheses of the form $h(\Gamma_0)=0$ provided $h(\cdot)$ is continuously differentiable and $\partial h(\Gamma_0)/\partial \operatorname{vec}\Gamma$ satisfies the block diagonality and rank conditions discussed above for $R$.

*Remark (c).* The limit distributions given in Theorem 4.1 depend in general on the matrices $R$ and $\Omega$. This parameter dependency is analogous to that which one typically finds in the finite sample distributions of multivariate tests—see Phillips (1986*b*) for an analysis of this problem with respect to Wald tests in the conventional multivariate linear model. However, here the problem persists asymptotically with the result that the statistics cannot be used to mount tests that are asymptotically similar, i.e. have the same size asymptotically for all values of the nuisance parameters. Some reductions in the parameter dependencies can be achieved in certain special cases as we shall now illustrate.

## 4.2. *Block Tests*

We will constrain the analysis to the special case $R=\operatorname{diag}(I_{n_2}, 0)$, so that the hypotheses tested are of the form $H_0$: $A=A_0$. In this case the limit distributions given in Theorem

4.1 have a manageable and intuitively interesting parameterization. Working from part (a) of Theorem 4.1 we find

$$g_R(B_1, J_2, 0) = g_1(B_1, B_2^*, 0) = \left(\int_0^1 dB_1' \otimes B_2^{*\prime}\right)\left\{\Omega_{11}^{-1} \otimes \left(\int_0^1 B_2^* B_2^{*\prime}\right)^{-1}\right\}\left(\int_0^1 dB_1 \otimes B_2^*\right). \tag{16}$$

Observe that

$$\int_0^1 \operatorname{var}\{dB_1 \otimes B_2^* \mid B_2^*(s), s \leqq r\} = \int_0^1 \Omega_{11} \otimes B_2^* B_2^{*\prime} = \Omega_{11} \otimes \int_0^1 B_2^* B_2^{*\prime}$$

so that this random matrix is a natural metric for the quadratic form (16).

To simplify (16) we transform coordinates as follows. Define the processes

$$\begin{bmatrix} \underline{B}_1 \\ \underline{B}_2 \end{bmatrix} = \begin{bmatrix} \Omega_{11}^{-1/2} & 0 \\ 0 & \Omega_{22}^{-1/2} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \equiv BM(V)$$

and

$$\underline{B}_2^* = \underline{B}_2 - \int_0^1 \underline{B}_2 k_1' \left(\int_0^1 k_1 k_1'\right)^{-1} k_1$$

where

$$V = \begin{bmatrix} I & \Omega_{11}^{-1/2}\Omega_{12}\Omega_{22}^{-1/2} \\ \Omega_{22}^{-1/2}\Omega_{21}\Omega_{11}^{-1/2} & I \end{bmatrix} = \begin{bmatrix} I & \underline{\Omega}_{12} \\ \underline{\Omega}_{21} & I \end{bmatrix}, \quad \text{say.}$$

Using Lemma 3.1 of Phillips (1989) we may write

$$\underline{B}_1 \equiv \underline{\Omega}_{12}\underline{B}_2 + (I - P_{12})^{1/2} W_1$$

where

$$P_{12} = \underline{\Omega}_{12}\underline{\Omega}_{21} = \Omega_{11}^{-1/2}\Omega_{12}\Omega_{22}^{-1}\Omega_{21}\Omega_{11}^{-1/2}$$

and $W_1 \equiv BM(I_{n_1})$ is independent of $\underline{B}_2$. Next we assume that $n_2 \geqq n_1$, that $\Omega_{12}$ has full rank $n_1$ (both assumptions will be relaxed later) and transform

$$\underline{B}_2 \to H'\underline{B}_2 = W_2$$

$$\underline{B}_2^* \to H'\underline{B}_2^* = W_2^*$$

where

$$H = [H_1, H_2] \in 0(n_2)$$

and

$$H_1 = \underline{\Omega}_{21}(\underline{\Omega}_{12}\underline{\Omega}_{21})^{-1/2} = \underline{\Omega}_{21} P_{12}^{-1/2}.$$

Then

$$\underline{\Omega}_{12}\underline{B}_2 = \underline{\Omega}_{12} H H' \underline{B}_2 = [P_{12}^{1/2} \quad 0] W_2. \tag{17}$$

The quadratic form (16) now reduces as follows:

$$g_I(B_1, B_2^*, 0) = \left(\int_0^1 d\underline{B}_1' \otimes \underline{B}_2^{*\prime}\right)\left\{I \otimes \left(\int_0^1 \underline{B}_2^* \underline{B}_2^{*\prime}\right)^{-1}\right\}\left(\int_0^1 d\underline{B}_1 \otimes \underline{B}_2^*\right)$$

$$\equiv \left(\int_0^1 d\underline{B}_1' \otimes W_2^{*\prime}\right)\left\{I \otimes \left(\int_0^1 W_2^* W_2^{*\prime}\right)^{-1}\right\}\left(\int_0^1 d\underline{B}_1 \otimes W_2^*\right)$$

and

$$\underline{B}_1 = (I - P_{12})^{1/2} W_1 + [P_{12}^{1/2} \quad 0] W_2.$$

Thus

$$g_I(B_1, B_2^*, 0) = h(\underline{B}_1, W_2^*) \tag{18}$$

where

$$h(B, M) = \left( \int_0^1 dB' \otimes N(M)' \right) \left( \int_0^1 dB \otimes N(M) \right)$$

and

$$N(M) = \left( \int_0^1 MM' \right)^{-1/2} M.$$

A final rotation of $W$ by diag $(Q, Q, I)$, where $Q$ is the orthogonal matrix of latent vectors of $P_{12}$, diagonalizes $P_{12}$, leaving only the latent roots.

In the above we have assumed that $\Omega_{21}$ (or equivalently $\underline{\Omega}_{21}$) has full column rank. If rank $(\underline{\Omega}_{21}) = p < n_1$ we simply rotate coordinates in $R^{n_1}$ so that the leading submatrix of $\underline{\Omega}_{21}$ has full column rank. The result stated above in (18) still applies but now $P_{12}$ is of course a singular matrix. The rotation in $R^{n_1}$ transforms $P_{12}$ to the block diagonal form

$$P_{12} = \text{diag}\,(\underline{P}_{12}, 0),$$

where $\underline{P}_{12}$ is $p \times p$. The distributional result (18) is the same whether we use this reduction or simply (18) as stated. Analogous problems arise when $n_1 > n_2$ since $P_{12}$ is then always singular.

These considerations now lead us to the following formal statement. The proofs for IVZ and IVK are similar to those for OLS. The interested reader is referred to our working paper (Phillips and Hansen (1988)) for details of the proofs and the necessary constructions.

**Theorem 4.2.** *Under* $H_0$,
(a) OLS: $g_I(B_1, B_2^*, 0) = h(\hat{W}_1, W_2^*)$ *where if* $n_2 \geqq n_1$,

$$\hat{W}_1 = (I - \Lambda_{12})^{1/2} W_1 + [\Lambda_{12}^{1/2}, 0] W_2$$

$$\Lambda_{12} = \text{diag}\,\{\textit{latent roots of } P_{12} = \underline{\Omega}_{12}\underline{\Omega}_{21}\}$$

*and if* $n_2 < n_1$,

$$\hat{W}_1 = \begin{bmatrix} (I - \Lambda_{21})^{1/2} & 0 \\ 0 & I \end{bmatrix} W_1 + \begin{bmatrix} \Lambda_{21}^{1/2} \\ 0 \end{bmatrix} W_2$$

$$\Lambda_{21} = \text{diag}\,\{\textit{latent roots of } P_{21} = \underline{\Omega}_{21}\underline{\Omega}_{12}\}.$$

(b) IVZ: $g_I(B_1, \tilde{B}_2^*, 0) = \tilde{g}_I = h(B_1, \tilde{B}_2^*)$

$$B_1 = (I - \underline{\Omega}_{12}^*\underline{\Omega}_{21}^* - \underline{\Omega}_{13}\underline{\Omega}_{31})^{1/2} W_1 + \underline{\Omega}_{12}^* W_2 + \underline{\Omega}_{13} W_3$$

$$\tilde{B}_2^* = (I - \underline{\Omega}_{23}\underline{\Omega}_{32})^{1/2} \tilde{W}_2^* + \underline{\Omega}_{23} W_3^*$$

$$\underline{\Omega}_{21}^* = (I - \underline{\Omega}_{23}\underline{\Omega}_{32})^{-1/2}(\underline{\Omega}_{21} - \underline{\Omega}_{23}\underline{\Omega}_{31}) = \underline{\Omega}_{12}^{*\prime}$$

$$\tilde{W}_2^* = \left( \int_0^1 W_2^{*\prime} W_3^{*\prime} \right) \left( \int_0^1 W_3^* W_3^{*\prime} \right)^{-1} W_3^*.$$

(c) IVK: $g_I(B_I, \bar{B}_2^*, 0) \equiv h(\hat{W}_1, \bar{W}_2^*)$

$$\bar{W}_2^* = \left(\int_0^1 W_2^* k_2^{*\prime}\right)\left(\int_0^1 k_2^* k_2^{*\prime}\right)^{-1} k_2^*$$

$$k_2^* = k_2 - \int_0^1 k_2 k_1'\left(\int_0^1 k_1 k_1'\right)^{-1} k_1.$$

*In* (a), (b) *and* (c),

$$W(r) = (\underset{n_1}{W_1'}, \underset{n_2}{W_2'}, \underset{n_3}{W_3'})' \equiv BM(I_n)$$

$$W^*(r) = W(r) - \int_0^1 Wk_1'\left(\int_0^1 k_1 k_1'\right)^{-1} k_1(r)$$

$$\underline{\Omega}_{ab} = \Omega_{aa}^{-1/2}\Omega_{ab}\Omega_{bb}^{-1/2}.$$

**Corollary 4.3.** (a) *If* $\underline{\Omega}_{31} = 0$ *and* $\underline{\Omega}_{32} = 0$ *then* $\tilde{g}_I \equiv h(B_1, \tilde{B}_2^*) \equiv h(\hat{W}_1, \tilde{W}_2^*)$ *where* $\hat{W}_1$ *and* $\tilde{W}_2^*$ *are defined in parts* (a) *and* (b) *of Theorem* 4.2. (b) *If* $\underline{\Omega}_{31} = 0$ *and if* $y_{2t}$ *and* $y_{3t}$ *are cointegrated then* $\tilde{g}_I \equiv \chi^2_{n_1 n_2}$.

*Remark* (*a*). Result (a) of Theorem 4.2 generalizes Lemma 5.6 in Park and Phillips (1988). Observe that when $P_{12} = 0$ we have

$$g_I(B_1, B_2^*; 0) = \left(\int_0^1 dW_1' \otimes N_2'\right)\left(\int_0^1 dW_1 \otimes N_2\right)$$

$$\equiv \chi^2_{n_1 n_2}$$

where $N_2 = N(W_2^*)$ since

$$\int_0^1 dW_1 \otimes N_2 \equiv N\left(0, I \otimes \int_0^1 N_2 N_2'\right) \equiv N(0, I_{n_1 n_2}).$$

On the other hand when $P_{12} = I$ we have

$$g_I(B_1, B_2^*; 0) = \left(\int_0^1 dW_{21}' \otimes N_2'\right)\left(\int_0^1 dW_{21} \otimes N_2\right).$$

where $W_{21} = [I_{n_1} \;\; 0] W_2$. Since $N_2 = N(W_2)$ the limit distribution in this case is a form of unit root distribution. In general, the limit distribution may be regarded as depending on a matrix linear combination of a "unit root" type of stochastic integral and an independent multivariate normal variate. The weights in this linear combination are delivered by the matrix coefficient of determination $P_{12}$.

*Remark* (*b*). Theorem 4.2 shows that the asymptotic dependence on nuisance parameters is more complicated for IVZ based statistics than for those based on OLS. In general, we find that the limit distribution depends on the long-run covariance structure of the innovation processes that drive the structural equation, the regressors and the instruments. Simplifications in the dependence occur as this covariance structure itself simplifies. Some leading cases are given in the corollary.

*Remark* (*c*). The condition $\Omega_{31}=0$ may be interpreted as a second-order orthogonality condition for the instruments. Note that the first-order orthogonality condition (C2) discussed earlier ensures the consistency of the IVZ estimator $\tilde{A}$. The second-order condition $\Omega_{31}=0$ sets the long-run correlation between the equation errors and the instrument errors to zero. The effect of this second-order orthogonality is to reduce parameter dependencies in the limit distribution $\tilde{g}_I$. When $\Omega_{32}=0$, as in part (a) of Corollary 4.3, the instruments are, in effect, long-run uncorrelated with the equation errors *and* the regressor errors. In this case it is only the stochastic trend in the instrument vector $y_{3t}$ that does the work of an instrument and the only parameter dependency in the limit distribution $\tilde{g}_I$ that is left is $P_{12}$, the long-run coefficient of determination between the equation errors and the regressor errors. When $P_{12}=0$ the regressors behave in the long run as if they were exogenous and we find $\tilde{g}_I \equiv \chi^2_{n_1 n_2}$.

*Remark* (*d*). When $y_{2t}$ and $y_{3t}$ are cointegrated, the limit Brownian motions $B_2$ and $B_3$ are related by the equation

$$B_2=\Omega_{23}\Omega_{33}^{-1}B_3=GB_3, \quad \text{say.}$$

This may occur when $y_{3t}$ has been chosen to fulfill the classic role of an instrument in simultaneous equations theory in which a "reduced form" equation for $y_{2t}$ would have the form

$$y_{2t}=Gy_{3t}+\Pi_{21}\Delta k_{1t}+v_{2t},\ v_{2t}\equiv I(0) \tag{3''}$$

in place of (3) or (3′). We observe that in this case the covariance matrix $\Omega$ is singular.

Corollary 4.3 part (b) gives the special case when $\Omega_{31}=0$. In this case, since $\Omega_{21}=0$ also, we find that $y_{2t}$ and $y_{3t}$ are in effect long run exogenous and hence $\tilde{g}_I \equiv \chi^2_{n_1 n_2}$ as given.

*Remark* (*e*). When $n_1=1$ the limit distribution given in Theorem 4.2 has been tabulated in the preprint of Park and Phillips (1988) for the cases $n_2=1$, 2, 3 and with the deterministic regressors $k_{1t}=1$, $k'_{1t}=(1, t)$. The tabulations are given for a grid of values of the scalar coefficient $P_{12}=p_{12}$ over the interval $0<p_{12}<1$. Such tabulations do not seem to be very useful in the general case given here. They would involve the matrix of coefficients $P_{12}$ (or rather its latent roots) in the OLS case and even more involved dependencies in the IVZ case. As a result, another approach will be explored in the next section.

*Remark* (*f*). Hall (1989) has advocated using lagged values of the dependent variable as instruments to construct a univariate unit root test in models with finite-order MA errors. This is covered by our own framework and is equivalent to setting $n_1=n_2=1$, $y_{2t}=y_{1t-1}$, $y_{3t}=y_{1t-k}$ for some $k>l$ (when the errors are MA($l$)). This implies $B_1=B_2=\tilde{B}_2=B_3$ and Theorem 4.2(b) then yields

$$\tilde{g}_I=h(B_3, B_3^*)=h(W, W^*)$$

where $W\equiv BM(1)$. The functional $h(W, W^*)$ gives the asymptotic distribution of the squared value of the Dickey-Fuller $t$-statistic allowing for general deterministic trends in the regression (see Ouliaris, Park and Phillips (1989)).

## 5. ASYMPTOTIC $\chi^2$ CRITERIA

The source of the nuisance parameter dependencies in the limit distributions studied in the previous two sections is the dependence between the limit Brownian motions $B_1$ and

$B_2$. This dependence may, in turn, be interpreted as a form of conventional simultaneous equations bias arising from the endogeneity of the regressors $y_{2t}$ in (2). However, as we have seen from the analysis of the IVZ statistics, traditional methods of dealing with this bias, like instrumental variables, do not eliminate it. The only case so far studied in which the dependency disappears occurs when the regressor $y_{2t}$ is exogenous—a case where simultaneous methods are hardly necessary.

This problem has recently been studied in Phillips (1988*b*). It is shown there that the dependencies in the limit distributions are removed when full maximum likelihood methods of estimation are employed. In the present context this requires joint estimation of (2) and (3) and this includes full estimation of the generating mechanism of the innovations. In the present section we develop a nonparametric procedure that is asymptotically equivalent to full maximum likelihood.

Let $\hat{\Omega}$ be any consistent estimator of $\Omega$ and define:

$$u_{1t}^{+} = u_{1t} - \Omega_{12}\Omega_{22}^{-1}u_{2t}$$

$$y_{1t}^{+} = y_{1t} - \Omega_{12}\Omega_{22}^{-1}u_{2t}$$

$$\hat{y}_{1t}^{+} = y_{1t} - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\Delta y_{2t}.$$

Note that

$$\begin{bmatrix} u_{1t}^{+} \\ u_{2t} \end{bmatrix} = \begin{bmatrix} I & -\Omega_{12}\Omega_{22}^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} = J_{bb}u_{bt}$$

which has long-run covariance matrix

$$\Omega_{bb}^{+} = J_{bb}\Omega_{bb}J_{bb}' = \begin{bmatrix} \Omega_{11\cdot 2} & 0 \\ 0 & \Omega_{22} \end{bmatrix}$$

where

$$\Omega_{11\cdot 2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$$

and where we use the subscript "$b$" to signify that subscripts "1" and "2" are taken together.

Now define the following estimators of $\Gamma$ based on $\hat{y}_{1t}^{+}$:

$$\hat{\Gamma}^{+} = (\Sigma_1^T \hat{y}_{1t}^{+}x_t')(\Sigma_1^T x_t x_t')^{-1}$$

$$\bar{\Gamma}^{+} = (\Sigma_1^T \hat{y}_{1t}^{+}\bar{x}_t')(\Sigma_1^T \bar{x}_t\bar{x}_t')^{-1}$$

and the modified (bias-corrected) estimator

$$\hat{\Gamma}^{+*} = [\Sigma_1^T \hat{y}_{1t}^{+}x_t' - T(\hat{J}_{1b}\hat{\Delta}_{2b}', 0)](\Sigma_1^T x_t x_t')^{-1}$$

where

$$\hat{J}_{1b} = [I, -\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}].$$

In the case of the IVZ estimator we need also to consider the effects of the instrument innovations $u_{3t}$. Accordingly we define

$$u_{1t}^{++} = u_{1t} - \Omega_{1a}\Omega_{aa}^{-1}u_{at}$$

$$y_{1t}^{++} = y_{1t} - \Omega_{1a}\Omega_{aa}^{-1}u_{at}$$

$$\hat{y}_{1t}^{++} = y_{1t} - \hat{\Omega}_{1a}\hat{\Omega}_{aa}^{-1}\Delta y_{at}.$$

We have

$$\begin{bmatrix} u_{1t}^{++} \\ u_{at} \end{bmatrix} = \begin{bmatrix} I & -\Omega_{1a}\Omega_{aa}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{at} \end{bmatrix} = Ju_t$$

which has long-run covariance matrix

$$\Omega^{+} = J\Omega J' = \begin{bmatrix} \Omega_{11\cdot a} & 0 \\ 0 & \Omega_{aa} \end{bmatrix}$$

with

$$\Omega_{11\cdot a} = \Omega_{11} - \Omega_{1a}\Omega_{aa}^{-1}\Omega_{a1}.$$

In these formulae we use the subscript "$a$" to signify elements corresponding to "2" and "3" jointly. We further define

$$\tilde{\Gamma}^{+} = (\Sigma \hat{y}_{1t}^{++}\tilde{x}_t')(\Sigma \tilde{x}_t\tilde{x}_t')^{-1}$$
$$\tilde{\Gamma}^{+*} = [\Sigma_1^T \hat{y}_{1t}^{++}x_t - T(\hat{J}_{1c}\hat{\Delta}_{3c}', 0)](\Sigma_1^T z_t z_t')^{-1}(\Sigma_1^T z_t x_t')(\Sigma_1^T \tilde{x}_t\tilde{x}_t)^{-1}$$

where

$$\hat{J}_{1c} = [I, -\hat{\Omega}_{1a}\hat{\Omega}_{aa}^{-1}]$$

and "$c$" signifies "1" and "$a$" taken together.

From these new estimators of $\Gamma$ we construct the following $G$-statistics using the formulae as given in (13) and (14):

$$\underset{\sim}{G}_R = G_R(\hat{\Gamma}^{+*}, \hat{\Omega}_{11\cdot 2}),\ G_R(\bar{\Gamma}^{+}, \bar{\Omega}_{11\cdot 2}),\ G_R(\tilde{\Gamma}^{+*}, \tilde{\Omega}_{11\cdot a})$$

We call these new $G$-statistics fully modified Wald tests. We have:

**Theorem 5.1.** *Under* $H_0$, $\underset{\sim}{G}_R \Rightarrow \chi_g^2$.

*Remark* (*a*). Theorem 5.1 shows that the fully modified Wald tests behave as asymptotic $\chi^2$ criteria. This greatly facilitates statistical testing and eliminates the difficulties of nuisance parameter dependencies that were discussed in earlier sections.

*Remark* (*b*). The results in Theorem 5.1 are equivalent to those of Wald tests based on full maximum likelihood estimation of (2) and (3). The latter, which is discussed in detail in Phillips (1988*b*), requires formulation and full estimation of the error-generating mechanism for $u_t$. The present tests avoid this by the use of a nonparametric consistent estimate of the long-run covariance matrix $\Omega$. This estimate is used to purge the error $u_{1t}$ in the regression equation (2) of its dependence on the error processes that drive the regressors $y_{2t}$ and the instruments $y_{3t}$.

*Remark* (*c*). Note that when $u_t$ is iid $N(0, \Omega)$ we have

$$E(u_{1t} \mid u_{2t}) = \Omega_{12}\Omega_{22}^{-1}u_{2t} \tag{19}$$

and

$$u_{1t}^{+} = u_{1t} - \Omega_{12}\Omega_{22}^{-1}u_{2t} \equiv N(0, \Omega_{11\cdot 2}).$$

The estimator $\hat{A}^{+}$ is "bias-corrected" (asymptotically) and takes into account (19). In fact (19) is eliminated asymptotically by using the nonparametric estimate

$$\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\Delta y_{2t}.$$

Maximum likelihood methods, on the other hand, explicitly take (19) into account since the likelihood conditional on $u_{2t}$ involves (19) directly. As discussed in Phillips (1988*b*) this is equivalent in the present simple case to including $\Delta y_{2t}$ as a regressor in (2).

## 6. EXPERIMENTAL EVIDENCE

### 6.1. *IV, bias-corrected and fully-modified estimators: a review of existing evidence*

We have published separately (see Hansen and Phillips (1989)) a study of the small sample properties of instrumental variable, Park–Phillips bias corrected, and our fully-modified estimators via Monte Carlo simulation methods. The Data Generating Process (DGP) used in that study was adopted from the study of Banerjee *et al.* (1986). We tried six different estimation methods:

[1] OLS,
[2] Cointegrated instruments,
[3] Spurious $I(1)$ instruments,
[4] Spurious deterministic instruments,
[5] Park–Phillips bias corrected, and
[6] Fully modified.

Sample size was fixed at 100 and we varied three parameters, controlling long-run endogeneity, serial correlation, and the signal-to-noise ratio.

Comparing the uncorrected estimates [1]–[4], OLS performed best (in terms of minimum mean squared errors) for high signal-to-noise ratio, while the IV techniques performed better for low signal-to-noise ratios. The bias-corrected least squares technique dominates these estimators, but was in turn itself dominated by the fully-modified procedure.

The paper cited above also compared the distributions of estimated $t$-statistics for OLS, Park–Phillips, and fully-modified procedures. The difference is dramatic: while the variance of the Park–Phillips $t$-statistics ranged from 8 to 11, the variance of the fully-modified $t$-statistic ranged from 2 to 3.

These results support the asymptotic theory developed in Sections 3 and 4 that IV techniques, even with "spurious" instruments, can be used in $I(1)$ cointegrating regressions, yet the problems caused by endogeneity persist in IV estimation. On the other hand, the fully-modified statistics developed in Section 5 were found to perform rather well in these simulations and seem promising as candidates for empirical research.

### 6.2. *Fully modified semi-parametric estimation and Hendry error-correction parametric estimation*

#### (i) *Asymptotics*

In a recent paper, Phillips (1988*c*) has compared our fully-modified estimation procedure to the single equation error-correction methodology advocated by David Hendry in empirical time-series research. For an exposition of the latter we refer the reader to two articles by Hendry and Richard (1982, 1983). In this methodology, the starting point is a general unrestricted single equation regression of the form

$$y_{1t} = \hat{a}'y_{2t} + \hat{\gamma}'x_t + \hat{w}_t. \tag{20}$$

To relate this format to our own model we set $n_1 = 1$, $\pi = 0$ and ignore possible trend components $(k_{1t})$ in the fitted regression for ease of exposition. Elements of $x_t$ are chosen parsimoniously to render the residual $\hat{w}_t$ effectively orthogonal to lagged variables. We take a stylized version of this method in which (20) has the explicit form

$$y_{1t} = \hat{a}'y_{2t} + \textstyle\sum_{m=1}^{p} \hat{\gamma}'_{1m}\Delta y_{1t-m} + \sum_{m=0}^{p} \hat{\gamma}'_{2m}\Delta y_{2t-m} + \hat{w}_t. \tag{21}$$

If $p \to \infty$ and $p/T \to 0$ as $T \to \infty$ then (21) is an empirical attempt to asymptotically fit the following regression with distributed lags of infinite order

$$y_{1t} = a'y_{2t} + \sum_{m=1}^{\infty} \gamma_{1m}' \Delta y_{1t+m} + \sum_{m=0}^{\infty} \gamma_{2m}' \Delta y_{2t-m} + \eta_t, \tag{22}$$

where $\eta_t$ is orthogonal to the past history of $\{\Delta y_{1t-1}\}$ and $\{\Delta y_{2t}\}$. If

$$u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \begin{pmatrix} y_{1t} - a'y_{2t} \\ \Delta y_{2t} \end{pmatrix}$$

is generated by

$$u_t = \varepsilon_t + \sum_{k=1}^{\infty} \theta_k \varepsilon_{t-k} = \varepsilon_t + \theta(L)\varepsilon_{t-1}$$

with

$$E(\varepsilon_t) = 0, \qquad E(\varepsilon_t \varepsilon_t') = \begin{bmatrix} \sigma_{11} & \sigma_{21}' \\ \sigma_{21} & \Sigma_{22} \end{bmatrix}, \qquad E(\varepsilon_t \varepsilon_s') = 0, \qquad t \neq s$$

then $\eta_t$ is given by

$$\eta_t = \varepsilon_{1t} - \sigma_{21}' \Sigma_{22}^{-1} \varepsilon_{2t}$$

and the process $\xi_t = (\eta_t, u_{2t}')'$ has long-run covariance matrix

$$2\pi f_{\xi\xi}(0) = \begin{bmatrix} \sigma_{11 \cdot 2} & \sigma_{11 \cdot 2}\theta_{21}(1)' \\ \sigma_{11 \cdot 2}\theta_{21}(1) & \Omega_{22} \end{bmatrix}$$

where

$$\theta(L) = \begin{bmatrix} \theta_{11}(L) & \theta_{12}(L) \\ \theta_{21}(L) & \theta_{22}(L) \end{bmatrix}.$$

The partial-sum process constructed from $\xi_t$ then has the following asymptotic behaviour:

$$T^{-1/2} \Sigma_1^{[Tr]} \xi_t \Rightarrow \begin{pmatrix} B_\eta(r) \\ B_2(r) \end{pmatrix} = BM(2\pi f_{\xi\xi}(0)).$$

We now see that least squares on (21) gives rise to the limit theory

$$\begin{aligned} T(\hat{a} - a) &\Rightarrow \left( \int_0^1 B_2 B_2' \right)^{-1} \int_0^1 B_2 dB_\eta \\ &= \left( \int_0^1 B_2 B_2' \right)^{-1} \left[ \int_0^1 B_2 dW + \int_0^1 B_2 dB_2' \Omega_{22}^{-1} \theta_{21}(1) \right] \sigma_{11 \cdot 2} \end{aligned} \tag{23}$$

where

$$W(r) \equiv BM(\sigma_{11 \cdot 2}^{-1} - \theta_{21}(1)' \Omega_{22}^{-1} \theta_{21}(1))$$

and is independent of $B_2$. In the special case for which $\theta_{21}(1) = 0$ (23) is a mixture normal and $\hat{a}$ is asymptotically median unbiased. In general, however, (23) has a "unit root" distributional component that imports both bias and inefficiency into the limit distribution.

As shown in Phillips (1988*c*), of the two single-equation strategies for the estimation of the cointegrating vector only the semi-parametric fully modified estimator given in Section 5 achieves the asymptotic efficiency of systems maximum likelihood. The parametric Hendry approach comes very close to attaining the same asymptotic behaviour but will, in general, be both biased and inefficient (i.e. not equivalent to full maximum likelihood on the system). Asymptotic theory may be misleading in small samples. One may expect, for instance, that a parametric procedure may be superior in spite of its asymptotic bias because of poor finite sample performance of the semi-parametric procedure. We now turn to Monte Carlo methods to make an assessment of these issues.

(ii) *Finite sample simulations*

The data generating process we used was

$$y_{1t} = ay_{2t} + \pi + u_{1t}$$
$$y_{2t} = y_{2t-1} + u_{2t}, \qquad t = 1, \ldots, T$$
$$\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = u_t = \varepsilon_t + \theta\varepsilon_{t-1}, \qquad \varepsilon_t \equiv \text{iid } N(0, \Sigma).$$

We set

$$a = 2, \qquad \pi = 0, \qquad T = 50$$
$$\theta = \begin{bmatrix} 0{\cdot}3 & -0{\cdot}4 \\ \theta_{21} & 0{\cdot}6 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 1 & \sigma_{21} \\ \sigma_{21} & 1 \end{bmatrix}$$

and allowed $\theta_{21}$ and $\sigma_{21}$ to vary. This example is analyzed in Phillips (1988*c*) and is a special case of the general model discussed above. The asymptotic theory depends critically upon the parameter $\theta_{21}$.

We calculated the distributions of estimates and $t$-statistics for the cointegrating parameter obtained by OLS, Hendry and fully-modified methods. The nuisance parameters for the fully-modified procedure were estimated with a Bartlett triangular window of lag length 5, using the OLS residuals $\hat{u}_{1t}$ to calculate $\hat{\Omega}_{21}$ and $\hat{\Delta}_{21}$. For the OLS $t$-statistic we used the long-run covariance estimate $\hat{\Omega}_{11}$ to facilitate comparisons. For the Hendry procedure we included in the regression the covariates ($\Delta y_{2t}$, $\Delta y_{2t-1}$, $\Delta y_{2t-2}$, $\Delta y_{1t-1}$, $\Delta y_{1t-2}$). The fact that five covariates were chosen was designed to coincide with the choice of 5 lags for the fully modified semi-parametric corrections. (In the latter, the two-sided nature of the covariance matrix estimates generates eleven parameters. The triangular window, however, reduces the effective window size to one-half of eleven, or 5·5.) No attempt was made to alter these choices once the experiment had been started. This may be somewhat unfair to the Hendry procedure where judgment on $p$ in (21) is part of empirical practice.

The results are summarized in Tables I and II, and Figures 1 through 4. Table I records the Monte Carlo means and standard deviations of $(\hat{a} - a)$ for the ordinary least

TABLE I

*Mean (standard deviation) of $\hat{a} - a$*

| | $\theta_{21} = 0{\cdot}8$ | $\theta_{21} = 0{\cdot}4$ | $\theta_{21} = 0{\cdot}0$ |
|---|---|---|---|
| $\sigma_{21} = -0{\cdot}8$ | | | |
| OLS | −0·137 (0·125) | −0·090 (0·089) | −0·055 (0·061) |
| ECM | −0·062 (0·106) | −0·021 (0·066) | −0·003 (0·041) |
| FM | −0·025 (0·127) | −0·028 (0·079) | −0·025 (0·052) |
| $\sigma_{21} = -0{\cdot}4$ | | | |
| OLS | −0·067 (0·081) | −0·057 (0·079) | −0·040 (0·061) |
| ECM | −0·051 (0·086) | −0·030 (0·077) | −0·007 (0·060) |
| FM | −0·042 (0·094) | −0·027 (0·081) | −0·015 (0·063) |
| $\sigma_{21} = 0{\cdot}4$ | | | |
| OLS | −0·024 (0·040) | −0·020 (0·046) | −0·011 (0·050) |
| ECM | −0·023 (0·046) | −0·019 (0·053) | −0·009 (0·060) |
| FM | −0·023 (0·048) | −0·012 (0·052) | 0·004 (0·060) |
| $\sigma_{21} = 0{\cdot}8$ | | | |
| OLS | −0·015 (0·025) | −0·010 (0·028) | −0·004 (0·036) |
| ECM | −0·009 (0·024) | −0·008 (0·030) | −0·005 (0·039) |
| FM | −0·016 (0·028) | −0·005 (0·030) | 0·015 (0·043) |

TABLE II

*Mean (standard deviation) of* $t_{(\hat{a}-a)}$

| | $\theta_{21}=0{\cdot}8$ | $\theta_{21}=0{\cdot}4$ | $\theta_{21}=0{\cdot}0$ |
|---|---|---|---|
| $\sigma_{21}=-0{\cdot}8$ | | | |
| OLS | −1·616 (1·268) | −1·240 (1·105) | −0·930 (1·00) |
| ECM | −1·259 (2·040) | −0·563 (1·701) | −0·078 (1·40) |
| FM | −0·388 (1·432) | −0·449 (1·092) | −0·456 (0·896) |
| $\sigma_{21}=-0{\cdot}4$ | | | |
| OLS | −1·156 (1·32) | −0·986 (1·25) | −0·754 (1·149) |
| ECM | −1·058 (1·69) | −0·636 (1·57) | −0·163 (1·388) |
| FM | −0·729 (1·49) | −0·516 (1·35) | −0·335 (1·193) |
| $\sigma_{21}=0{\cdot}4$ | | | |
| OLS | −0·711 (1·19) | −0·520 (1·21) | −0·267 (1·24) |
| ECM | −0·664 (1·29) | −0·478 (1·34) | −0·213 (1·37) |
| FM | −0·606 (1·26) | −0·267 (1·30) | 0·096 (1·36) |
| $\sigma_{21}=0{\cdot}8$ | | | |
| OLS | −0·575 (0·955) | −0·302 (0·979) | −0·098 (1·04) |
| ECM | −0·445 (1·15) | −0·339 (1·25) | −0·184 (1·36) |
| FM | −0·519 (0·922) | −1·102 (0·962) | −0·418 (1·12) |

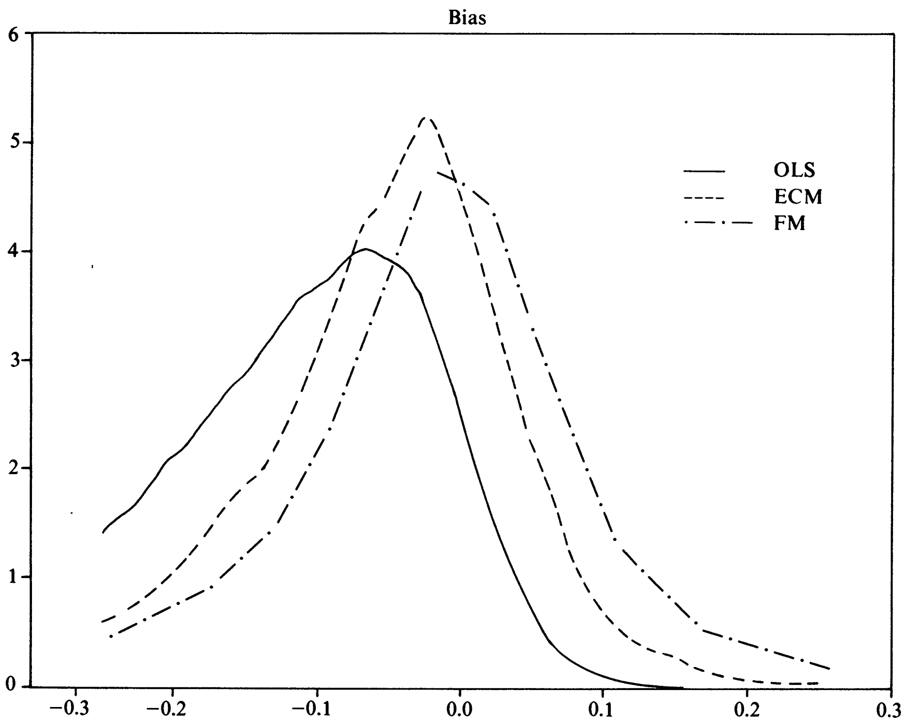


FIGURE 1

squares (OLS), Hendry error-correction (ECM) and fully-modified (FM) estimators. (All simulations used 30,000 replications.) In general, OLS is the most biased estimator. Both the ECM and FM procedures perform well. As predicted, the ECM displays moderate bias for $\theta_{21}=0{\cdot}8$, yet is virtually unbiased at $\theta_{21}=0$. The fully-modified procedure shows a small but persistent bias in finite samples and seems generally preferable to the ECM method.

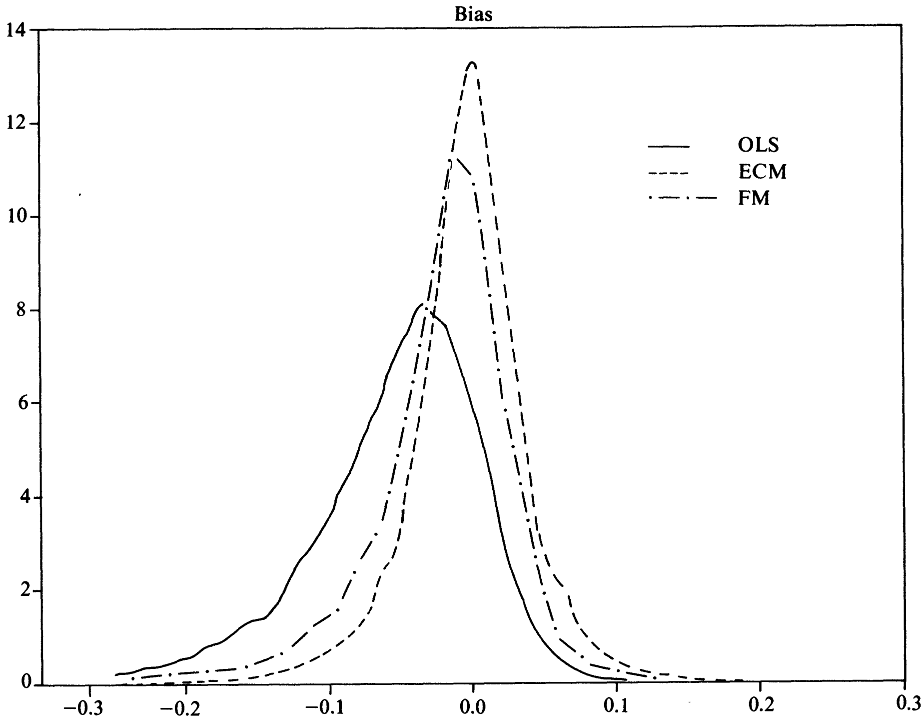


FIGURE 2

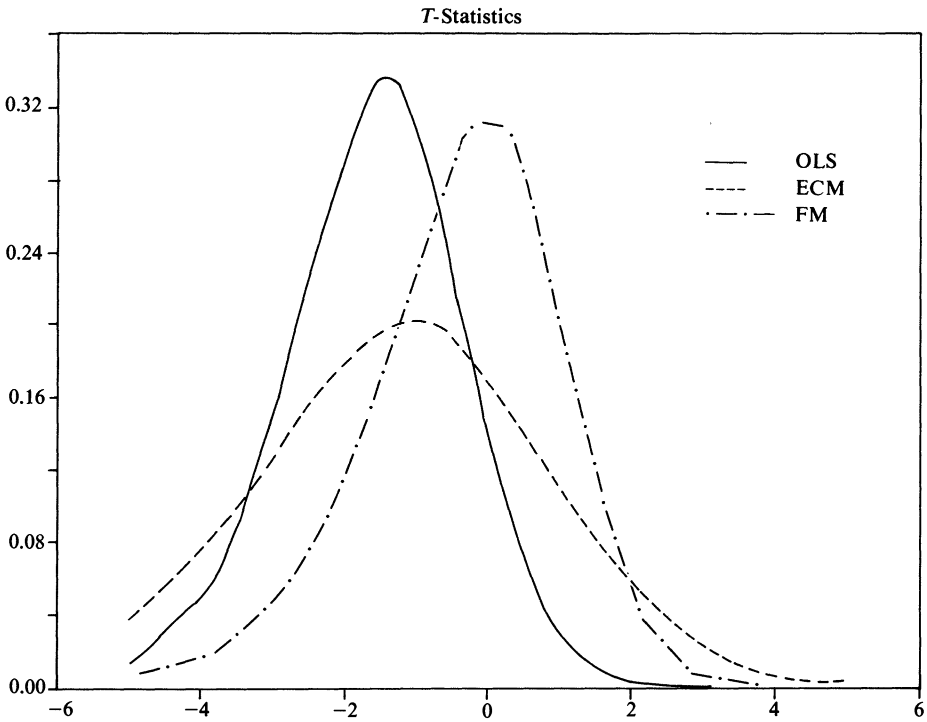


FIGURE 3

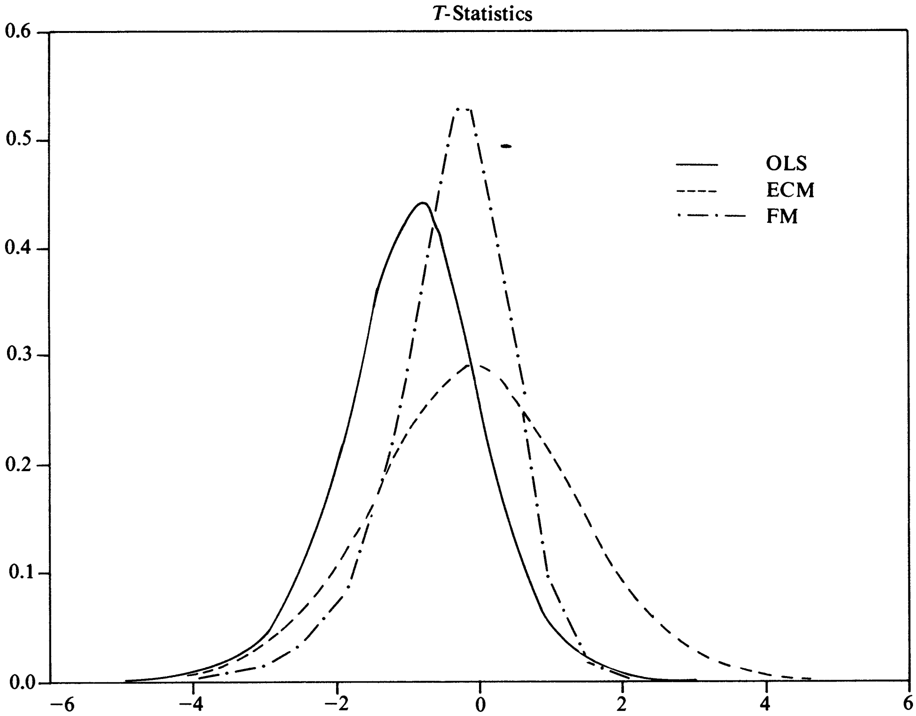


FIGURE 4

Figures 1 and 2 display estimated probability density functions (pdf's) for the estimators for $\sigma_{21} = -0{\cdot}8$ ($\theta_{21} = 0{\cdot}8$ in Figure 1 and $\theta_{21} = 0$ in Figure 2.) These densities were estimated using a normal kernel with a bandwidth of 0·2. Readers can see how the distributions display thick left tails and are fairly peaked at the mode. In Figure 1 ($\theta_{21} = 0{\cdot}8$), the FM distribution is better centred than the ECM; the reverse applies in Figure 2 ($\theta_{21} = 0$). This reinforces the theoretical results above. Thus, $\theta_{21} = 0$ is an important pre-condition for the Hendry ECM method to work well in large samples. But when this condition does hold, the parametric nature of the ECM method gives it a natural advantage over our semi-parametric approach.

In Table II are recorded the means and standard deviations of the distributions of the $t$-statistics. FM performs better than ECM in both bias and standard deviation for $\theta_{21} \neq 0$. When $\theta_{21} = 0$, however, the ECM $t$-statistic is less biased (for $\sigma_{21} < 0$) but its variance is still substantially higher. This is due to the fact that the inclusion of a limited number of lag terms has not fully eliminated serial correlation in the residuals. The FM procedure, in contrast, achieves a distribution which roughly approximates a biased standard normal. Figures 3 and 4 display estimated pdf's for the $t$-statistics under $\sigma_{21} = -0{\cdot}8$ and $\theta_{21} = 0{\cdot}8$ and 0. These estimates used a normal kernel with a bandwidth of 0·4. The figures show clearly the bias effect in the ECM distribution for $\theta_{21} > 0$, its excessive variance for all parameter values, and the relatively successful performance of the FM $t$-statistic.

Overall, both the Hendry error-correction and the fully-modified estimators seem to work quite well, considering that the sample size used is only 50. This is encouraging support for the use of asymptotic theory in integrated regressions. Some skepticism about the usefulness of asymptotic theory has emerged over the past few years after early simulation studies (such as Banerjee *et al.* (1986)) found that the super-consistency of OLS in cointegrating regressions was misleading in small samples. The implication of such studies was that asymptotic theory seemed to provide poor approximations in sample sizes that are typical in economic data. Our simulations reveal that the reverse is true. Asymptotics are not only relevant but also seem to provide good discriminatory power among differing statistical procedures even for samples as small as 50. The key ingredient in our analysis is a fully developed asymptotic distribution theory. Super-consistency in itself provides little useful information about sampling behaviour. Now that a limit distribution theory has been worked out, however, it seems fair to conclude from our simulations that it provides a reliable general guide to sampling performance, points to the most influential parameters and helps in selecting estimators and tests.

## 7. CONCLUSIONS AND FURTHER WORK

The present paper helps to complete the programme of study initiated in Phillips and Durlauf (1986) and Park and Phillips (1988, 1989). Our attention has concentrated on problems of statistical inference in multivariate linear regressions with integrated processes. By developing a theory which accommodates quite general IV estimators we have been able to isolate the sources of nuisance parameter dependencies in the limit distributions that have persisted in earlier work and have been an obstacle to the development of operational inferential procedures. These obstacles are resolved in the present treatment through semiparametric corrections that lead to a class of fully modified Wald tests. The new statistics have limiting $\chi^2$ distributions under the null and therefore greatly facilitate inference in $I(1)$ regression models. In effect, the new tests provide a semiparametric version of the optimal inference procedures (based on maximum likelihood methods) that have been developed in other ongoing work—see Phillips (1988*b*).

Our methods also provide a partial alternative to the ECM methodology that is of growing popularity in empirical research and that has been developed over a number of years in the research of Hendry (1986, 1987). The ECM methodology is parametric in nature and has proved successful in a variety of empirical applications. As shown in Phillips (1988*c*), there is a close relationship between our semiparametric fully-modified methods and the parametric ECM approach. So close, in fact, that the methods are asymptotically equivalent in some cases. In other cases (characterized by feedback among the innovations) our fully-modified methods are preferable in terms of asymptotic behaviour. The simulations that we report here in Section 6 show that these conclusions from asymptotic theory carry over remarkably well in finite samples.

Our focus of interest in this paper has been multivariate cointegrating regressions. IV techniques may be usefully employed in other integrated regressor contexts such as unit root vector autoregressions, tests for unit roots and tests for cointegration. Some of the ideas and methods suggested here are also applicable in nonlinear models in tests of nonlinear hypotheses. These are topics that the authors currently have under investigation.

## APPENDIX

As noted in the text, the estimators of $A$ in (2) are invariant to the replacement of (3) and (4) with the alternative generating mechanisms (3′) and (4′). Since estimates of $\Pi$ in (2) are not invariant to this replacement we shall take (3) and (4) to be the generating mechanisms throughout this appendix. The extra generality that applies in the case of estimates of $A$ may simply be taken for granted.

The following preliminary result will be useful. Its proof relies on simple manipulations of the type given in earlier work—see Phillips and Durlauf (1986) and Park and Phillips (1987*a*). Define $x_{Tt} = W_T^{-1}x_t$, $\tilde{x}_{Tt} = W_T^{-1}\tilde{x}_t$, $\bar{x}_{Tt} = W_T^{-1}\bar{x}_t$ and $y_t^* = y_t - (\Sigma_1^T y_t k_{1t}')(\Sigma_1^T k_{1t} k_{1t}')^{-1} k_{1t}$.

**Lemma A1**

(a) $T^{-1}\Sigma_1^T\begin{bmatrix} T^{-1/2}y_t \\ \delta_T k_t \end{bmatrix}[T^{-1/2}y_t', k_t'\delta_T] \Rightarrow \int_0^1 JJ'$

(b) $T^{-1/2}y^*_{[Tr]} \Rightarrow B^*(r)$

(c) $x_{T[Tr]} \Rightarrow J_2(r)$

(d) $\tilde{x}_{T[Tr]} \Rightarrow \tilde{J}_2(r)$

(e) $\bar{x}_{T[Tr]} \Rightarrow \bar{J}_2(r)$

(f) $T^{-1/2}\Sigma_1^T x_{Tt}u_{1t}' \Rightarrow \int_0^1 J_2 dB_1' + (\Delta_{21}', 0)'$

(g) $T^{-1/2}\Sigma_1^T \tilde{x}_{Tt}u_{1t}' \Rightarrow \int_0^1 \tilde{J}_2 dB_1' + (\Delta_{31}'F_{23}^{*\prime}, 0)'$

(h) $T^{-1/2}\Sigma_1^T \bar{x}_{Tt}u_{1t}' \Rightarrow \int_0^1 \bar{J}_2 dB_1'$

*where*

$$J(r)' = [B(r)', k(r)'], \qquad J_2(r)' = [B_2(r)', k_1(r)'].$$

**Lemma A2.** $\int_0^1 JJ' > 0$ *a.s.*

*Proof.* In view of assumption (8), we need to show that

$$\int_0^1 B^{**}B^{**\prime} > 0 \text{ a.s.} \tag{A1}$$

where

$$B^{**}(r) = B(r) - \int_0^1 Bk'\left(\int_0^1 kk'\right)^{-1} k(r).$$

Note that $B^{**}$ is itself a full rank Gaussian process. Indeed

$$B^{**}(r) \equiv N(0, \Omega v(r))$$

where

$$v(r) = r - 2\left(\int_0^1 (r \wedge s)k(s)ds\right)\left(\int_0^1 kk'\right)^{-1} k(r)$$
$$+ \int_0^1\int_0^1 (s_1 \wedge s_2)k(s_1)\left(\int_0^1 kk'\right)^{-1} k(r)k(r)'\left(\int_0^1 kk'\right)^{-1} k(s_2)ds_1 ds_2.$$

Define $W = \Omega^{-1/2}B^{**}$ which is vector "detrended" Brownian motion. Partition $W = (W_1, W_2, \ldots W_n)'$. Each element $W_i$ is independent of the other elements and is identically distributed in an $L_2[0, 1]$ Hilbert space with inner product $\int_0^1 g_1 g_2$. Observe that

$$\Pr\left\{\left|\int_0^1 WW'\right| = 0\right\}$$
$$\leq \sum_{i=1}^n \Pr\{W_i \text{ lies in the span of } (W_1, \ldots, W_{i-1}, W_{i+1}, \ldots, W_n)\}$$
$$= n\Pr\{W_1 \text{ lies in the span of } (W_2, \ldots, W_n)\}$$
$$= nE[\Pr\{W_1 \text{ lies in the span of } (W_2, \ldots, W_n) | (W_2, \ldots, W_n)\}]$$
$$\leq nE[\Pr\{W_1 \text{ lies in an } (n-1)\text{-dimensional manifold of } L_2[0, 1)\}]$$
$$= nE[0] = 0$$

which establishes (A1). ‖

*Proof of Theorem* 3.1. This is a simple application of the results in Lemmas (A1), (A2) and the continuous mapping theorem (CMT). Thus, in the case of (*b*) we have

$$T^{-1/2}(\tilde{\Gamma} - \Gamma)W_T = (T^{-1/2}\Sigma_1^T u_{1t}\tilde{x}_{Tt}')(T^{-1}\Sigma_1^T \tilde{x}_{Tt}\tilde{x}_{Tt}')^{-1}$$
$$\Rightarrow \left[\int_0^1 dB_1\tilde{J}_2' + (\Delta_{31}'F_{23}^{*\prime}, 0)\right]\left(\int_0^1 \tilde{J}_2\tilde{J}_2'\right)^{-1} \tag{A2}$$

by (d) and (g) of Lemma (A1), joint convergence and an application of the CMT. ||

*Remark.* Results such as (A2) above rely on the fact that certain random matrices are positive definite almost surely (a.s.). Here we have

$$\int_0^1 \tilde{J}_2\tilde{J}_2' = \int_0^1 J_2Z'\left(\int_0^1 ZZ'\right)^{-1}\int_0^1 ZJ_2'. \tag{A3}$$

In view of Lemma (A2) all submatrices of $\int_0^1 JJ'$ are positive definite (a.s.) and in particular

$$\int_0^1 ZZ' > 0.$$

The rank of the matrix (A3) therefore depends solely on the rank of $\int_0^1 J_2Z'$ since the order condition for IVZ, viz. $n_3 \geqq n_2$, is assumed to hold—see (1). Now since

$$\int_0^1 J_2Z' = \begin{pmatrix} \int_0^1 B_2B_3' & \int_0^1 B_2k_1' \\ \int_0^1 k_1B_3' & \int_0^1 k_1k_1' \end{pmatrix},$$

and $\int_0^1 k_1k_1' > 0$, the rank of (A3) depends on the rank of $\int_0^1 B_2^*B_3^{*\prime}$. The required result follows from the next lemma.

**Lemma A3.**

$$\operatorname{rank}\left(\int_0^1 B_2^*B_3^{*\prime}\right) = n_2 \text{ a.s.} \tag{A4}$$

$$\operatorname{rank}\left(\int_0^1 B_2^*k_2^{*\prime}\right) = n_2 \text{ a.s.} \tag{A5}$$

where $k_2^* = k_2 - \int_0^r k_2k_1'(\int_0^1 k_1k_1')^{-1}k_1$.

*Proof.* We may write (cf. Phillips (1989) Lemma 3.1)

$$B_2^*(r) \equiv \Omega_{32}'\Omega_{33}^{-1}B_3^*(r) + \Omega_{22\cdot 3}^{1/2}W_2^*$$

where

$$W_2^* = W_2(r) - \int_0^1 W_2k_1'\left(\int_0^1 k_1k_1'\right)^{-1}k_1,$$

and $W_2 \equiv BM(I_{n_2})$ and is independent of $B_3$, and $\Omega_{22\cdot 3} = \Omega_{22} - \Omega_{23}\Omega_{33}^{-1}\Omega_{32}$. Thus for any $n_3 \times n_3$ matrix $G$,

$$\int_0^1 B_2^*B_3^{*\prime}G = \Omega_{32}'\Omega_{33}^{-1}\int_0^1 B_3^*B_3^{*\prime}G + \Omega_{22\cdot 3}^{1/2}\int_0^1 W_2^*B_3^{*\prime}G. \tag{A6}$$

Lemma A2 yields that rank $(\int_0^1 B_3^*B_3^{*\prime}) = n_3$, a.s. Now set $G$ to equal

$$G = \left\{\int_0^1\int_0^1 (s_1 \wedge s_2)B_3^*(s_1)B_3^*(s_2)'ds_1ds_2\right\}^{-1/2}.$$

Each row of $\int_0^1 W_2^*B_3^{*\prime}G = \int_0^1 W_2B_3^{*\prime}G$ is independently distributed as a multivariate normal random vector with covariance matrix $I_{n_3}$, ensuring that the matrix $\int_0^1 W_2B_3^{*\prime}G$ has rank $n_2$ a.s. Moreover, since this matrix is stochastically independent of the first member on the right side of (A6) we conclude that (A6) has rank $n_2$ a.s., completing the proof of (A4).

To prove (A5) we note that

$$\int_0^1 B_2^*k_2^{*\prime} = \Omega_{22}^{1/2}\int_0^1 W_2k_2^{*\prime},$$

where $W_2 \equiv BM(I_{n_2})$. Each row of $\int_0^1 W_2k_2^{*\prime}$ is independently distributed as

$$N(0, Q), \quad Q = \int_0^1\int_0^1 k_2^*(r)k_2^*(s)'(r \wedge s)drds.$$

Since $Q>0$ by (8), $\int_0^1 W_2 k_2^{*\prime}$ has rank $n_2$ and (A5) follows. ‖

*Remark.* In an entirely analogous fashion we have

$$\operatorname{rank}\left(\int_0^1 B_2 B_3'\right) = n_2 \quad \text{a.s.}$$

$$\operatorname{rank}\left(\int_0^1 B_2 k_2'\right) = n_2 \quad \text{a.s.}$$

*Proof of Theorem* 3.2. Again, this is a simple application of the results in Lemma (A1), joint convergence and the CMT. For example, in the case of (b) we have

$$\begin{aligned}
T^{1/2}(\tilde{\Gamma}^* - \Gamma) W_T &= (T^{-1/2}\Sigma_1^T u_{1t} z_{Tt}' - (\hat{\Delta}_{31}', 0))(T^{-1}\Sigma_1^T z_{Tt} z_{Tt}')^{-1} \\
&\quad \cdot (T^{-1}\Sigma_1^T z_{Tt} x_{Tt}')(T^{-1}\Sigma_1^T \tilde{x}_{Tt}\tilde{x}_{Tt}')^{-1} \\
&\Rightarrow \left[\int_0^1 dB_1 Z' \left(\int_0^1 ZZ'\right)^{-1} \int_0^1 ZJ_2'\right]\left(\int_0^1 \tilde{J}_2\tilde{J}_2'\right)^{-1} \\
&= \left(\int_0^1 dB_1 \tilde{J}_2'\right)\left(\int_0^1 \tilde{J}_2\tilde{J}_2'\right)^{-1}
\end{aligned}$$

where

$$z_{Tt} = (T^{-1/2}y_{3t}', k_{1t}'\delta_{1T})'$$

and the result follows immediately. ‖

*Proof of Theorem* 4.1. Since $R^+$ is block diagonal as discussed in the text and $W_T$ is diagonal we have

$$\begin{aligned}
(W_T \otimes I_{n_1})(R \operatorname{vec} \underline{\Gamma} - r) &= (W_T \otimes I_{n_1}) R \operatorname{vec}(\underline{\Gamma} - \Gamma) \\
&= (W_T \otimes I_{n_1}) R^+ \operatorname{vec}(\underline{\Gamma} - \Gamma)' \\
&= R^+ \operatorname{vec}[(\underline{\Gamma} - \Gamma) W_T]' \\
&= R \operatorname{vec}[(\underline{\Gamma} - \Gamma W_T].
\end{aligned}$$

Therefore under $H_0$

$$\begin{aligned}
G_R(\underline{\Gamma}, \underline{\Omega}_{11}) &= (R \operatorname{vec}\underline{\Gamma} - r)'[R(\underline{\Omega}_{11} \otimes \underline{M})R']^{-1}(R \operatorname{vec}\underline{\Gamma} - r) \\
&= \operatorname{vec}[(\underline{\Gamma} - \Gamma) W_T]' R'(W_T^{-1} \otimes I_{n_1})[R(\underline{\Omega}_{11} \otimes \underline{M})R']^{-1}(W_T^{-1} \otimes I_{n_1}) \operatorname{vec}[(\underline{\Gamma} - \Gamma) W_T] \\
&= \operatorname{vec}[T^{1/2}(\underline{\Gamma} - \Gamma) W_T]' R'[R(\underline{\Omega}_{11} \otimes T W_T \underline{M} W_T)R']^{-1} R \operatorname{vec}[T^{-1/2}(\underline{\Gamma} - \Gamma) W_T].
\end{aligned}$$

Next,

$$T W_T \underline{M} W_T = (T^{-1} W_T^{-1} \Sigma_1^T \underline{x}_t \underline{x}_t' W_T^{-1})^{-1} \Rightarrow \left(\int_0^1 \underline{J}_2 \underline{J}_2'\right)^{-1} \tag{A7}$$

where

$$\underline{x}_t = x_t, \tilde{x}_t, \bar{x}_t$$

$$\underline{J}_2 = J_2, \tilde{J}_2, \bar{J}_2.$$

The stated results follow directly from Theorem 3.2, (A7) and the CMT. ‖

*Proof of Theorem* 5.1. We demonstrate the argument by proving $G_R(\hat{\Gamma}^{+*}, \hat{\Omega}_{11\cdot 2}) \Rightarrow \chi_g^2$. The other results follow in a similar way. Observe that

$$\begin{aligned}
T^{1/2}(\hat{\Gamma}^{+*} - \Gamma) W_T &= [T^{-1/2}\Sigma u_{1t}^+ x_{Tt}' - (\hat{J}_{1b}\hat{\Delta}_{2b}', 0)](T^{-1}\Sigma x_{Tt} x_{Tt}')^{-1} \\
&\quad - \Delta J_{12}(T^{-1/2}\Sigma u_{2t} x_{Tt}')(T^{-1}\Sigma x_{Tt} x_{Tt}')^{-1}
\end{aligned}$$

where

$$\Delta J_{12} = \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \Omega_{12}\Omega_{22}^{-1} = o_p(1).$$

Therefore

$$T^{1/2}(\hat{\Gamma}^{++*}-\Gamma)W_T = J_{1b}\left[T^{-1/2}\Sigma u_{bt}x'_{Tt}-(\Delta'_{2b},0)\right](T^{-1}\Sigma x_{Tt}x'_{Tt})^{-1}+o_p(1)$$
$$\Rightarrow J_{1b}\int_0^1 dB_b J'_2\left(\int_0^1 J_2J'_2\right)^{-1}$$

Now note that

$$\begin{bmatrix} J_{1b}B_b \\ B_2 \end{bmatrix} = J_{bb}B_b \equiv BM(\Omega^+_{bb}).$$

Thus $J_{1b}B_b$ and $B_2$ are independent Brownian motions and

$$J_{1b}\int_0^1 dB_b J'_2\left(\int_0^1 J_2J'_2\right)^{-1/2} \equiv N(0, \Omega_{11\cdot 2}\otimes I).$$

Next under $H_0$ we have

$$T^{1/2}(W_T\otimes I_{n_1})(R\operatorname{vec}\hat{\Gamma}^{++*}-r) = R\operatorname{vec}[T^{1/2}(\hat{\Gamma}^{++*}-\Gamma)W_T]$$
$$\Rightarrow R\int_0^1\left[J_{1b}dB_b\otimes\left(\int_0^1 J_2J'_2\right)^{-1}J_2\right].$$

Also

$$TR(\hat{\Omega}_{11\cdot 2}\otimes W_T\hat{M}W_T)R' \Rightarrow R\left(\Omega_{11\cdot 2}\otimes\left(\int_0^1 J_2J'_2\right)^{-1}\right)R'.$$

We deduce that

$$G_R(\hat{\Gamma}^{++*},\hat{\Omega}_{11\cdot 2}) = (R\operatorname{vec}\hat{\Gamma}^{++*}-r)'\{R(\hat{\Omega}_{11\cdot 2}\otimes\hat{M})R'\}^{-1}(R\operatorname{vec}\hat{\Gamma}^{++*}-r)$$
$$\Rightarrow\left\{\int_0^1 dB'_bJ'_{1b}\otimes J'_2\left(\int_0^1 J_2J'_2\right)^{-1}\right\}R'\left\{R\left(\Omega_{11\cdot 2}\otimes\left(\int_0^1 J_2J'_2\right)^{-1}\right)R'\right\}^{-1}$$
$$\cdot R\left\{\int_0^1 J_{1b}dB_b\otimes\left(\int_0^1 J_2J'_2\right)^{-1}J_2\right\}. \quad \text{(A9)}$$

Observe that conditional on $F_2=\sigma(B_2(r), 0\leqq r\leqq 1)$ we have

$$R\left\{\int_0^1 J_{1b}dB_b\otimes\left(\int_0^1 J_2J'_2\right)^{-1}J_2\right\}\Bigg|_{F_2} \equiv N\left(0,\left\{R\left(\Omega_{11\cdot 2}\otimes\left(\int_0^1 J_2J'_2\right)^{-1}\right)R'\right\}\right)$$

so that conditional on $F_2$ (A9) is $\chi^2_g$. Since this distribution is independent of $F_2$ the result holds unconditionally and we deduce that $G_R(\hat{\Gamma}^{+*},\hat{\Omega}_{11\cdot 2})\Rightarrow\chi^2_g$, as required. ||

*Acknowledgement.* Joon Park, Mark Watson and three referees made helpful comments on an earlier version of the paper. Our thanks also go to Glena Ames for her skill and effort in keyboarding the manuscript and to the NSF for support under Grant No. SES 8519595. The paper was first drafted in February 1988 while the first author was living on Waiheke Island and visiting the University of Auckland in New Zealand. It was revised in February 1989.

## REFERENCES

ANDREWS, D. W. K. (1987), "Heteroskedasticity and autocorrelation consistent covariance matrix estimation" (mimeograph, Yale University).

BANERJEE, A., DOLADO, J. J., HENDRY, D. F. and SMITH, G. W. (1986), "Exploring equilibrium relationships in econometrics through static models: Some Monte Carlo evidence", *Oxford Bulletin of Economics and Statistics*, **48**, 253-277.

HALL, A. (1989), "Testing for a unit root in the presence of moving average errors", *Biometrika* (forthcoming).

HANSEN, B. E. and PHILLIPS, P. C. B. (1989), "Estimation and inference in models of cointegration: A simulation study", *Advances in Econometrics* (forthcoming).

HENDRY, D. F. (1986), "Econometric modeling with cointegrated variables: An overview", *Oxford Bulletin of Economics and Statistics*, **48**, 201-212.

HENDRY, D. F. (1987), "Econometric methodology: A personal perspective," in Bewley, T. (ed.), *Advances in Econometrics, Vol. 2*. (Cambridge: Cambridge University Press).

HENDRY, D. F. and RICHARD, J-F. (1982), "On the formulation of empirical models in dynamic econometrics", *Journal of Econometrics*, **20**, 3-33.

HENDRY, D. F. and RICHARD, J-F. (1983), "The econometric analysis of economic time series (with discussion)", *International Statistical Review*, **51**, 111-163.

OULIARIS, S., PARK, J. Y. and PHILLIPS, P. C. B. (1989), "Testing for a unit root in the presence of a maintained trend", in Raj, B. (ed.), *Advances in Econometrics and Modeling.* (Amsterdam: Kluwer Academic).

NEWEY, W. and WEST, K. D. (1987). "A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica*, **55**, 703-708.

PARK, J. Y. and PHILLIPS, P. C. B. (1988), "Statistical inference in regressions with integrated processes: Part 1", *Econometric Theory*, **4**, 468-498.

PARK, J. Y. and PHILLIPS, P. C. B. (1989), "Statistical inference in regressions with integrated processes: Part 2", *Econometric Theory*, **5**, 95-132.

PHILLIPS, P. C. B. (1986*a*), "Understanding spurious regressions in econometrics", *Journal of Econometrics*, **33**, 311-340.

PHILLIPS, P. C. B. (1986*b*), "The exact distribution of the Wald statistic" *Econometrica*, **54**, 881-895.

PHILLIPS, P. C. B. (1987*a*), "Time series regression with a unit root", *Econometrica*, **55**, 277-301.

PHILLIPS, P. C. B. (1988*a*), "Multiple regression with integrated processes", in Prabhu, N. U. (ed.), *Statistical Inference from Stochastic Processes, Contemporary Mathematics Vol. 80* (Providence: American Mathematical Society).

PHILLIPS, P. C. B. (1988*b*), "Optimal inference in cointegrated systems," (Cowles Foundation Discussion Paper No. 866).

PHILLIPS, P. C. B. (1988*c*), "Reflections on econometric methodology", *Economic Record* (in press).

PHILLIPS, P. C. B. (1989), "Partially identified econometric models", *Econometric Theory*, **5**, 181-240.

PHILLIPS, P. C. B. and HANSEN, B. E. (1988), "Statistical inference in instrumental variables regression with $I(1)$ processes" (Cowles Foundation Discussion Paper No. 869).

PHILLIPS, P. C. B. and DURLAUF, S. N. (1986), "Multiple time series with integrated variables", *Review of Economic Studies*, **53**, 473-496.

PHILLIPS, P. C. B. and PERRON, P. (1987), "Testing for a unit root in time series regression", *Biometrika*, **75**, 335-346.

STOCK, J. H. (1987), "Asymptotic properties of least squares estimators of cointegrating vectors", *Econometrica*, **55**, 1035-1056.

SIMS, C. A., STOCK, J. and WATSON, M. W. (1987), "Inference in linear time series models with unit roots," (mimeograph, University of Minnesota).