ON THE INTERACTIONS OF UNIT ROOTS AND EXOGENEITY

David F. Hendry
Nuffield College,
Oxford.

March 31, 2008

Key Words and Phrases: cointegration; Monte Carlo; unit roots; weak exogeneity.

ABSTRACT

The paper considers the impact on estimation and inference of interactions between the existence of unit roots in a data generation process and the presence or absence of weak and strong exogeneity of conditioning variables for the parameters of interest in individual cointegrated linear relationships. The asymptotic distributions of estimators for single equation conditional linear relations are analyzed in conjunction with a Monte Carlo study. The results confirm the important role of weak exogeneity in single equation estimation from integrated-cointegrated data; highlight the advantages of using an asymptotic analysis to understand the complicated interactions observed; and reveal the accuracy of the limiting distributions in characterizing finite sample behaviour.

1 INTRODUCTION

Since the proposal of Engle and Granger [1987], many studies have considered estimating single-equation models linking cointegrated variables. Both static and dynamic models have been considered and a wide variety of outcomes has been found for the behaviour of alternative methods (see, *inter alia*, Phillips and Durlauf [1986], Stock [1987], Gonzalo [1989], Phillips and Loretan [1991], Kiviet and Phillips [1992] and Banerjee, Dolado, Galbraith and Hendry, 1993). This paper investigates the role of both weak and strong exogeneity for the parameters of interest in postulated cointegrating equations. In particular, the analysis highlights the effects on estimation and inference of interactions between the validity or otherwise of weak and/or strong exogeneity of contemporaneous regressors in individual cointegrated linear relationships when there are unit roots in the data generation process (DGP).

A simple bivariate DGP for two $\mathsf{I}(1)$ variables captures the salient features of the problem, and leads to eight distinct cases of interest. The asymptotic distributions of estimators and hypothesis tests for single-equation conditional linear relations are analyzed in conjunction with a Monte Carlo study. This approach reveals both the complicated interactions which result and, despite super consistency, the important role of weak exogeneity in sustaining valid single-equation inference in cointegrated processes. Of course, weak exogeneity is insufficient to sustain efficient inference, either by itself or in conjunction with a diagonal long-run covariance matrix. Nor is weak exogeneity necessary asymptotically, since the relevant failure may be due to links with variables integrated of a lower order. Conversely, the coincidence of the equation to be estimated with the conditional expectation, which in turn matches the corresponding equation of the DGP, is insufficient for valid inference in the absence of weak exogeneity. The relevance of these asymptotic results to finite samples is investigated in a Monte Carlo.

Section [2] reviews the concepts of weak and strong exogeneity. Section [3] presents the DGP and delineates the eight cases to be investigated in detail. The next section derives the covariance matrices required for the asymptotic distributions of the estimators and other statistics, and section [5] describes the vector Brownian motion processes. Section [6] obtains the limiting distributions of the estimators and tests. Section [8] provides a Monte Carlo study of the finite sample behaviour of the estimators and tests to illustrate the interactions between the presence or absence of weak exogeneity and alternative regression specifications. Section [9] discusses testing for weak exogeneity and section [10] concludes.

2 WEAK AND STRONG EXOGENEITY

The concept of weak exogeneity was proposed in Richard [1980] and analyzed by Engle, Hendry and Richard [1983], building on Koopmans [1950] and Barndorff-Nielsen [1978]: Ericsson [1992] provides an excellent exposition. The sequential joint density at time t of the two variables $\mathbf{x}_t = (y_t, z_t)'$ conditional on $\mathbf{X}_{t-1} = (\mathbf{X}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{t-1})$, when $\mathbf{X}_0$ denotes the matrix of initial conditions, is:

$$\mathsf{D}_{\mathsf{x}}\left(y_t, z_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right) \text{ for } \boldsymbol{\theta} = (\theta_1, ..., \theta_n)' \in \boldsymbol{\Theta} \subseteq \mathbb{R}^n \tag{1}$$

An investigator is interested in modelling the determination of y_t. Generally, z_t is endogenous in the framework of the joint density, so both variables must be modelled. However, under certain conditions on the system (1), z_t may not need to be analyzed to learn how y_t is determined — and the weak exogeneity of z_t for the parameters of interest in the model of y_t defines such conditions. Since a joint density can always be factorized into the product of a conditional density and a marginal density, weak exogeneity is intended to ensure that analyzing only the former sustains inference without loss of information about the parameters of interest in the system. To be empirically relevant, the analysis must hold for $t = 1, ..., T$.

Transform from the original parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ to the set $\boldsymbol{\phi} \in \boldsymbol{\Phi}$ given by:

$$\boldsymbol{\phi} = \mathbf{f}\left(\boldsymbol{\theta}\right) \text{ where } \boldsymbol{\phi} \in \boldsymbol{\Phi} \text{ and } \boldsymbol{\theta} \in \boldsymbol{\Theta} \tag{2}$$

and $\mathbf{f}\left(\cdot\right)$ defines a one-one reparameterization of θs into ϕs. Choose $\boldsymbol{\phi}$ such that $\boldsymbol{\phi}' = \left(\boldsymbol{\phi}_1', \boldsymbol{\phi}_2'\right)$, where $\boldsymbol{\phi}_i$ has n_i elements ($n_1 + n_2 = n$) corresponding to the factorization of the joint density (1) into a conditional density and a marginal density:

$$\mathsf{D}_{\mathsf{x}}\left(y_t, z_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right) = \mathsf{D}_{\mathsf{y|z}}\left(y_t \mid z_t, \mathbf{X}_{t-1}, \boldsymbol{\phi}_1\right) \mathsf{D}_{\mathsf{z}}\left(z_t \mid \mathbf{X}_{t-1}, \boldsymbol{\phi}_2\right). \tag{3}$$

Such a factorization can always be achieved if $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are defined to support it, although the resulting parameters may then be linked.

Denote the parameters of interest by $\boldsymbol{\psi}$, a vector of k elements, then z_t is weakly exogenous for $\boldsymbol{\psi}$ if:

(I) $\boldsymbol{\psi} = \mathbf{g}\left(\boldsymbol{\phi}_1\right)$ alone, where the function $\mathbf{g}\left(\boldsymbol{\phi}_1\right)$ need not be one-one; and:

(II) $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are variation free, so the parameter space of $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$ is:

$$(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2) \in \boldsymbol{\Phi}_1 \times \boldsymbol{\Phi}_2 \text{ when } \boldsymbol{\Phi}_1 \times \boldsymbol{\Phi}_2 = \{(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2) : \boldsymbol{\phi}_1 \in \boldsymbol{\Phi}_1 \text{ and } \boldsymbol{\phi}_2 \in \boldsymbol{\Phi}_2\}$$

Conditions (I) and (II) ensure that ψ can be learned from ϕ_1, and together exclude the possibility that ψ depends on ϕ_2 either directly (I), or indirectly (II), so no information about the parameters of interest can be derived from the marginal model. Hence we can learn ψ uniquely and completely from the conditional model. The concept of weak exogeneity is, therefore, close to that of S-ancillarity in Barndorff-Nielsen [1978]. A failure of either condition (I) or (II) precludes inference from the conditional model alone without loss of information, although the analysis does not specify how that information loss is manifested.

Next, z_t is strongly exogenous for ψ if z_t is weakly exogenous for ψ, and:

(III) $\mathsf{D_z}\left(z_t|\mathbf{X}_{t-1},\phi_2\right) = \mathsf{D_z}\left(z_t|\mathbf{Z}^1_{t-1},\mathbf{X}_0,\phi_2\right)$.

When (III) is satisfied, z_t does not depend upon $\mathbf{Y}_{t-1}$ so y does not Granger cause z, following Granger [1969]. This condition sustains marginalizing $\mathsf{D_z}\left(z_t|\mathbf{X}_{t-1},\phi_2\right)$ with respect to $\mathbf{Y}^1_{t-1}$, but does not concern conditioning. Consequently, Granger causality alone is neither necessary nor sufficient for weak exogeneity and cannot validate inference procedures.

Under parameter constancy, these definitions apply to the whole sample period $(1,\ldots,T)$:

$$\begin{aligned}\mathsf{D_X}\left(\mathbf{Y}^1_T,\mathbf{Z}^1_T \mid \mathbf{X}_0,\boldsymbol{\theta}\right) &= \prod_{t=1}^{T}\mathsf{D_x}\left(y_t,z_t \mid \mathbf{X}_{t-1},\boldsymbol{\theta}\right) \\ &= \prod_{t=1}^{T}\mathsf{D_{y|z}}\left(y_t \mid z_t,\mathbf{X}_{t-1},\phi_1\right)\prod_{t=1}^{T}\mathsf{D_z}\left(z_t \mid \mathbf{X}_{t-1},\phi_2\right)\end{aligned} \tag{4}$$

When z_t is strongly exogenous for ψ, then from (III):

$$\begin{aligned}\prod_{t=1}^{T}\mathsf{D_z}\left(z_t \mid \mathbf{X}_{t-1},\phi_2\right) &= \prod_{t=1}^{T}\mathsf{D_z}\left(z_t \mid \mathbf{Z}^1_{t-1},\mathbf{X}_0,\phi_2\right) \\ &= \mathsf{D_Z}\left(\mathbf{Z}^1_T \mid \mathbf{X}_0,\phi_2\right)\end{aligned} \tag{5}$$

which is the joint (marginal) density of z. Hence from (4):

$$\mathsf{D_X}\left(\mathbf{Y}^1_T,\mathbf{Z}^1_T \mid \mathbf{X}_0,\boldsymbol{\theta}\right) = \mathsf{D_{Y|Z}}\left(\mathbf{Y}^1_T \mid \mathbf{Z}^1_T,\mathbf{X}_0,\phi_1\right)\mathsf{D_Z}\left(\mathbf{Z}^1_T \mid \mathbf{X}_0,\phi_2\right). \tag{6}$$

Thus, the full-sample joint density factorizes into the product of density functions for $\mathbf{Y}^1_T|\mathbf{Z}^1_T$ and $\mathbf{Z}^1_T$, which thereby sustains full-sample conditioning. All of these results hold for y_t, z_t being vectors.

Formulations of weak exogeneity conditions and tests for various parameters of interest in cointegrated systems are discussed in Boswijk [1992a,b], Dolado [1992], Hendry and Mizon [1993], Johansen [1992a,b], Johansen and Juselius [1990] and Urbain [1992]. We establish the specific necessary conditions directly in each specification in the next section.

3 A BIVARIATE COINTEGRATED SYSTEM

We consider the following bivariate DGP for the $\mathsf{I}(1)$ vector $\mathbf{x}_t = (y_t : z_t)'$:

$$y_t = \beta z_t + w_{1t} \tag{7}$$

$$z_t = \lambda y_{t-1} + w_{2t} \tag{8}$$

where:

$$\begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \rho & 1 \end{pmatrix}\begin{pmatrix} w_{1t-1} \\ w_{2t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \tag{9}$$

and:

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \sim \mathsf{IN}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \gamma\sigma_1\sigma_2 \\ \gamma\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right] = \mathsf{IN}\left[\mathbf{0}, \mathbf{\Sigma}\right] \tag{10}$$

The DGP in (7)-(10) defines a (co-)integrated vector process in triangular form (see Phillips, 1991) which can be written in many ways, of which the following error-correction form is perhaps the most useful:

$$\begin{aligned} y_t &= \beta z_t + \epsilon_{1t} \\ \Delta z_t &= \lambda \Delta y_{t-1} + \rho\left(y_{t-1} - \beta z_{t-1}\right) + \epsilon_{2t} \end{aligned} \tag{11}$$

where $\boldsymbol{\epsilon}_t = (\epsilon_{1t} : \epsilon_{2t})'$ is distributed as in (10).

The parameters of the DGP are $(\beta, \lambda, \rho, \gamma, \sigma_1, \sigma_2)$ where:

(i) $\beta \neq 0$ determines the presence of cointegration between y_t and z_t;
(ii) $\lambda \neq 0$ determines Granger-causality of Δy on Δz;
(iii) $\rho \neq 0$ determines a failure of weak exogeneity of z_t for (β, σ_1);
(iv) $\gamma \neq 0$ determines the presence of contemporaneity; and:
(v) σ_2/σ_1 determines the signal-noise ratio for (7)

When cointegration holds, β and σ_1 can be normalized at unity without loss of generality, as is assumed henceforth and we also set σ_2 to unity in what follows to focus on the exogeneity issues. The investigator is assumed to be interested in estimating (7) and determining the parameter of interest β, which characterizes the long-run relationship between y_t and z_t.

Let $\mathcal{I}_{t-1}$ denote available lagged information (the σ -field generated by $\mathbf{X}_{t-1}$). Then, from (10) and (11), the conditional expectation of y_t given $(z_t, \mathcal{I}_{t-1})$ is:

$$\mathsf{E}\left[y_t \mid z_t, \mathcal{I}_{t-1}\right] = \beta z_t + \gamma \Delta z_t - \gamma\rho\left(y_{t-1} - \beta z_{t-1}\right) - \gamma\lambda\Delta y_{t-1} \tag{12}$$

For some parameter values in the DGP, the conditional expectation will coincide with (7), whereas for other parameter configurations, (7) and (12) will differ. In the latter case, it may be unsurprising that estimation of (7) is not fully informative, and that weak exogeneity is violated. However, an important aspect of cointegrated $\mathsf{I}(1)$ processes is that a coincidence between the equation to be estimated and the conditional expectation of the dependent variable given all available information is not sufficient to justify least squares estimation. This remains so even when the error is an innovation against the complete information set and is a surprising implication which highlights the important role of weak exogeneity in justifying single equation inference. The following analysis illustrates the practical importance of weak exogeneity failures in this context, and will stress that no serious difficulties arise from either Granger causality alone, or contemporaneity by itself, provided weak exogeneity holds.

Eight configurations of parameter values will be considered. In the first three, weak exogeneity holds or is easily obtained in a single equation specification. In the next three, weak exogeneity is violated. In the last two, weak exogeneity is again violated, but in ways which seek to refute alternative possible characterizations of the conditions for valid single equation inference in $\mathsf{I}(1)$ systems.

(a) When $\lambda = \rho = \gamma = 0$, (7) is a valid regression equation between $\mathsf{I}(1)$ variables defined by the conditional expectation:

$$\mathsf{E}\left[y_t \mid z_t, \mathcal{I}_{t-1}\right] = \beta z_t \tag{13}$$

Conditions (I)-(III) in section [2] are satisfied, so z_t is both weakly and strongly exogenous for the parameter of interest β.

(b) When $\lambda = \rho = 0$, but $\gamma \neq 0$, then (7) suffers from 'simultaneity bias' in that z_t and w_{1t} are correlated. A valid regression equation is given by the conditional expectation:

$$\mathsf{E}\left[y_t \mid z_t, \mathcal{I}_{t-1}\right] = \beta z_t + \gamma \Delta z_t \tag{14}$$

In (14), z_t is both weakly and strongly exogenous for the parameters of interest (β, γ) as (I)-(III) are satisfied. Thus, the addition of the impact variable Δz_t 'corrects' for the contemporaneous correlation between ϵ_{1t} and ϵ_{2t}, and restores valid single equation inference. Equation (7) is mis-specified but remains a cointegrating relation.

(c) When $\rho = \gamma = 0$, but $\lambda \neq 0$, then y Granger causes z. Now, z_t cannot be strongly exogenous for β, but could be weakly exogenous. The conditional expectation is:

$$\mathsf{E}\left[y_t \mid z_t, \mathcal{I}_{t-1}\right] = \beta z_t \tag{15}$$

and the second equation is uninformative about β, allowing single equation inference without loss of information on the basis of (I)-(II). Equation (7) is correctly specified.

(d) When $\lambda = \gamma = 0$, but $\rho \neq 0$, then there is a failure of weak exogeneity of z_t for β, even though the conditional expectation yields:

$$\mathsf{E}\left[y_t \mid z_t, \mathcal{I}_{t-1}\right] = \beta z_t \tag{16}$$

Nevertheless, z_t is never weakly exogenous for the parameter of interest β when $\rho \neq 0$ since:

$$\Delta z_t = \rho\left(y_{t-1} - \beta z_{t-1}\right) + \epsilon_{2t} \tag{17}$$

so a more efficient analysis is feasible by jointly estimating (7) (or (16)) and (17). Thus, we have a case where (7) coincides with the conditional expectation which is the DGP equation, but weak exogeneity is violated by a failure of (II). Below, we investigate the effects of that loss of information.

(e) When $\lambda = 0$ but $\rho \neq 0$ and $\gamma \neq 0$, weak exogeneity of z_t for β is clearly violated, as the conditional expectation is:

$$\mathsf{E}\left[y_t \mid z_t, \mathcal{I}_{t-1}\right] = \beta z_t + \gamma\left[\Delta z_t - \rho\left(y_{t-1} - \beta z_{t-1}\right)\right] \tag{18}$$

Condition (II) is violated. Moreover, in such a situation, correcting for Δz_t as in (14) need not improve matters, since $(y_{t-1} - \beta z_{t-1})$ becomes an omitted regressor in place of w_{2t}, thereby replacing a white-noise variable by an autocorrelated one. When $\rho\gamma = -1$, the error correction vanishes from (18) which thereby holds in first differences (i.e. there is a common factor of unity) but y and z remain cointegrated through the marginal process: analyzing only the conditional model would now lose all long-run information.

(f) When $\rho = 0$ but $\lambda \neq 0$ and $\gamma \neq 0$, the conditional expectation is:

$$\mathsf{E}\left[y_t \mid z_t, \mathcal{I}_{t-1}\right] = \beta z_t + \gamma\left[\Delta z_t - \lambda \Delta y_{t-1}\right] \tag{19}$$

Condition (II) is violated again and consequently z_t is not weakly exogenous for (β, γ, λ). Nor are regressions of type (14) valid either, the omitted regressor now being Δy_{t-1}.

(g) This is the special case of (e) (i.e. $\lambda = 0$) when $\rho + \gamma = 0$. As explained in section [4], the long-run covariance matrix becomes diagonal under such a condition, but weak exogeneity is violated. Dolado [1992] has argued for using the diagonality of the long-run covariance matrix as a criterion for validating single equation inference in (7) and this case is a counter-example to that claim (see sections [6] and [8]).

(h) The final case, cited by Dolado [1992], is almost the converse of (g): the long-run covariance matrix is diagonal and fully efficient inference results asymptotically although weak exogeneity fails. The model is close to (7)-(10), namely:

$$y_t = \beta z_t + \epsilon_{1t} \tag{20}$$

$$\Delta z_t = \delta\left(\Delta y_{t-1} - \beta \Delta z_{t-1}\right) + \epsilon_{2t} = u_{2t} \tag{21}$$

where $\lambda = \gamma = 0$ and $\boldsymbol{\epsilon}_t \sim \mathsf{IN}\left(\mathbf{0}, \mathbf{I}\right)$. The main difference from (7)-(8) is that the parameter of interest β enters the Δz_t equation through a term of order I(-1), since $(\Delta y_t - \beta \Delta z_t) = \Delta\epsilon_{1t}$ is the first difference of the I(0) error on (7). This induces diagonality in the long-run covariance matrix and the cross-linking due to the presence of ϵ_{1t} in both equations is asymptotically negligible. However, when $\gamma \neq 0$, we find:

$$\mathsf{E}\left[y_t \mid z_t, \mathcal{I}_{t-1}\right] = \beta z_t + \gamma\left[\Delta z_t - \delta\left(\Delta y_{t-1} - \beta \Delta z_{t-1}\right)\right] \tag{22}$$

so several parameter cross-linkages occur.

The conditional expectations are all special cases of (12), but in many instances do not coincide with (7). Even when the model is the conditional expectation, the validity of weak exogeneity depends on a joint analysis of the system: in each of (d), (e) and (f) correct specification of the conditional expectation is insufficient to sustain single equation analyses.

To keep the asymptotic and Monte Carlo analyses manageable, we focus on the eight cases (a)-(h) in all of which $\rho\lambda = 0$. This simplifies the derivations of long-run covariance matrices and vector Brownian motion processes as discussed in the next two sections. We will consider three models, namely estimating cointegrating regressions in the form of (7); these corrected for γ, in the form of (14); and the relevant conditional expectation $\mathsf{E}\left[y_t|z_t, \mathcal{I}_{t-1}\right]$. The last is always 'correctly specified' in terms of including all the relevant regressors and having an innovation error, but as stressed, the conditioning variables need not be weakly exogenous for its parameters. Such a regression yields an autoregressive-distributed lag model of the form in (12) written unrestrictedly as:

$$y_t = \alpha_0 z_t + \alpha_1 z_{t-1} + \alpha_2 y_{t-1} + \alpha_3 y_{t-2} + \nu_t \tag{23}$$

The long-run solution to (23) is $\mathsf{E}\left[y_t - \kappa z_t\right] = 0$, where:

$$\kappa = \left(\alpha_0 + \alpha_1\right) / \left(1 - \alpha_2 - \alpha_3\right) = \beta.$$

This could be derived from direct estimation of (23), or equivalently in error-correction form:

$$\Delta y_t = \left(\beta + \gamma\right)\Delta z_t - \left(\gamma\rho + 1\right)\left(y_{t-1} - \beta z_{t-1}\right) - \gamma\lambda\Delta y_{t-1} + \nu_t \tag{24}$$

It is more convenient in the Monte Carlo to use the Bewley [1979] transformation which yields (see Banerjee *et al.* [1993] for an exposition):

$$y_t = \delta_0 z_t + \delta_1 \Delta z_t + \delta_2 \Delta y_t + \delta_3 \Delta y_{t-1} + \nu_t^* \tag{25}$$

where:

$$\delta_0 = \beta; \delta_1 = \frac{\gamma\left(1 - \rho\beta\right)}{\left(1 + \gamma\rho\right)}; \delta_2 = \frac{\gamma\rho}{\left(1 + \gamma\rho\right)}; \delta_3 = \frac{-\gamma\lambda}{\left(1 + \gamma\rho\right)};\ \text{ and } \nu_t^* = \frac{\nu_t}{\left(1 + \gamma\rho\right)}.$$

Due to the endogeneity of Δy_t, (25) must be estimated by instrumental variables using y_{t-1} as the identifying instrument. Then the coefficient of z_t delivers the long-run parameter. However, because instrumental variables has no existing integer moments when just identified, the resulting estimates could potentially manifest outliers (see Sargan [1982], Maasoumi [1978] and Hendry [1990] for analyses and simulation results).

4 COVARIANCE MATRICES

Two covariance matrices of importance to the distributions in sections [5] and [6] are derived here. For a mean-zero weakly stationary stochastic process $\{\mathbf{u}_t\}$, the first is the unconditional covariance matrix $\boldsymbol{\Phi}_u$ defined by $\boldsymbol{\Phi}_u = \mathsf{E}\left[\mathbf{u}_t\mathbf{u}_t'\right]$. The second is the long-run covariance matrix $\boldsymbol{\Omega}_u$ given by:

$$\boldsymbol{\Omega}_u = \lim_{T\to\infty} \mathsf{E}\left[T^{-1}\left(\sum_{t=1}^{T}\mathbf{u}_t\right)\left(\sum_{s=1}^{T}\mathbf{u}_s'\right)\right] \tag{26}$$

which is the vector expression analogous to $\lim_{T\to\infty} \mathsf{E}\left[T^{-1}\left(\sum_{t=1}^{T} u_t\right)^2\right]$ in the scalar case, so that $\boldsymbol{\Omega}_u$ is the variance matrix of the limiting distribution of the mean of $\{\mathbf{u}_t\}$. We derive these for the DGP in section [3] when $\rho\lambda = 0$ since there is a marked simplification in that situation.

First, when $\lambda = 0$ and $\rho \neq 0$, the DGP can be written as:

$$y_t = \beta z_t + u_{1t} \tag{27}$$

$$\Delta z_t = u_{2t} \tag{28}$$

where $u_{1t} = w_{1t}$ and $u_{2t} = \Delta w_{2t}$, so that:

$$\mathbf{u}_t = \mathbf{R}\mathbf{u}_{t-1} + \boldsymbol{\epsilon}_t \ \text{ where } \ \mathbf{R} = \begin{pmatrix} 0 & 0 \\ \rho & 0 \end{pmatrix} \ \text{ and } \ \boldsymbol{\epsilon}_t \sim \mathsf{IN}\left(\mathbf{0}, \boldsymbol{\Sigma}\right) \tag{29}$$

The unconditional covariance matrix $\boldsymbol{\Phi}_u$ of $\mathbf{u}_t$ is:

$$\begin{aligned} \boldsymbol{\Phi}_u &= \mathsf{E}\left[\left(\mathbf{R}\mathbf{u}_{t-1} + \boldsymbol{\epsilon}_t\right)\left(\mathbf{u}_{t-1}'\mathbf{R}' + \boldsymbol{\epsilon}_t'\right)\right] \\ &= \mathbf{R}\mathsf{E}\left[\mathbf{u}_{t-1}\mathbf{u}_{t-1}'\right]\mathbf{R}' + \mathsf{E}\left[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\right] \\ &= \mathbf{R}\boldsymbol{\Phi}_u\mathbf{R}' + \boldsymbol{\Sigma} \end{aligned} \tag{30}$$

using stationarity. When the error variances are unity, the elements of $\boldsymbol{\Phi}_u = \{\phi_{ij}\}$ in the present bivariate case are:

$$\boldsymbol{\Phi}_u = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \rho^2\phi_{11} \end{pmatrix} + \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix} = \begin{pmatrix} 1 & \gamma \\ \gamma & 1+\rho^2 \end{pmatrix}. \tag{31}$$

Next, the long-run covariance matrix is (the appendix provides fuller details):

$$\begin{aligned} \boldsymbol{\Omega}_u &= \lim_{T\to\infty} \mathsf{E}\left[T^{-1}\left(\sum_{t=1}^{T}\mathbf{u}_t\right)\left(\sum_{s=1}^{T}\mathbf{u}_s'\right)\right] &= \boldsymbol{\Phi}_u + \boldsymbol{\Upsilon}' + \boldsymbol{\Upsilon} \\ &= \left(\mathbf{I}-\mathbf{R}\right)^{-1}\boldsymbol{\Phi}_u + \boldsymbol{\Phi}_u\left(\mathbf{I}-\mathbf{R}'\right)^{-1} - \boldsymbol{\Phi}_u &= \left(\mathbf{I}-\mathbf{R}\right)^{-1}\boldsymbol{\Sigma}\left(\mathbf{I}-\mathbf{R}'\right)^{-1} \end{aligned}$$

where the last expression uses (30). For the process in (27)-(29):

$$\boldsymbol{\Upsilon}' = \mathbf{R}\left(\mathbf{I}-\mathbf{R}\right)^{-1}\boldsymbol{\Phi}_u = \begin{pmatrix} 0 & 0 \\ \rho & \gamma\rho \end{pmatrix} \tag{32}$$

and:

$$\boldsymbol{\Omega}_u = \begin{pmatrix} 1 & \gamma+\rho \\ \gamma+\rho & 1+2\rho\gamma+\rho^2 \end{pmatrix} \tag{33}$$

Thus, $\boldsymbol{\Omega}_u$ is diagonal if and only if $\gamma + \rho = 0$, for which a sufficient condition is $\gamma = \rho = 0$.

Second, when $\rho = 0$ but $\lambda \neq 0$, the DGP can be written as:

$$y_t = \beta z_t + u_{1t} \tag{34}$$

$$\Delta z_t = \lambda\beta\Delta z_{t-1} + \lambda\Delta u_{1t-1} + u_{2t} = \zeta_{2t} = \lambda\beta\zeta_{2t-1} + \lambda\Delta u_{1t-1} + \epsilon_{2t} \tag{35}$$

where $u_{1t} = w_{1t}$ and $u_{2t} = \Delta w_{2t}$ as before. Letting $\boldsymbol{\zeta}_t = (u_{1t} : \zeta_{2t})'$:

$$\boldsymbol{\zeta}_t = \mathbf{R}_1\boldsymbol{\zeta}_{t-1} + \mathbf{R}_2\boldsymbol{\zeta}_{t-2} + \boldsymbol{\epsilon}_t \tag{36}$$

where:

$$\mathbf{R}_1 = \begin{pmatrix} 0 & 0 \\ \lambda & \beta\lambda \end{pmatrix} \text{ and } \mathbf{R}_2 = \begin{pmatrix} 0 & 0 \\ -\lambda & 0 \end{pmatrix} \tag{37}$$

and again $\boldsymbol{\epsilon}_t \sim \mathsf{IN}(\mathbf{0}, \boldsymbol{\Sigma})$. Now $\boldsymbol{\Omega}_\zeta = (\mathbf{I} - \mathbf{R}_1 - \mathbf{R}_2)^{-1}\boldsymbol{\Sigma}(\mathbf{I} - \mathbf{R}_1 - \mathbf{R}_2)^{-1\prime}$ so:

$$\boldsymbol{\Omega}_\zeta = \begin{pmatrix} 1 & \gamma(1-\beta\lambda)^{-1} \\ \gamma(1-\beta\lambda)^{-1} & (1-\beta\lambda)^{-2} \end{pmatrix} \tag{38}$$

and hence $\boldsymbol{\Omega}_\zeta$ is diagonal if and only if $\gamma = 0$. The calculation of $\boldsymbol{\Phi}_\zeta$ is best achieved by expressing (36) as a stacked first-order (4×4) system and solving as in (30), which yields:

$$\boldsymbol{\Phi}_\zeta = \begin{pmatrix} 1 & \gamma \\ \gamma & \left(1-\beta^2\lambda^2\right)^{-1}\left[1 + 2\lambda^2(1+\beta\gamma)(1-\beta\lambda)\right] \end{pmatrix} \tag{39}$$

so that:

$$\boldsymbol{\Upsilon}'_\zeta = (1-\beta\lambda)^{-1}\begin{pmatrix} 0 & 0 \\ \beta\gamma\lambda & \left(1-\beta^2\lambda^2\right)^{-1}\left[\lambda\beta - \lambda^2(1+\beta\gamma)(1-\beta\lambda)^2\right] \end{pmatrix}. \tag{40}$$

The elements of $\boldsymbol{\Upsilon}_\zeta$ are denoted by $\tau_{\zeta ij}$, and in both (32) and (40), $\tau_{21} = 0$. These formulae are used in the next two sections.

5 VECTOR BROWNIAN MOTION

Consider a general bivariate $\mathsf{I}(1)$ process:

$$\mathbf{p}_t = \mathbf{p}_{t-1} + \mathbf{v}_t \text{ where } \mathbf{p}_0 = \mathbf{0} \tag{41}$$

and $\mathbf{v}_t$ is a 'well behaved' mean-zero, weakly stationary stochastic process with unconditional covariance $\mathsf{E}[\mathbf{v}_t\mathbf{v}'_t] = \boldsymbol{\Phi}_v$ and non-singular long-run covariance $\boldsymbol{\Omega}_v = \boldsymbol{\Phi}_v + \boldsymbol{\Upsilon}_v + \boldsymbol{\Upsilon}'_v$, as in section [4]. Sufficient conditions on $\{\mathbf{v}_t\}$ for this paper are that it is a stationary, linear, mixing process with finite integer moments of up to fourth order. The analysis in this section draws on a number of results in Phillips and Durlauf [1986], Phillips [1986, 1987, 1988, 1991], and Park and Phillips [1988, 1989]: see Banerjee *et al.* [1993] for an exposition.

First, we standardize the process using $\boldsymbol{\Omega}_v^{-1} = \mathbf{K}_v\mathbf{K}'_v$ so that $\mathbf{K}'_v\boldsymbol{\Omega}_v\mathbf{K}_v = \mathbf{I}$ and hence:

$$\mathbf{m}_t = \mathbf{K}'_v\mathbf{p}_t = \mathbf{K}'_v\mathbf{p}_{t-1} + \mathbf{K}'_v\mathbf{v}_t \text{ where } \mathbf{K}'_v\mathbf{v}_t = \mathbf{e}_t. \tag{42}$$

For a general block symmetric matrix $\boldsymbol{\Omega}$ where:

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}$$

then:

$$\mathbf{K}' = \begin{pmatrix} \boldsymbol{\Omega}_{11.2}^{-\frac{1}{2}} & -\boldsymbol{\Omega}_{11.2}^{-\frac{1}{2}}\boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1} \\ \mathbf{0} & \boldsymbol{\Omega}_{22}^{-\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} \mathbf{H} & -\mathbf{HC} \\ \mathbf{0} & \boldsymbol{\Omega}_{22}^{-\frac{1}{2}} \end{pmatrix} \tag{43}$$

where $\boldsymbol{\Omega}_{11\cdot 2} = \left(\boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}\right) = \mathbf{H}^{-2}$, and $\mathbf{C} = \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}$. Then $\mathbf{m}_T/\sqrt{T}$ converges weakly to a standardized vector Brownian motion denoted $BM\left(\mathbf{I}\right)$, or more generally:

$$T^{-\frac{1}{2}}\sum_{t=1}^{[Tr]}\mathbf{e}_t \Rightarrow \mathbf{B}\left(r\right) \text{ for } r \in [0,1] \text{ as } T \to \infty \tag{44}$$

where $\Rightarrow$ denotes weak convergence and $[Tr]$ is the integer part of Tr. In the bivariate case, $\mathbf{B}\left(r\right) = \left(B_1\left(r\right) : B_2\left(r\right)\right)'$ and the $B_i\left(r\right)$ are the independent standardized Wiener processes associated with accumulating the $\{e_{it}\}$.

Consider the case when $\mathbf{e}_t \sim \mathsf{IN}\left(\mathbf{0}, \mathbf{I}\right)$: then using a component by component analysis of the standardized m_{it}, in the present bivariate case (see e.g. Banerjee and Hendry, 1992):

$$T^{-2}\sum_{t=1}^{T}\mathbf{m}_t\mathbf{m}_t' \Rightarrow \int_0^1 \mathbf{B}\left(r\right)\mathbf{B}\left(r\right)'\,\mathrm{d}r = \begin{pmatrix} \int_0^1 B_1\left(r\right)^2\mathrm{d}r & \int_0^1 B_1\left(r\right)B_2\left(r\right)\mathrm{d}r \\ \int_0^1 B_1\left(r\right)B_2\left(r\right)\mathrm{d}r & \int_0^1 B_2\left(r\right)^2\mathrm{d}r \end{pmatrix} \tag{45}$$

and:

$$T^{-1}\sum_{t=1}^{T}\mathbf{m}_{t-1}\mathbf{e}_t' \Rightarrow \int_0^1 \mathbf{B}\left(r\right)\mathrm{d}\mathbf{B}\left(r\right)' = \begin{pmatrix} \int_0^1 B_1\left(r\right)\mathrm{d}B_1\left(r\right) & \int_0^1 B_1\left(r\right)\mathrm{d}B_2\left(r\right) \\ \int_0^1 B_2\left(r\right)\mathrm{d}B_1\left(r\right) & \int_0^1 B_2\left(r\right)\mathrm{d}B_2\left(r\right) \end{pmatrix} \tag{46}$$

Next, consider expressions of the form:

$$T^{-1}\sum_{t=1}^{T}\mathbf{m}_t\mathbf{e}_t' = T^{-1}\sum_{t=1}^{T}\mathbf{m}_{t-1}\mathbf{e}_t' + T^{-1}\sum_{t=1}^{T}\mathbf{e}_t\mathbf{e}_t' \Rightarrow \int_0^1 \mathbf{B}\left(r\right)\mathrm{d}\mathbf{B}\left(r\right)' + \mathbf{I} \tag{47}$$

Thus, the error covariance matrix is added on if the cross-product under analysis is a contemporaneous rather than a lagged one.

Returning to the unstandardized and potentially autocorrelated process $\{\mathbf{v}_t\}$, then $\mathbf{p}_T/\sqrt{T}$ converges weakly to the vector Brownian motion $BM\left(\boldsymbol{\Omega}\right)$, or more generally:

$$T^{-\frac{1}{2}}\sum_{t=1}^{[Tr]}\mathbf{v}_t \Rightarrow \mathbf{V}\left(r\right) \text{ for } r \in [0,1] \text{ as } T \to \infty \tag{48}$$

Let $\mathbf{V}\left(r\right) = \left(V_1\left(r\right) : V_2\left(r\right)\right)'$ in the bivariate case, where $V_1\left(r\right)$ and $V_2\left(r\right)$ are not independent in general. Corresponding to (45)-(47), we have (see e.g. Park and Phillips [1988, 1989]):

$$T^{-2}\sum_{t=1}^{T}\mathbf{p}_t\mathbf{p}_t' \Rightarrow \int_0^1 \mathbf{V}\left(r\right)\mathbf{V}\left(r\right)'\,\mathrm{d}r \tag{49}$$

$$T^{-1}\sum_{t=1}^{T}\mathbf{p}_{t-1}\mathbf{v}_t' \Rightarrow \int_0^1 \mathbf{V}\left(r\right)\mathrm{d}\mathbf{V}\left(r\right)' + \boldsymbol{\Upsilon}_v \tag{50}$$

where $\boldsymbol{\Upsilon}_v$ is non-zero when $\{\mathbf{v}_t\}$ is autocorrelated; and as $\mathsf{E}\left[\mathbf{v}_t\mathbf{v}_t'\right] = \boldsymbol{\Phi}_v$:

$$T^{-1}\sum_{t=1}^{T}\mathbf{p}_t\mathbf{v}_t' \Rightarrow \int_0^1 \mathbf{V}\left(r\right)\mathrm{d}\mathbf{V}\left(r\right)' + \boldsymbol{\Upsilon}_v + \boldsymbol{\Phi}_v \tag{51}$$

The vector Brownian motion could be standardized using $\mathbf{B}(r) = \mathbf{K}_v'\mathbf{V}(r)$, such that $\mathbf{B}(r)$ is $BM(\mathbf{I})$. Multiplying out $\mathbf{K}_v'\mathbf{V}(r)$, we have:

$$\mathbf{K}_v'\mathbf{V}(r) = \begin{pmatrix} \mathbf{H} & -\mathbf{HC} \\ \mathbf{0} & \mathbf{\Omega}_{22}^{-\frac{1}{2}} \end{pmatrix} \begin{pmatrix} \mathbf{V}_1(r) \\ \mathbf{V}_2(r) \end{pmatrix} = \begin{pmatrix} \mathbf{HV}_1(r) - \mathbf{HCV}_2(r) \\ \mathbf{\Omega}_{22}^{-\frac{1}{2}}\mathbf{V}_2(r) \end{pmatrix} \tag{52}$$

Thus, in the bivariate case:

$$B_1(r) = h(V_1(r) - cV_2(r)) \text{ and } B_2(r) = \sigma_2^{-1/2}V_2(r) \tag{53}$$

Since the standardized vector Brownian motion $\mathbf{B}(r)$ has independent components, $\mathbf{V}_2(r)$ and $\left(\mathbf{V}_1(r) - \mathbf{\Omega}_{12}\mathbf{\Omega}_{22}^{-1}\mathbf{V}_2(r)\right)$ are independent also. Applying these results, we have the following for $\{\mathbf{u}_t\}$ and $\{\boldsymbol{\zeta}_t\}$ defined respectively by (29) and (36) in section [4]. First, when $\lambda = 0$, from (33):

$$\mathbf{\Omega}_u^{-1} = \frac{1}{(1-\gamma^2)} \begin{pmatrix} 1 + 2\rho\gamma + \rho^2 & -(\gamma+\rho) \\ -(\gamma+\rho) & 1 \end{pmatrix}$$

so that:

$$\mathbf{K}_u' = \begin{pmatrix} \left(\frac{\omega_{22}}{\eta}\right)^{\frac{1}{2}} & -(\gamma+\rho)(\omega_{22}\eta)^{-\frac{1}{2}} \\ 0 & \omega_{22}^{-\frac{1}{2}} \end{pmatrix} \tag{54}$$

where $\eta = (1-\gamma^2)$ and $\omega_{22} = (1 + 2\rho\gamma + \rho^2)$ and hence:

$$B_1(r) = \left(\frac{\omega_{22}}{\eta}\right)^{\frac{1}{2}} \left(V_1(r) - \left[\frac{(\gamma+\rho)}{\omega_{22}}\right] V_2(r)\right) \text{ and } B_2(r) = \frac{V_2(r)}{\sqrt{\omega_{22}}} \tag{55}$$

Next, when $\rho = 0$ but $\lambda \neq 0$, from (38):

$$\mathbf{K}_\zeta' = \begin{pmatrix} \eta^{-\frac{1}{2}} & -\gamma(1-\beta\lambda)\eta^{-\frac{1}{2}} \\ 0 & (1-\beta\lambda) \end{pmatrix} \tag{56}$$

again yielding two independent standardized processes, where:

$$B_1(r) = \eta^{-\frac{1}{2}} \left(V_1(r) - \left[\frac{\eta}{(1-\beta\lambda)}\right] V_2(r)\right) \tag{57}$$

with $\omega_{22} = 1/(1-\beta\lambda)^2$ and $B_2(r)$ as in (55).

6 LIMITING DISTRIBUTIONS

6.1 OLS Estimation: $\lambda = 0$

We begin by considering the ordinary least-squares (OLS) estimator of (7) when $\lambda = 0$. Subsequent derivations follow similar lines. Since $w_{1t} = \epsilon_{1t}$ in (7), we use the results in section [5], and a functional central limit theorem (see e.g. Phillips [1986, 1987]). The first step follows from (49) and (51), since $\tau_{21} = 0$ from (32); the second since $\phi_{u21} = \gamma$ from (31); and the last line requires:

$$V_1(r) = \left(\frac{\eta}{\omega_{22}}\right)^{\frac{1}{2}} B_1(r) + \left(\frac{\gamma+\rho}{\omega_{22}}\right) V_2(r) \text{ and } B_2(r) = \frac{V_2(r)}{\sqrt{\omega_{22}}}$$

from (55):

$$
\begin{aligned}
T\left(\hat{\beta}-\beta\right) &= \left(T^{-2}\sum_{t=1}^{T} z_t^2\right)^{-1}\left(T^{-1}\sum_{t=1}^{T} z_t \epsilon_{1t}\right) \\
&\Rightarrow \left(\int_0^1 V_2\left(r\right)^2 \mathrm{d}r\right)^{-1}\left(\int_0^1 V_2\left(r\right)\mathrm{d}V_1\left(r\right)+\phi_{u21}\right) \\
&= \left(\omega_{22}^{-1}\int_0^1 V_2\left(r\right)^2 \mathrm{d}r\right)^{-1}\left(\omega_{22}^{-1}\int_0^1 V_2\left(r\right)\mathrm{d}V_1\left(r\right)+\frac{\gamma}{\omega_{22}}\right) \qquad (58)\\
&= \left(\int_0^1 B_2\left(r\right)^2 \mathrm{d}r\right)^{-1}\left[\left(\frac{\eta^{\frac{1}{2}}}{\omega_{22}}\right)\int_0^1 B_2\left(r\right)\mathrm{d}B_1\left(r\right)\right] \\
&\quad + \left(\int_0^1 B_2\left(r\right)^2 \mathrm{d}r\right)^{-1}\left[\left(\frac{\gamma+\rho}{\omega_{22}}\right)\int_0^1 B_2\left(r\right)\mathrm{d}B_2\left(r\right)+\frac{\gamma}{\omega_{22}}\right].
\end{aligned}
$$

OLS remains super consistent, but the Monte Carlo results will demonstrate the relevance to finite sample inference of the normalized limiting distributions derived from (58).

Several of the components of (58) have distributions which are related to the normal distribution. First, since $B_1\left(r\right)$ and $B_2\left(r\right)$ are independent Wiener processes, conditionally on $B_2\left(\cdot\right)$ (see Park and Phillips [1988, 1989] or Banerjee and Hendry [1992] for an exposition):

$$\left(\int_0^1 B_2\left(r\right)^2 \mathrm{d}r\right)^{-1}\int_0^1 B_2\left(r\right)\mathrm{d}B_1\left(r\right) \sim \mathsf{N}\left[0, \left(\int_0^1 B_2\left(r\right)^2 \mathrm{d}r\right)^{-1}\right] \qquad (59)$$

Consequently, the distribution in (59) is a linear mixture of normals centered on zero. Further:

$$\left(\int_0^1 B_2\left(r\right)^2 \mathrm{d}r\right)^{-\frac{1}{2}}\left(\int_0^1 B_2\left(r\right)\mathrm{d}B_1\left(r\right)\right) \sim \mathsf{N}\left[0,1\right] \qquad (60)$$

and hence (60) holds unconditionally. Second (see Dickey and Fuller [1979, 1981]):

$$DF_{\alpha} = \left(\int_0^1 B_2\left(r\right)^2 \mathrm{d}r\right)^{-1}\int_0^1 B_2\left(r\right)\mathrm{d}B_2\left(r\right) \qquad (61)$$

and

$$DF_{\mathsf{t}} = \left(\int_0^1 B_2\left(r\right)^2 \mathrm{d}r\right)^{-\frac{1}{2}}\int_0^1 B_2\left(r\right)\mathrm{d}B_2\left(r\right) \qquad (62)$$

are the Dickey-Fuller $T\left(\hat{\alpha}-1\right)$ and t-distributions respectively for testing for a unit root in the univariate marginal process for $\{z_t\}$. Finally, (see Fuller, 1976):

$$\int_0^1 B_2\left(r\right)\mathrm{d}B_2\left(r\right) \sim \frac{1}{2}\left(\chi^2\left(1\right)-1\right)$$

so the numerator in (61) is $\frac{1}{2}\left(\chi^2\left(1\right)-1\right)$. The fact that $\mathsf{P}\left(\chi^2\left(1\right)\leq 1\right)\simeq 0.7$ suggests that the last term in (58) will impart a negative shift to the distribution.

To understand all the implications of (58), we consider the special cases (a), (b), (d), (e) and (g) in turn, then examine (c) and (f) (where $\lambda \neq 0$) and finally (h). Table I summarizes the states of nature under consideration, using * to denote a non-zero value.

Table I: Parameter values

parameter	a	b	c	d	e	f	g	h
λ	0	0	*	0	0	*	0	0
ρ	0	0	0	*	*	0	$-\gamma$	δ
γ	0	*	0	0	*	*	*	0

(a) When $\lambda = \rho = \gamma = 0$, then $\eta = \omega_{22} = 1$ and $\phi_{u12} = 0$, so (58) collapses to:

$$T\left(\hat{\beta} - \beta\right) \Rightarrow \left(\int_0^1 B_2(r)^2 \,\mathrm{d}r\right)^{-1} \left(\int_0^1 B_2(r)\,\mathrm{d}B_1(r)\right) \tag{63}$$

where from (59):

$$T\left(\hat{\beta} - \beta\right) \sim \mathsf{N}\left[0, \left(\int_0^1 B_2(r)^2 \,dr\right)^{-1}\right] \tag{64}$$

Thus, the distribution of $T\left(\hat{\beta} - \beta\right)$ in (64) is a mixture of normals centered on zero and any finite sample bias must be $o_p\left(T^{-1}\right)$. Although the limiting variance is stochastic, since:

$$\left(T^{-2}\sum_{t=1}^{T} z_t^2\right) \Rightarrow \omega_{22}\left(\int_0^1 B_2(r)^2 \,\mathrm{d}r\right) \tag{65}$$

and $\sigma_1^2 = \omega_{22} = 1$, the conventional coefficient standard error (ESE) should accurately estimate the sampling standard deviation (SSD) in (64). Below, we consider the ratio of SSD to ESE.

(b) When $\lambda = \rho = 0$ but $\gamma \neq 0$, then $\eta = \left(1 - \gamma^2\right)$, $\omega_{22} = 1$ and $\phi_{u12} = \gamma$ so (58) becomes for $T\left(\hat{\beta} - \beta\right)$:

$$\left(\int_0^1 B_2(r)^2 \,\mathrm{d}r\right)^{-1} \left[\eta^{\frac{1}{2}} \int_0^1 B_2(r)\,\mathrm{d}B_1(r) + \gamma\left(\int_0^1 B_2(r)\,\mathrm{d}B_2(r) + 1\right)\right] \tag{66}$$

However, the final term inside $[\cdot]$ in (66) is $\frac{1}{2}\gamma\left(\chi^2(1) + 1\right)$ which has the same sign as γ. The first term in $[\cdot]$ remains a mixture of normals but scaled by $\sqrt{(1-\gamma^2)}$, so relative to (a), when $\gamma > 0$, the distribution is shifted rightwards and is non-normal. The conventional coefficient standard error no longer correctly estimates the sampling standard deviation, and hypothesis tests could be distorted. These, and the estimation of (14), are considered below.

(d) When $\lambda = \gamma = 0$ but $\rho \neq 0$, then $\eta = 1$, $\omega_{22} = 1 + \rho^2$ and $\phi_{u12} = 0$ so (58) becomes for $T\left(\hat{\beta} - \beta\right)$:

$$\left(\int_0^1 B_2(r)^2 \,\mathrm{d}r\right)^{-1} \frac{1}{(1+\rho^2)} \left[\int_0^1 B_2(r)\,\mathrm{d}B_1(r) + \rho\int_0^1 B_2(r)\,\mathrm{d}B_2(r)\right] \tag{67}$$

This is a mixture of the normal and Dickey-Fuller distributions, and for sufficiently large $\rho > 0$, the last term will impart a negative shift to the distribution (conversely for negative ρ). Again, inference is liable to be distorted.

(e) When $\lambda = 0$ but $\gamma \neq 0$ and $\rho \neq 0$, then $\eta = \left(1 - \gamma^2\right)$, $\omega_{22} = 1 + 2\gamma\rho + \rho^2$ and $\phi_{u12} = \gamma$ so (58) remains as it is. The terms inside $\omega_{22}^{-1}\left[\frac{1}{2}(\gamma + \rho)\left(\chi^2(1) - 1\right) + \gamma\right]$ will partially offset each other, so having both a failure of weak exogeneity and simultaneity may induce less distortion than either alone. This matches the conclusions in (b) and (d) that $(\gamma \neq 0, \rho = 0)$ and $(\gamma = 0, \rho \neq 0)$ impart different directions of skewness to the limiting distribution.

(g) When $\lambda = 0$ but $\gamma = -\rho$, so $\gamma + \rho = 0$ we have an example of type (e) chosen to create a diagonal long-run covariance matrix while violating weak exogeneity. From (58), a bias is anticipated despite the independence of $V_1(r)$ and $V_2(r)$. This bias is in accord with (18), which also explains why an augmented regression should not improve matters even though it might be thought to do so from (58) being written as (at the point $\gamma = -\rho$):

$$T\left(\hat{\beta} - \beta\right) \Rightarrow \left(\int_0^1 B_2(r)^2 \,\mathrm{d}r\right)^{-1} \left[\left(\frac{\eta^{\frac{1}{2}}}{\omega_{22}}\right)\int_0^1 B_2(r)\,\mathrm{d}B_1(r) + \frac{\gamma}{\omega_{22}}\right]. \tag{68}$$

The limiting non-normality arises because $\mathbf{\Phi}$ is not diagonal even though $\mathbf{\Omega}$ is. These cases reveal the separate roles played by weak exogeneity, simultaneity and the diagonality of the long-run covariance matrix in determining the non-centrality of the limiting distribution and the independence of $B_1(r)$ and $B_2(r)$.

If an investigator knew that $\gamma = -\rho$, full information maximum likelihood estimation of (12) and the marginal process from (11) would be feasible due to the orthogonality of the errors and the non-linear cross-equation restrictions, since the system takes the form:

$$\begin{aligned} \Delta y_t &= (\beta - \rho) \Delta z_t + \left(\rho^2 - 1\right)(y_{t-1} - \beta z_{t-1}) + \xi_{1t} \\ \Delta z_t &= \rho (y_{t-1} - \beta z_{t-1}) + \epsilon_{2t} \end{aligned} \tag{69}$$

where $\xi_{1t} = \epsilon_{1t} - \mathsf{E}[\epsilon_{1t}|\epsilon_{2t}]$ so $\mathsf{E}[\xi_{1t}\epsilon_{2t}] = 0$.

6.2 Augmented OLS Estimation: $\lambda = 0$

The possible failure of augmentation just noted is important in the context of using regressions such as equation (14) in which an attempt is made to correct for 'simultaneity' by adding Δz_t as a regressor. In terms of the original DGP (7)-(10) when $\lambda = 0$, an alternative reparameterization is:

$$y_t = \beta z_t + \gamma \Delta z_t + w_{1t} \tag{70}$$

$$\Delta z_t = w_{2t} \tag{71}$$

where:

$$\begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \rho & 0 \end{pmatrix} \begin{pmatrix} w_{1t-1} \\ w_{2t-1} \end{pmatrix} + \begin{pmatrix} \xi_{1t} \\ \xi_{2t} \end{pmatrix} \tag{72}$$

and $\boldsymbol{\xi}_t \sim \mathsf{IN}[\mathbf{0}, \mathbf{I}]$ with ξ_{1t} as above. Consequently, from section [4]:

$$\mathbf{\Omega}_w = \begin{pmatrix} 1 & \rho \\ \rho & 1+\rho^2 \end{pmatrix}, \ \mathbf{\Phi}_w = \begin{pmatrix} 1 & 0 \\ 0 & 1+\rho^2 \end{pmatrix} \text{ and } \mathbf{\Upsilon}_w = \begin{pmatrix} 0 & \rho \\ 0 & 0 \end{pmatrix} \tag{73}$$

so that $\eta = 1,\ \omega_{22} = 1 + \rho^2$ and $\phi_{u12} = \tau_{21} = 0$. The parameter estimates of (14) are:

$$\begin{aligned} \begin{pmatrix} \sqrt{T}(\hat{\gamma} - \gamma) \\ T\left(\hat{\beta} - \beta\right) \end{pmatrix} &= \begin{pmatrix} T^{-1}\sum_{t=1}^{T} \left(\Delta z_t^2\right) & T^{-\frac{3}{2}} \sum_{t=1}^{T} z_t \Delta z_t \\ T^{-\frac{3}{2}} \sum_{t=1}^{T} z_t \Delta z_t & T^{-2}\sum_{t=1}^{T} z_t^2 \end{pmatrix}^{-1} \begin{pmatrix} T^{-\frac{1}{2}} \sum_{t=1}^{T} \Delta z_t \xi_{1t} \\ T^{-1}\sum_{t=1}^{T} z_t \xi_{1t} \end{pmatrix} \\ &\Rightarrow \begin{pmatrix} \sigma^2_{\Delta z} & 0 \\ 0 & \int_0^1 W_2(r)^2 \, \mathrm{d}r \end{pmatrix}^{-1} \begin{pmatrix} \mathsf{N}\left[0, \sigma^2_{\Delta z}\right] \\ \int_0^1 W_2(r) \, \mathrm{d}W_1(r) \end{pmatrix} \end{aligned} \tag{74}$$

where $W_1(r),\ W_2(r)$ are the Wiener processes associated with accumulating the w_{it}. Thus $T\left(\hat{\beta} - \beta\right)$:

$$\begin{aligned} &\Rightarrow \left(\int_0^1 W_2(r)^2 \, \mathrm{d}r\right)^{-1} \left[\int_0^1 W_2(r) \, \mathrm{d}W_1(r)\right] \\ &= \left(\int_0^1 B_2(r)^2 \, \mathrm{d}r\right)^{-1} \frac{1}{(1+\rho^2)} \left[\int_0^1 B_2(r) \, \mathrm{d}B_1(r) + \rho \int_0^1 B_2(r) \, \mathrm{d}B_2(r)\right] \end{aligned} \tag{75}$$

since $W_1(r) = \omega_{22}^{-\frac{1}{2}} B_1(r) + (\rho/\omega_{22}) W_2(r)$. This outcome is now compared to the corresponding result for estimating (7) in (58).

(a) Since $\gamma = 0$ and $\rho = 0$, there can be no gain from fitting (14) relative to (7), since (75) coincides with (63) when $\rho = 0$: equally, asymptotically there is no loss.
(b) Since $\rho = 0$ and $\gamma \neq 0$, a distinct improvement will occur relative to (66) because the distribution in (75) reverts to a linear mixture of normals. This matches the weak exogeneity of z_t for the parameters of the conditional model (14). Thus, inference proceeds as in (a), so by itself $\gamma \neq 0$ does not induce a weak exogeneity failure.
(d) As $\gamma = 0$ anyway in this case, (75) is the same as (67) so there is no change from the conclusions for (7).
(e) Now both $\rho \neq 0$ and $\gamma \neq 0$ so the (left) skewness in the distribution could be exacerbated by correcting one of the two 'problems'. Conversely, as (75) coincides with (67), the correction has been 'successful' and asymptotically mimics $\gamma = 0$.
(g) The same implications hold as in (e).

6.3 OLS Estimation: $\rho = 0$

Now we turn to the cases where $\rho = 0$ but $\lambda \neq 0$: these have weak but not strong exogeneity. In such a situation, a similar analysis to (58) applies but using $\mathbf{\Omega}_\zeta$ in (38). Since $\omega_{22} = (1-\beta\lambda)^{-2}$, $\eta = \left(1-\gamma^2\right)$ and $\phi_{\zeta 12} = \gamma$, estimating (7) by OLS yields:

$$\begin{aligned} T\left(\hat{\beta}-\beta\right) &= \left(T^{-2}\sum_{t=1}^{T} z_t^2\right)^{-1}\left(T^{-1}\sum_{t=1}^{T} z_t \epsilon_{1t}\right) \\ &\Rightarrow \left(\int_0^1 V_2\left(r\right)^2 \mathrm{d}r\right)^{-1}\left(\int_0^1 V_2\left(r\right)\mathrm{d}V_1\left(r\right)+\phi_{\zeta 12}\right) \\ &= (1-\beta\lambda)\left(\int_0^1 B_2\left(r\right)^2 \mathrm{d}r\right)^{-1}\left[\left(1-\gamma^2\right)^{\frac{1}{2}}\int_0^1 B_2\left(r\right)\mathrm{d}B_1\left(r\right)\right] \\ &\quad +\gamma\left(1-\beta\lambda\right)\left(\int_0^1 B_2\left(r\right)^2 \mathrm{d}r\right)^{-1}\left(\int_0^1 B_2\left(r\right)\mathrm{d}B_2\left(r\right)+\left(1-\beta\lambda\right)\right) \end{aligned} \quad (76)$$

since $\tau_{21} = 0$ again, when:

$$V_1\left(r\right) = \left(1-\gamma^2\right)^{\frac{1}{2}} B_1\left(r\right)+\gamma\left(1-\beta\lambda\right)V_2\left(r\right) \text{ and } B_2\left(r\right) = V_2\left(r\right)\left(1-\beta\lambda\right)$$

from (57). As with (58), we find a mixture of a conditionally normal and a Dickey-Fuller distribution in (76), as well as a non-centrality effect when $\gamma \neq 0$. We can now consider the special cases (c) and (f).
(c) When $\rho = \gamma = 0$, but $\lambda \neq 0$, for estimating (7), (76) becomes:

$$T\left(\hat{\beta}-\beta\right) \Rightarrow (1-\beta\lambda)\left(\int_0^1 B_2\left(r\right)^2 \mathrm{d}r\right)^{-1}\left(\int_0^1 B_2\left(r\right)\mathrm{d}B_1\left(r\right)\right). \quad (77)$$

Thus, despite the fact that y Granger causes z, since $B_1\left(r\right)$ and $B_2\left(r\right)$ are independent Wiener processes, the distribution of the OLS estimator is a mixture of normals, centered on zero: the violation of strong exogeneity, when weak exogeneity is maintained, does not seriously affect inference in this unit-root model. Using (60) and matching (64), conditional on B_2:

$$T\left(\hat{\beta}-\beta\right) \sim \mathsf{N}\left[0, \left(\omega_{22}\int_0^1 B_2\left(r\right)^2 \mathrm{d}r\right)^{-1}\right] \quad (78)$$

so from (65), the conventional coefficient standard error correctly estimates the sampling standard deviation. However:
(f) When $\rho = 0$, but $\lambda \neq 0$ and $\gamma \neq 0$, then weak exogeneity is violated and (76) reveals that the limiting distribution of the OLS estimator is non-normal, and is not centered on zero, although the last two terms could partially offset each other. Inference could be seriously distorted, as investigated in [8].

6.4 Augmented OLS Estimation: $\rho = 0$

In the present setting, correcting for Δz_t as in (14) could again have an important effect. Reparameterize the DGP when $\rho = 0$ as:

$$y_t = \beta z_t + \gamma^* \Delta z_t + s_{1t} \tag{79}$$

$$\Delta z_t = s_{2t} \tag{80}$$

where $\gamma^* = \gamma / (1 + \varphi)$, $s_{1t} = e_{1t} - \lambda\gamma\Delta y_{t-1}$ and $s_{2t} = \epsilon_{2t} + \lambda\Delta y_{t-1}$. Then:

$$\begin{aligned} \begin{pmatrix} s_{1t} \\ s_{2t} \end{pmatrix} &= \lambda\beta \begin{pmatrix} 0 & -\gamma \\ 0 & 1 \end{pmatrix} \begin{pmatrix} s_{1t-1} \\ s_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \\ &+\lambda \begin{pmatrix} -\gamma & -\gamma^2 \\ 1 & \gamma \end{pmatrix} \begin{pmatrix} e_{1t-1} \\ e_{2t-1} \end{pmatrix} - \lambda \begin{pmatrix} -\gamma & -\gamma^2 \\ 1 & \gamma \end{pmatrix} \begin{pmatrix} e_{1t-2} \\ e_{2t-2} \end{pmatrix} \end{aligned} \tag{81}$$

where $e_{1t} = \epsilon_{1t} - \gamma^* \epsilon_{2t}$, so that $\mathsf{E}[e_{1t}\epsilon_{2t}] = 0$ and $\mathsf{E}\left[e_{1t}^2\right] = \left(1 - \gamma^2\right)$. From the appendix analysis:

$$\mathbf{\Omega}_s = (1 - \beta\lambda)^{-2} \begin{pmatrix} (1 - \beta\lambda)^2 \left(1 - \gamma^2\right) + \gamma^2\beta^2\lambda^2 & -\gamma\beta\lambda \\ -\gamma\beta\lambda & 1 \end{pmatrix} \tag{82}$$

and:

$$\mathbf{\Phi}_s = \begin{pmatrix} 1 - \gamma^2 (1 - \varphi) & -\gamma\varphi \\ -\gamma\varphi & 1 + \varphi \end{pmatrix} \tag{83}$$

when $\varphi = \lambda^2 \left(1 - \lambda^2\beta^2\right)^{-1} \left[\beta^2 + 2(1 - \lambda\beta)(1 + \beta\gamma)\right]$. The structure of $\mathbf{s}_t$ as a non-diagonal first-order autoregression with a second-order moving-average error is such that $\mathbf{\Upsilon}_s$ entails $\tau_{21} \neq 0$ when $\gamma\lambda \neq 0$ since:

$$\tau_{21} = \frac{-\gamma\lambda \left[\beta - \lambda (1 - \lambda\beta)^2 (1 + \beta\gamma)\right]}{\left[(1 - \lambda\beta)^2 (1 + \lambda\beta)\right]}.$$

Consequently:

$$\begin{aligned} \begin{pmatrix} \sqrt{T}(\hat{\gamma} - \gamma) \\ T\left(\hat{\beta} - \beta\right) \end{pmatrix} &= \begin{pmatrix} T^{-1}\sum_{t=1}^{T}\left(\Delta z_t^2\right) & T^{-\frac{3}{2}}\sum_{t=1}^{T} z_t\Delta z_t \\ T^{-\frac{3}{2}}\sum_{t=1}^{T} z_t\Delta z_t & T^{-2}\sum_{t=1}^{T} z_t^2 \end{pmatrix}^{-1} \begin{pmatrix} T^{-\frac{1}{2}}\sum_{t=1}^{T}\Delta z_t s_{1t} \\ T^{-1}\sum_{t=1}^{T} z_t s_{1t} \end{pmatrix} \\ &\Rightarrow \begin{pmatrix} \sigma_{\Delta z}^2 & 0 \\ 0 & \int_0^1 W_2(r)^2 \mathrm{d}r \end{pmatrix}^{-1} \begin{pmatrix} \mathsf{N}\left[0, \delta^2\right] \\ \int_0^1 W_2(r)\,\mathrm{d}W_1(r) + \mu \end{pmatrix} \end{aligned} \tag{84}$$

where $\mu = \phi_{12s} + \tau_{21}$. Thus, 'correcting' for simultaneity complicates the limiting distribution, and increases its dependence on nuisance effects. Their impact will become clear in the Monte Carlo.

6.5 OLS Estimation: $\lambda = 0, \delta \neq 0$

(h) When $\gamma = 0$, $\mathbf{\Omega}$ is diagonal but $\delta \neq 0$ implies that z_t is not weakly exogenous for β. In fact, both $\mathbf{\Omega}_u$ and $\mathbf{\Phi}_u$ are diagonal so that:

$$T\left(\hat{\beta} - \beta\right) \Rightarrow \left(\int_0^1 B_2(r)^2\, dr\right)^{-1} \int_0^1 B_2(r)\,\mathrm{d}B_1(r) \tag{85}$$

This confirms the asymptotic efficiency of OLS applied to (20) despite the presence of β in (21). The example contrasts vividly with (g) where a similar weak exogeneity failure (but with a levels rather than a difference feedback in the Δz_t process) induces a large bias in OLS. The asymptotic result is due to β in (21) being a coefficient on an I(-1) term.

However, the result in (85) is fragile in that $\gamma \neq 0$ now induces a 'double' failure of weak exogeneity and the limiting distribution is markedly affected: indeed, it has the same form as (76) (see (90) below) and similar conclusions follow as in section [6.3].

7 INFERENCE

We now consider tests of specification hypotheses of the form $\mathsf{H}_0 : \beta = \beta^*$ in the eight cases.

7.1 Inference: $\lambda = 0$

(a) From (64) and (60) when $\gamma = \rho = 0$, the t-test of $\mathsf{H}_0 : \beta = \beta^*$ based on $\mathsf{t}_{\beta*} = \left(\hat{\beta} - \beta^*\right) / \mathsf{SE}\left[\hat{\beta}\right]$ should be asymptotically $\mathsf{N}[0,1]$ when the null is true. In fact:

$$\hat{\sigma}_1^2 = T^{-1} \sum_{t=1}^{T} \left(y_t - \hat{\beta} z_t\right)^2 = T^{-1} \sum_{t=1}^{T} \left(w_{1t} - T\left(\hat{\beta} - \beta\right) T^{-1} z_t\right)^2 \Rightarrow \sigma_1^2 \tag{86}$$

since the terms involving z_t are asymptotically negligible: this result is true independently of ρ and γ. Thus, from (65):

$$\mathsf{t}_{\beta*} = \left(T^{-2} \sum_{t=1}^{T} z_t^2\right)^{-\frac{1}{2}} \frac{T\left(\hat{\beta} - \beta^*\right)}{\hat{\sigma}_1} \Rightarrow \mathsf{N}[0,1] \tag{87}$$

(b) Although there is a bias in the limiting distribution of $T\left(\hat{\beta} - \beta\right)$ in (66) when $\gamma \neq 0$, $\rho = 0$, this is $O_p\left(T^{-1}\right)$ and hence negligible in large samples. Nevertheless, the impact on inference does not vanish even asymptotically. Rather:

$$\begin{aligned} \mathsf{t}_{\beta*} &= \left(T^{-2} \sum_{t=1}^{T} z_t^2\right)^{-\frac{1}{2}} \left(\frac{T\left(\hat{\beta} - \beta^*\right)}{\hat{\sigma}_1}\right) \\ &\Rightarrow \left(1-\gamma^2\right)^{\frac{1}{2}} \left(\int_0^1 B_2(r)^2 \,\mathrm{d}r\right)^{-\frac{1}{2}} \left[\int_0^1 B_2(r)\,\mathrm{d}B_1(r)\right] \\ &\quad + \gamma \left(\int_0^1 B_2(r)^2 \,\mathrm{d}r\right)^{-\frac{1}{2}} \left[\int_0^1 B_2(r)\,\mathrm{d}B_2(r) + 1\right] \\ &= \left(1-\gamma^2\right)^{\frac{1}{2}} \mathsf{N}[0,1] + \gamma DF_{\mathsf{t}} + \gamma \left(\int_0^1 B_2(r)^2 \,\mathrm{d}r\right)^{-\frac{1}{2}} \end{aligned} \tag{88}$$

Compared to (a), the distribution in (88) is non-normal when $\gamma \neq 0$ and conventional hypothesis tests will not have the correct size.
(d) When $\rho \neq 0$, the $\mathsf{t}_{\beta*}$ statistic will again not have an asymptotic normal distribution:

$$\begin{aligned} \mathsf{t}_{\beta*} &\Rightarrow \frac{1}{\left(1+\rho^2\right)} \left(\int_0^1 B_2(r)^2 \,\mathrm{d}r\right)^{-\frac{1}{2}} \left[\int_0^1 B_2(r)\,\mathrm{d}B_1(r) + \rho \int_0^1 B_2(r)\,\mathrm{d}B_2(r)\right] \\ &= \left(1+\rho^2\right)^{-1} \left(\mathsf{N}[0,1] + \rho DF_{\mathsf{t}}\right) \end{aligned} \tag{89}$$

This is proportional to a weighted average of the normal and Dickey-Fuller 't' distribution.
(e) From (58) when $\gamma \neq 0$ and $\rho \neq 0$, a result like (88) occurs, but with different weights for the three components.

7.2 Inference: $\rho = 0$

(c) From (78) when $\gamma = 0$, even if $\lambda \neq 0$ using (65) for the estimated standard error, t-tests of H_0 will be asymptotically $\mathsf{N}[0, 1]$.
(f) As in (88) when $\gamma \neq 0$ and $\lambda \neq 0$, from (76) the same three terms recur with new weights.
The Monte Carlo evidence in section [8] reveals how well these asymptotic results describe finite sample outcomes for both estimation and inference and how large some of the inference distortions can be.

8 A MONTE CARLO STUDY

The Monte Carlo study uses the same DGP as (7)-(10), for specific values of the parameters (λ, ρ, γ) where $\beta = \sigma_1 = \sigma_2 = 1$. The study was undertaken recursively using PC-NAIVE with 10,000 replications (see Hendry, Neale and Ericsson, 1991) across sample sizes $T = 20, \ldots, 100$ for the three regression models (7), (12) and (24) (the last estimated as (25) using instrumental variables). Numerical values of the parameters were selected to illustrate the theoretical derivations, and comprised combinations of $\lambda,\ \rho$ and γ from (0; 0.5) to cover the eight cases (a)-(h), with $\gamma = -\frac{1}{2}$ in (g). We first focus on the full-sample biases in estimating β by the 3 methods in the 8 states of nature.

Table II: Finite Sample Biases and Rejection Frequencies at T=100

	a	b	c	d	e	f	g	h^*
OLS	0.000	0.026	-0.0007	-0.020	0.0001	-0.0003	-0.066	0.023
MCSE	0.0003	0.0004	0.0002	0.0003	0.0002	0.0002	0.0006	0.0003
R	1.07	1.18	1.08	1.13	0.82	0.96	1.77	1.06
P (t)	0.048	0.101	0.049	0.092	0.017	0.030	0.407	0.092
F	0.50	0.79	0.48	0.24	0.52	0.49	0.07	0.78
AOLS	-	0.0001	-	-	-0.015	-0.007	-0.034	-0.01
MCSE	-	0.0003	-	-	0.0002	0.0002	0.0005	0.0003
R	-	1.08	-	-	0.97	1.07	1.46	1.01
P (t)	-	0.048	-	-	0.046	0.053	0.205	0.039
F	-	0.51	-	-	0.21	0.32	0.20	0.49
DOLS	-	0.0002	-	-0.020	-0.009	-0.0003	-0.037	0.0005
MCSE	-	0.0003	-	0.0003	0.0002	0.0002	0.0005	0.0003
R	-	1.09	-	1.15	1.12	1.12	1.20	1.11
P (t)	-	0.052	-	0.092	0.071	0.068	0.129	0.066
F	-	0.50	-	0.24	0.31	0.46	0.21	0.51

Table II records the results. AOLS denotes the augmented OLS regression in (14); DOLS is the instrumental variables estimator of (25); MCSE is the Monte Carlo standard error of the bias; R is the ratio of the sampling standard deviation (SSD) in the experiments to the estimated standard error (ESE); P is the rejection frequency for the correct null hypothesis that $\beta = 1$; and F is the fraction of cases where the bias was positive and is reported since DOLS has no finite sample moments (see Sargan, 1982). The Monte Carlo standard errors of P are 0.002 at $P = 0.05$. When OLS was a ‘correctly specified’ special case of the other estimators, the latter were not computed (shown by –). Case h^* denotes that $\gamma \neq 0$. In every cases, an unrestricted intercept was included in estimated models.

In almost all cases, the distribution of the t-test of a correct null hypothesis was closer to symmetry than the distribution of the estimator, but was centered on the wrong location: its corresponding F value was very close to that shown in Table II.

(a) As the first column records, OLS is well behaved when $\gamma = \rho = 0$, being unbiased, with an essentially unbiased ESE for the SSD (since R is close to unity) and about a 5% rejection frequency of the hypothesis that $\beta = 1$. F confirms that OLS is median unbiased matching (60).

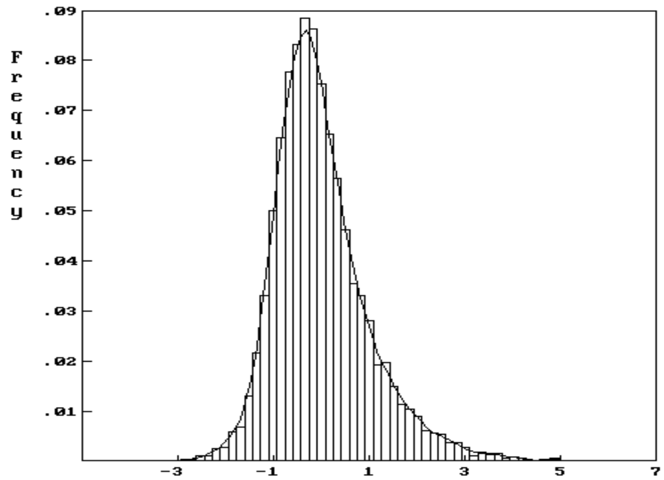

Figure 1: Standardized frequency distribution of OLS biases in case (b)

(b) Next, the introduction of cross correlation between the errors on the DGP ($\gamma \neq 0$) induces a noticeable positive bias in OLS (matching (61)), substantial skewness as measured by F, increases R, and leads to considerable over-rejection of the correct null that $\beta = 1$. Figure 1 records the standardized frequency distribution of the estimates, which manifests substantial skewness. Correcting by using (14) (i.e. AOLS) restores mean and median unbiasedness and brings R and P back to virtually the same values as OLS in case (a), matching (75) for $\rho = 0$. Further, the dynamic model (24) is well behaved. Since z_t is strongly exogenous for the parameters in both (14) and (24), single equation inference need lose no relevant information.

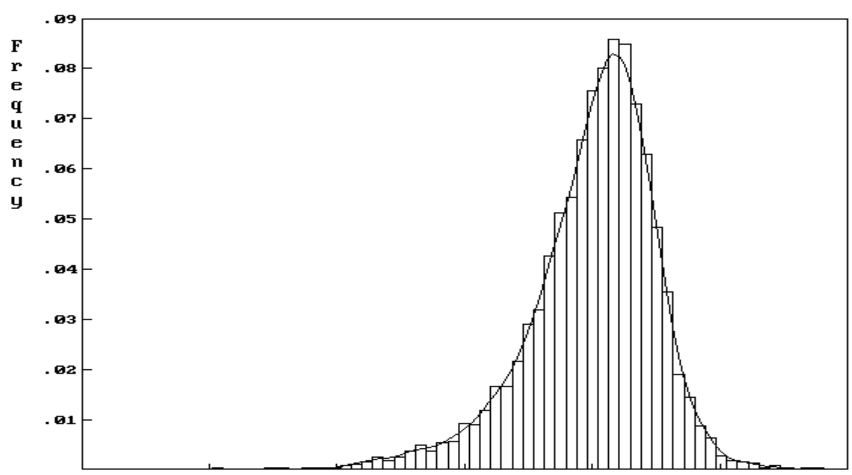

Figure 2: Standardized frequency distribution of OLS biases in case (d)

(d) The bias in OLS is negative when $\rho \neq 0$ and hence opposite to (b), matching (64). Again, P shows considerable over-rejection of the correct null on β. Figure 2 records the standardized frequency distribution to illustrate the skewness. Estimating the dynamic model does not help to correct such problems, since the outcome is almost the same as OLS. This is precisely the case in which both models coincide with the conditional expectation, but z_t is not weakly exogenous for the parameters in any of (7), (14) or (24). Thus, the information loss from ignoring the marginal model (17) is marked and distorts several aspects of inference.

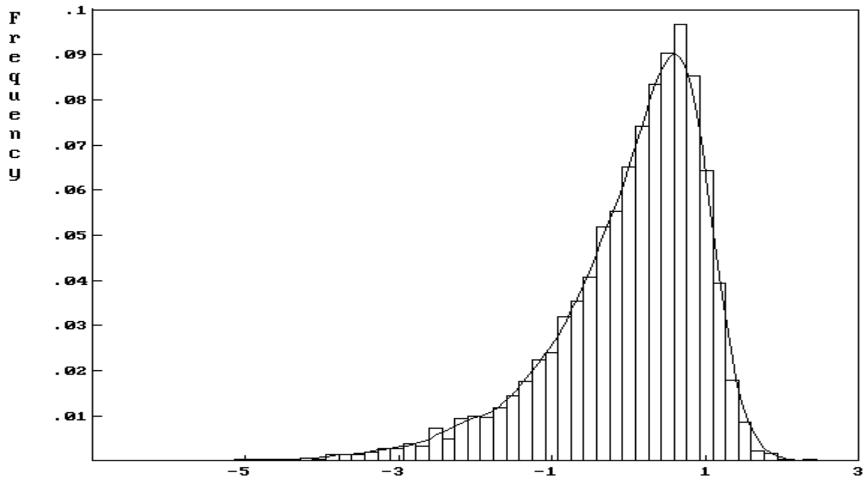

Figure 3: Standardized frequency distribution of OLS biases in case (g)

(e) This case would ordinarily be a puzzle in a simulation exercise: weak exogeneity is clearly violated, yet OLS is nearly mean and median unbiased. Such an outcome is explained by the partial offset of the terms in (58) for the parameter values deliberately chosen here. However, inference is distorted in that R is very low, as is the size of the 't'-test of $\beta^* = 1$. Correcting for $\gamma \neq 0$ induces a substantial negative bias, as anticipated from (75), although R and P are now better behaved. The dynamic model estimator based on (24) also remains negatively biased, despite coinciding with the conditional expectation in (16). Thus, once more a failure of weak exogeneity leads to a loss of information. Further, although (67) and (75) show the asymptotic equivalence of OLS in (d) and AOLS in (e), Table II reveals distinct finite sample differences, especially on R and P. We now consider the cases where $\rho = 0$ but $\lambda \neq 0$, so strong exogeneity can never occur.

(c) When $\gamma = 0$, there is a slight mean bias in OLS, but $R,\ P$ and F are all well behaved. Since weak exogeneity of z_t for β holds, and (77) shows that OLS is asymptotically a mixture of normals, the bias seems analogous to that which occurs in finite sample when estimating stationary dynamic models. AOLS and the dynamic model yielded similar results.

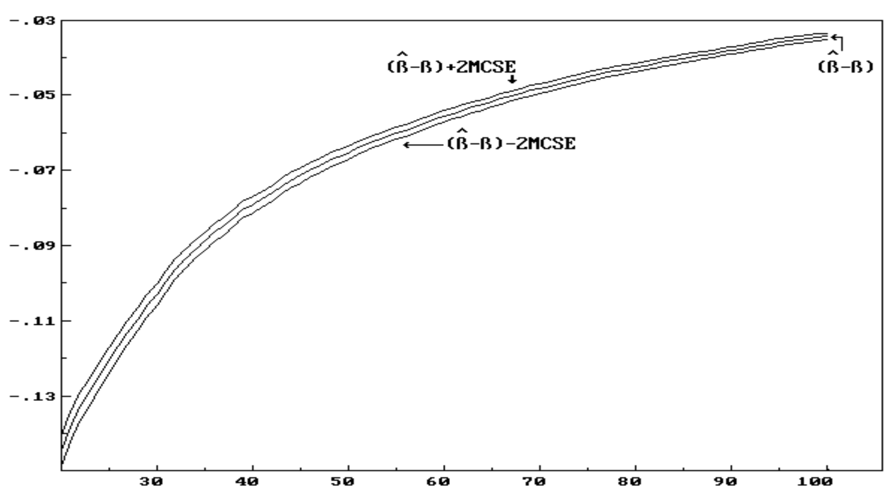

Figure 4: AOLS biases in case (g) with ± 2MCSE

(f) This case is again an enigma at first sight: OLS is nearly mean and median unbiased, and R is close to unity although P is low. The outcome matches the offset anticipated from (76) which attenuates the slight bias in (c). Augmenting OLS by adding Δz_t does not help: the bias is much larger as (82) suggested could occur. The dynamic model (now with five regressors) performs fairly well despite the failure of weak exogeneity but shows some median bias. Since $\rho = 0$, the coefficient δ_2 in (25) is zero, so OLS could be applied to the unrestricted variant of the conditional expectation (19). This yielded a mean bias of -0.0002 (0.0002), with $R = 1.09,\ P = 0.051$ and $F = 0.50$, so there is no obvious evidence of the weak exogeneity failure for these parameter values.
(g) We used $\gamma = -\frac{1}{2}$ and $\rho = \frac{1}{2}$ so $\gamma + \rho = 0$ as the example of type (e) with a diagonal long-run covariance matrix, while maintaining a violation of weak exogeneity. From (58), a negative bias can be anticipated despite the independence of $V_1(r)$ and $V_2(r)$. This indeed occurs, confirming that a diagonal long-run covariance matrix is not sufficient to sustain inference about a cointegration parameter. This also matches (18), as recorded in (68). Figure 3 shows the frequency distribution of the OLS bias and confirms the leftwards skewness. The over-rejection of the correct null is dramatic, and remains bad even for AOLS. As ever, DOLS helps somewhat, but cannot fully correct given the weak exogeneity failure.
(h) The finite sample outcomes when $\delta \neq 0$ but $\gamma = 0$ are not reported in detail. However, they differed from the asymptotic predictions in (85) since there was a significant bias of -0.0012 (0.0003) even at $T = 100$. More importantly, the outcome in (85) depends on $\gamma = 0$, and when that does not hold, both $\mathbf{\Omega}_u\ (= \mathbf{\Sigma})$ and $\mathbf{\Phi}_u$ cease to be diagonal. This is case h^* and the limiting distribution of $T\left(\hat{\beta} - \beta\right)$ becomes:

$$\left(\int_0^1 B_2(r)^2\,\mathrm{d}r\right)^{-1}\left[\left(1-\gamma^2\right)^{\frac{1}{2}}\int_0^1 B_2(r)\,\mathrm{d}B_1(r) + \gamma\left\{\int_0^1 B_2(r)\,\mathrm{d}B_2(r) + 1\right\}\right] \tag{90}$$

so an upward bias can be expected. The weak exogeneity status of z_t for β is not affected by the value of γ, although the outcome is, since (7) ceases to be 'correctly specified', albeit by the omission of an I(0) component. Column h^* in Table II records the finite sample outcomes which match (90) when $\gamma = 0.5$. There is a substantial upward bias and distinct over-rejection of the correct null that $\beta^* = 1$.

Corrections like (14) fail when $\gamma \neq 0$, since from (22), $\gamma\delta\left(\Delta y_{t-1} - \beta \Delta z_{t-1}\right) = \gamma\delta\Delta\epsilon_{1t-1}$ has been omitted from the conditional expectation. This induces a small amount of autocorrelation, as the entries for AOLS in (h) show: there is little mean or median bias but P reveals a small test size. The dynamic estimator has six regressors $(1, z_t, \Delta z_t, \Delta y_t, \Delta z_{t-1}, \Delta y_{t-1})$, and given (85), performs reasonably well despite the absence of weak exogeneity (but was too large to be estimated recursively).

The results above have all been at the largest sample size, namely $T = 100$. We next consider the recursively computed findings, and summarize these in four graphs, of which the first two are only illustrative. Figure 4 shows a typical outcome for AOLS estimating the bias in case (g), and is included to emphasize the high degree of precision of the Monte Carlo. These figures were created using Pc-Give (Doornik and Hendry, 1992).

The central line shows the bias. This is large when $T = 20$, and falls towards zero at a rate which is less rapid than T over the sample sizes used here. The bounding lines show the bias ± 2MCSE and confirm the accuracy of the bias estimates at all sample sizes considered.

Figure 5 shows the relation between the ESE and the SSD (which was summarized in Table II by the ratio R) for AOLS in experiment (g). The relationship of ESE to SSD is similar at all the sample sizes considered and this finding was usually true across the experiments, suggesting that R in Table II is a fairly representative measure. The main summary graphs across sample sizes are Figures 6 and 7, which record the biases in experiments (a), (b), (d), (e) and (c), (f), (g), (h) and (h*) respectively for all estimator/model combinations computed.

The substantial impact of the design variables on the biases is clear, as is the convergence of the biases to zero as well as the wide range of outcomes generated (biases which were large and positive have been plotted after a sign change to focus on absolute magnitudes). Since the estimated models are small, $T = 100$ must be judged a 'large' sample, so the outcomes at the smaller sample sizes (e.g. 40-60) may be more representative of empirical behaviour than those in Table II which were used to highlight the usefulness of the limiting distributions.

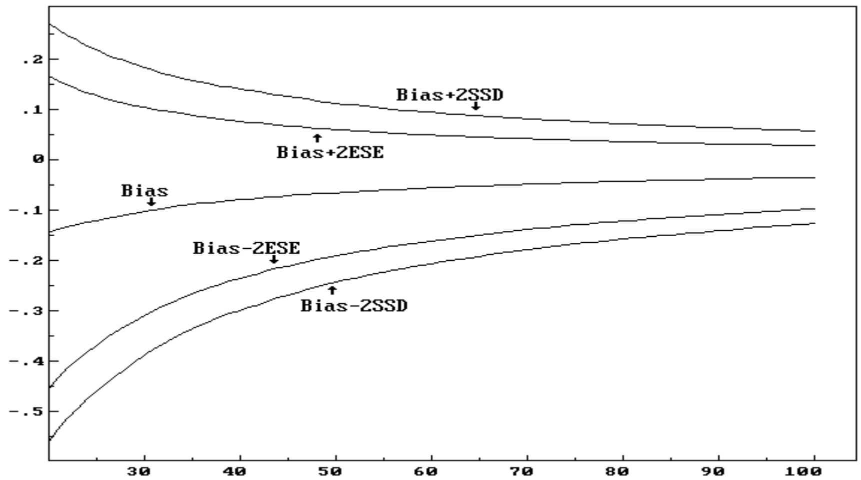

Figure 5: ESE and SSD for AOLS in case (g)

As both figures show, the biases are rather large at the smallest samples in many cases, particularly cases (b), (d) and (g). The unlabelled lines in Figure 6 are for OLS(a) and AOLS(b) where the biases are negligible throughout (under 2%). In Figure 7, OLS(h) denotes the estimator of (19) and the line immediately adjacent to it is OLS(c); the remaining unlabelled lines are for OLS(f), AOLS(h*) and DOLS(h*). Apart from the sampling variability in DOLS at very small samples, the bias lines never cross, so the qualitative conclusions in Table II about relative biases are relevant at all the sample sizes studied. Given the accuracy in Figure 4, the main differences between biases are significant. The biases in Figure 7 for case (g) are the most extreme, exceeding 25% of β for OLS at $T = 20$, yet this is when $\mathbf{\Omega}$ is diagonal. Further, the biases in case (h) are noticeable at small T, despite the asymptotic efficiency of OLS in that case, so there is a loss in small samples due to the failure of weak exogeneity.

Otherwise, for $T \geq 50$ and the parameter values chosen here, no biases exceed 10%. Since biases of that magnitude can occur in stationary dynamic models, there are no new estimation difficulties in I(1) systems for the models under analysis when weak exogeneity of the conditioning variables holds for the parameters of interest. However, the parameter points selected as illustrative are far from extreme and provide a relatively favourable state for methods which violate weak exogeneity: in particular, $\gamma\rho = -0.8$ in (24) would be more realistic, and would generate far larger biases than those shown in Table II.

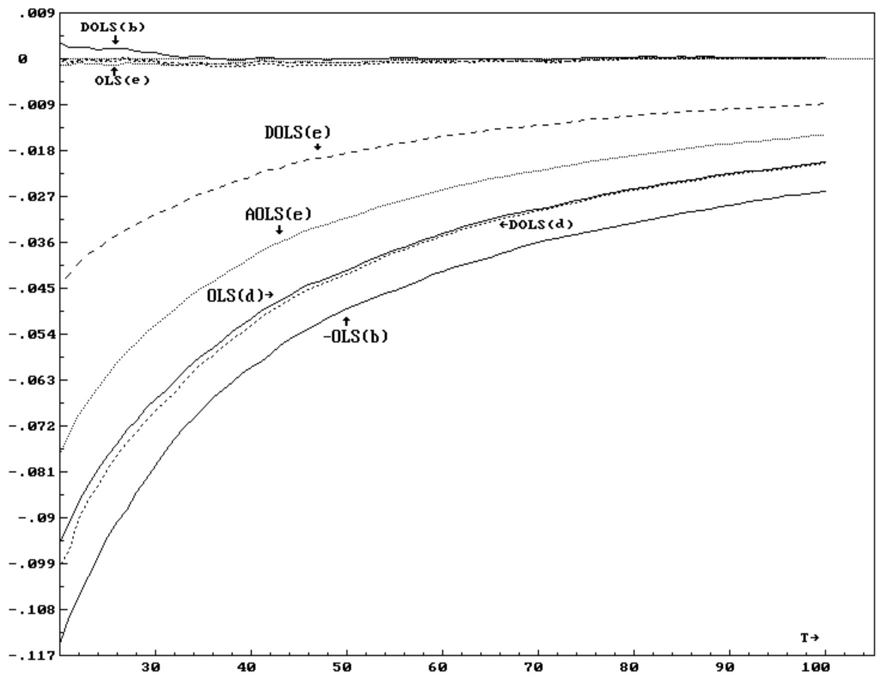

Figure 6: Recursively computed biases in cases (a), (b), (d), (e)

A less sanguine conclusion holds for inference when weak exogeneity fails. Although the non-centrality of the limiting distribution of the estimator only induces biases of $O_p\left(T^{-1}\right)$ which vanish quite rapidly in practice as figs. 6 and 7 reveal, the biases are of the same order as the estimated standard errors and hence inference is liable to be highly distorted. This can be seen in figs. 8-10 which plot the rejections of the correct null hypothesis of $\beta = 1$ using conventional t-tests at the 5% significance level, in cases (a), (b), (d), (e); (f), (h) and (h*) and case (g) respectively. There is no tendency for the sizes of the tests to converge on 0.05 when they are incorrect, and in case (g) the size distortion is both large and increases over the sample sizes considered. Thus, minor biases can distort inference even asymptotically when weak exogeneity fails to hold or the model is incorrectly specified in its I(0) components.

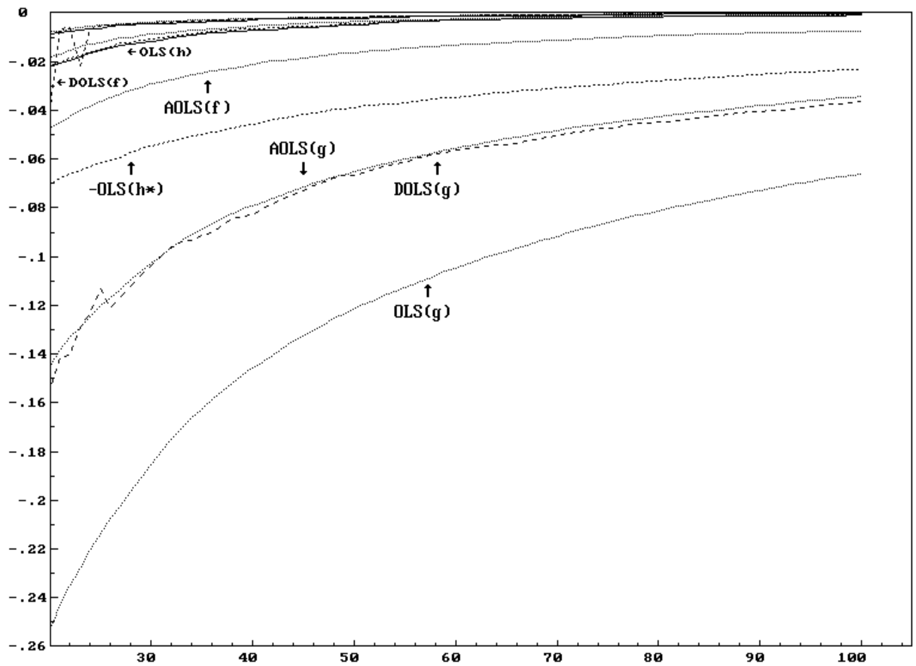

Figure 7: Recursively computed biases in cases (c), (f)-(h*)

9 TESTING WEAK EXOGENEITY

The eight special cases discussed in section [6] also serve to highlight important differences between various tests for weak exogeneity and estimator consistency respectively. Urbain [1992] draws a similar distinction. Here, tests for orthogonality between regressors and errors may or may not test for weak exogeneity. In one case (namely when $\rho = 0,\ \lambda = 0$ and $\gamma \neq 0$ as in (b)), since the fitted model (7) is mis-specified, such tests could reject with probability close to unity although z_t is in fact weakly exogenous for the parameters of the conditional model. Conversely, they may fail to reject beyond their size even though z_t is not weakly exogenous for the parameters of the conditional model (namely when $\rho \neq 0,\ \lambda = 0$ and $\gamma = 0$ as in (d)).

From (12), the optimal tests for the DGP in (7)-(10) are for $\mathsf{H}_\rho : \rho = 0$ when $\lambda = 0$ and $\gamma = 0$; and for $\mathsf{H}_\gamma : \gamma = 0$ when $\rho = 0$ but $\lambda \neq 0$. The former remains valid but inefficient when $\gamma \neq 0$.

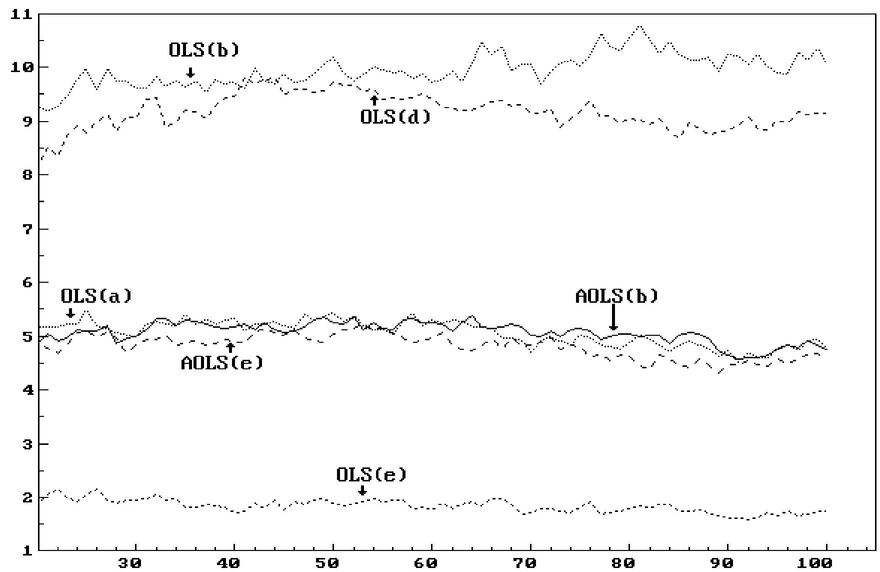

Figure 8: Recursively computed rejection frequencies in (a), (b), (d), (e)

10 CONCLUSION

The paper exposits the impact of weak and strong exogeneity failures on estimating conditional expectations, and linear approximations thereto, in co-integrated processes. Its main aim was to exposit the asymptotic analysis and illustrate the findings using a simple bivariate data generation process. The Monte Carlo results show a bewildering array of possible outcomes but the limiting distributions explain the vast majority of the outcomes. In almost all cases, there is an excellent match between the small sample outcomes and the asymptotic theory based on functionals of Wiener processes.

The results show the impact of weak exogeneity failures in integrated data. The diagonality of the long-run covariance matrix is not sufficient to sustain inference in all cases, especially in small samples, although efficient inference may be possible asymptotically in some cases where weak exogeneity fails but the long-run covariance matrix remains diagonal. Clearly, weak exogeneity alone is not sufficient either. Correcting for $\mathsf{I}(0)$ effects can have different consequences depending on the status of conditioning.

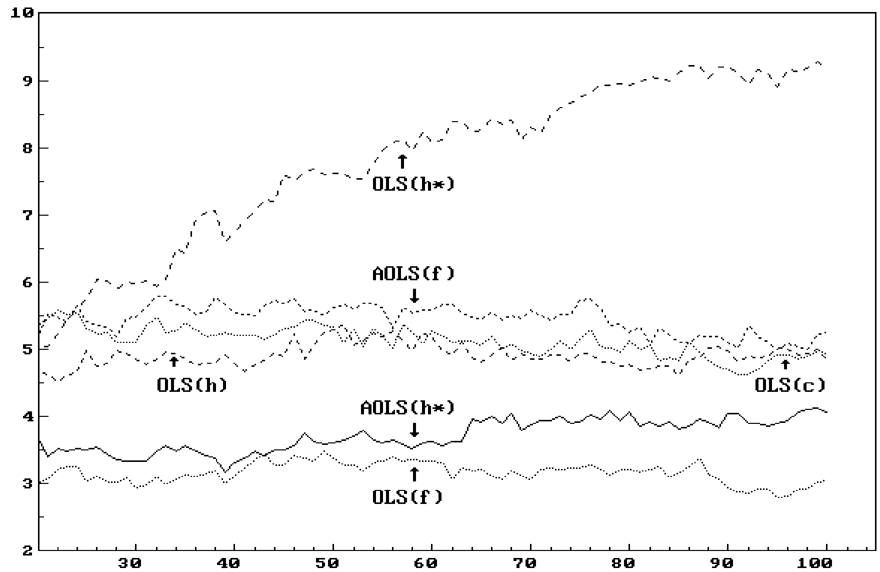

Figure 9: Recursively computed rejection frequencies in (c), (f), (h), (h*)

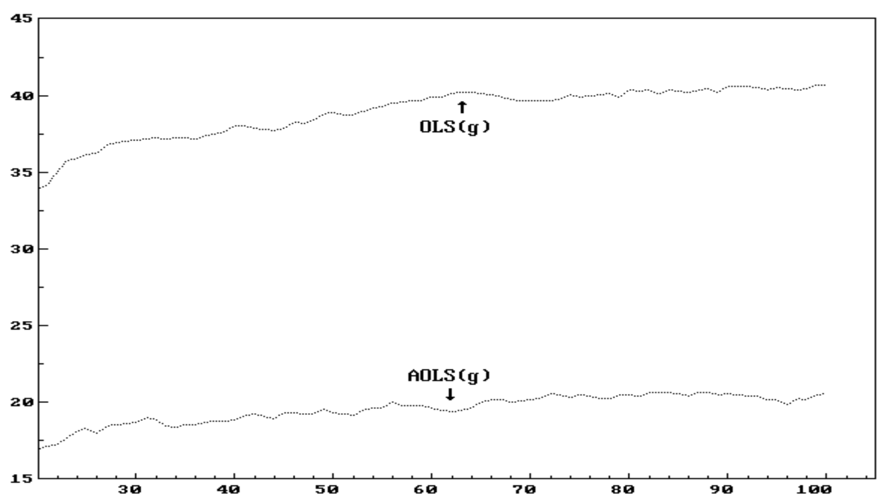

Figure 10: Recursively computed rejection frequencies in (g)

Granger-causality does not seriously impede inference when weak exogeneity holds, so strong exogeneity is not necessary to sustain inference. However, in co-integrated processes the absence of weak exogeneity can have adverse effects on estimation in small samples, and on inference asymptotically as well, even when the model under analysis coincides with the conditional expectation. Tests for weak exogeneity do not necessarily coincide with tests for orthogonality between regressors and errors: the latter may reveal other forms of mis-specification of the fitted model when it does not coincide with the conditional expectation, and may reveal no mis-specification when the fitted model coincides with the conditional expectation but weak exogeneity is violated. Weak exogeneity, as the basis for conditional inference about parameters of interest with no loss of information, seems to be at least as relevant in I(1) as in stationary processes.

11 APPENDIX

The DGP in (7)-(11) defines the (co-)integrated vector process:

$$y_t = \beta\lambda y_{t-1} + \beta w_{2t} + w_{1t} \tag{91}$$

$$z_t = \beta\lambda z_{t-1} + \lambda w_{1t-1} + w_{2t} \tag{92}$$

Differencing to remove the unit root in w_{2t} in (9), the system can be expressed as $\Delta \mathbf{x}_t = \boldsymbol{\nu}_t$ where:

$$\begin{aligned}\begin{pmatrix}\nu_{1t}\\ \nu_{2t}\end{pmatrix} &= \begin{pmatrix}\lambda\beta & 0\\ 0 & \lambda\beta\end{pmatrix}\begin{pmatrix}\nu_{1t-1}\\ \nu_{2t-1}\end{pmatrix} + \begin{pmatrix}\epsilon_{1t} + \beta\epsilon_{2t} - (1-\beta\rho)\,\epsilon_{1t-1}\\ \epsilon_{2t} + (\lambda+\rho)\,\epsilon_{1t-1} - \lambda\epsilon_{1t-2}\end{pmatrix}\\ &= \begin{pmatrix}\lambda\beta & 0\\ 0 & \lambda\beta\end{pmatrix}\begin{pmatrix}\nu_{1t-1}\\ \nu_{2t-1}\end{pmatrix} + \begin{pmatrix}1 & \beta\\ 0 & 1\end{pmatrix}\begin{pmatrix}\epsilon_{1t}\\ \epsilon_{2t}\end{pmatrix}\\ &\quad + \begin{pmatrix}\beta\rho - 1 & 0\\ \lambda+\rho & 0\end{pmatrix}\begin{pmatrix}\epsilon_{1t-1}\\ \epsilon_{2t-1}\end{pmatrix} + \begin{pmatrix}0 & 0\\ -\lambda & 0\end{pmatrix}\begin{pmatrix}\epsilon_{1t-2}\\ \epsilon_{2t-2}\end{pmatrix}\end{aligned} \tag{93}$$

or:

$$\boldsymbol{\nu}_t = \mathbf{B}_1\boldsymbol{\nu}_{t-1} + \mathbf{C}_0\boldsymbol{\epsilon}_t + \mathbf{C}_1\boldsymbol{\epsilon}_{t-1} + \mathbf{C}_2\boldsymbol{\epsilon}_{t-2}. \tag{94}$$

Thus, $\boldsymbol{\nu}_t$ is a stationary VARMA(1,2). Alternatively, in VAR form, we have:

$$\Delta\mathbf{x}_t = \boldsymbol{\pi}\boldsymbol{\Delta}\mathbf{x}_{t-1} + (\mathbf{S} - \mathbf{I})\,\mathbf{x}_{t-1} + \boldsymbol{\eta}_t \tag{95}$$

where:

$$\boldsymbol{\pi} = \begin{pmatrix}\lambda\beta & 0\\ \lambda & 0\end{pmatrix} = \begin{pmatrix}\beta\\ 1\end{pmatrix}[\lambda : 0]\,;\ \mathbf{S} = \begin{pmatrix}\beta\rho & \beta(1-\beta\rho)\\ \rho & (1-\beta\rho)\end{pmatrix} = \begin{pmatrix}\beta\\ 1\end{pmatrix}[\rho : (1-\beta\rho)]$$

and

$$(\mathbf{S} - \mathbf{I}) = \begin{pmatrix}-(1-\beta\rho)\\ \rho\end{pmatrix}(1 : -\beta)$$

Then:

$$\mathbf{x}_t = \mathbf{S}\mathbf{x}_{t-1} + \boldsymbol{\eta}_t^* = \mathbf{S}\sum_{k=0}^{t}\boldsymbol{\eta}_k^* + \mathbf{x}_0 \tag{96}$$

Many derivations in the text require the long-run covariance matrix of a stationary process. In general, the long-run covariance matrix $\boldsymbol{\Omega}_y$ of a mean-zero weakly stationary stochastic process $\{\mathbf{y}_t\}$ is given by the limit as $T \to \infty$ of:

$$\begin{aligned}\mathsf{E}\left[T^{-1}\left(\sum_{t=1}^{T}\mathbf{y}_t\right)\left(\sum_{s=1}^{T}\mathbf{y}_s'\right)\right] &= \mathsf{E}\left[T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{y}_t\mathbf{y}_s'\right]\\ &= \mathsf{E}\left[T^{-1}\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t'\right] + \mathsf{E}\left[T^{-1}\sum_{t=1}^{T}\sum_{s=t+1}^{T}\mathbf{y}_t\mathbf{y}_s'\right]\\ &\quad + \mathsf{E}\left[T^{-1}\sum_{t=1}^{T}\sum_{s=t+1}^{T}\mathbf{y}_s\mathbf{y}_t'\right]\\ &= \boldsymbol{\Phi} + \sum_{s=1}^{T-1}\mathsf{E}\left[\mathbf{y}_t\mathbf{y}_{t-s}'\right] + \sum_{s=1}^{T-1}\mathsf{E}\left[\mathbf{y}_{t-s}\mathbf{y}_t'\right]\\ &\to \boldsymbol{\Phi} + \boldsymbol{\Upsilon}' + \boldsymbol{\Upsilon} = \boldsymbol{\Omega}_y\end{aligned} \tag{97}$$

where $\mathbf{\Phi} = \mathsf{E}\left[\mathbf{y}_t\mathbf{y}_t'\right]$ is the unconditional covariance matrix.

When the process is a first-order autoregression: $\mathbf{y}_t = \mathbf{R}\mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$ where $\boldsymbol{\epsilon}_t \sim \mathsf{IN}\left(\mathbf{0}, \mathbf{\Sigma}\right)$, then $\mathbf{\Omega}_y$ in (97) becomes:

$$\left(\mathbf{I} + \mathbf{R} + \mathbf{R}^2 + \mathbf{R}^3 + \cdots + \mathbf{R}^{T-1}\right)\mathbf{\Phi} + \mathbf{\Phi}\left(\mathbf{I} + \mathbf{R} + \mathbf{R}^2 + \mathbf{R}^3 + \cdots + \mathbf{R}^{T-1}\right)' - \mathbf{\Phi} \tag{98}$$

which tends to $\left(\mathbf{I} - \mathbf{R}\right)^{-1}\mathbf{\Phi} + \mathbf{\Phi}\left(\mathbf{I} - \mathbf{R}'\right)^{-1} - \mathbf{\Phi}$, where $\mathbf{\Phi} = \mathbf{R}\mathbf{\Phi}\mathbf{R}' + \mathbf{\Sigma}$. However, a more convenient form of $\mathbf{\Omega}_y$, directly related to the spectral density at the origin, results from using:

$$\mathbf{\Sigma} = \mathbf{\Phi} - \mathbf{R}\mathbf{\Phi}\mathbf{R}' = \left(\mathbf{I} - \mathbf{R}\right)\mathbf{\Phi} + \mathbf{\Phi}\left(\mathbf{I} - \mathbf{R}'\right) - \left(\mathbf{I} - \mathbf{R}\right)\mathbf{\Phi}\left(\mathbf{I} - \mathbf{R}'\right) \tag{99}$$

so that on pre-multiplying $\mathbf{\Sigma}$ by $\left(\mathbf{I} - \mathbf{R}\right)^{-1}$ and post-multiplying by $\left(\mathbf{I} - \mathbf{R}'\right)^{-1}$ and using (98):

$$\mathbf{\Omega}_y = \left(\mathbf{I} - \mathbf{R}\right)^{-1}\mathbf{\Sigma}\left(\mathbf{I} - \mathbf{R}'\right)^{-1} \tag{100}$$

Similar principles apply to deriving the $\mathbf{\Omega}_y$ matrix for more general weakly stationary processes, such as those in the text. For example, in a first-order moving average:

$$\mathbf{y}_t = \boldsymbol{\epsilon}_t + \mathbf{D}\boldsymbol{\epsilon}_{t-1} \text{ where } \mathsf{E}\left[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\right] = \mathbf{\Sigma} \tag{101}$$

then:

$$\mathsf{E}\left[\mathbf{y}_t\mathbf{y}_t'\right] = \mathbf{\Psi} = \mathsf{E}\left[\left(\boldsymbol{\epsilon}_t + \mathbf{D}\boldsymbol{\epsilon}_{t-1}\right)\left(\boldsymbol{\epsilon}_t' + \boldsymbol{\epsilon}_{t-1}'\mathbf{D}'\right)\right] = \mathbf{D}\mathbf{\Sigma}\mathbf{D}' + \mathbf{\Sigma} \tag{102}$$

and:

$$\mathbf{\Omega}_y = \mathbf{\Psi} + \mathbf{D}\mathbf{\Sigma} + \mathbf{\Sigma}\mathbf{D}' = \mathbf{D}\mathbf{\Sigma}\mathbf{D}' + \mathbf{\Sigma} + \mathbf{D}\mathbf{\Sigma} + \mathbf{\Sigma}\mathbf{D}' = \left(\mathbf{I} + \mathbf{D}\right)\mathbf{\Sigma}\left(\mathbf{I} + \mathbf{D}'\right) \tag{103}$$

The long-run covariance matrix $\mathbf{\Omega}_\nu$ of $\{\boldsymbol{\nu}_t\}$ in (93) is singular because of cointegration, and analogously to (100) and (103) is given by:

$$\begin{aligned}
\mathbf{\Omega}_\nu &= \left(\mathbf{I} - \mathbf{B}_1\right)^{-1}\left(\mathbf{C}_0 + \mathbf{C}_1 + \mathbf{C}_2\right)\mathbf{\Sigma}\left(\mathbf{C}_0' + \mathbf{C}_1' + \mathbf{C}_2'\right)\left(\mathbf{I} - \mathbf{B}_1'\right)^{-1} \\
&= \left(1 - \beta\lambda\right)^{-2}\begin{pmatrix} \beta\rho & \beta \\ \rho & 1 \end{pmatrix}\begin{pmatrix} \sigma_1^2 & \gamma \\ \gamma & \sigma_2^2 \end{pmatrix}\begin{pmatrix} \beta\rho & \rho \\ \beta & 1 \end{pmatrix} \\
&= \left(1 - \beta\lambda\right)^{-2}\left(\rho^2\sigma_1^2 + 2\rho\gamma + \sigma_2^2\right)\begin{pmatrix} \beta^2 & \beta \\ \beta & 1 \end{pmatrix}
\end{aligned} \tag{104}$$

Thus, despite its initial complexity, $\mathbf{\Omega}_\nu$ simplifies greatly to a singular matrix, yet one which is dependent on all the parameters $(\beta, \lambda, \rho, \gamma, \sigma_1, \sigma_2)$ in the DGP.

However, the unconditional error variance matrix $\mathbf{\Phi}_\nu = \mathsf{E}\left[\boldsymbol{\nu}_t\boldsymbol{\nu}_t'\right]$ of $\{\boldsymbol{\nu}_t\}$ in (93) is not simple, namely:

$$\begin{aligned}
\mathbf{\Phi}_\nu &= \mathbf{B}_1\mathbf{\Phi}_\nu\mathbf{B}_1' + \mathbf{C}_0\mathbf{\Sigma}\mathbf{C}_0' + \mathbf{C}_1\mathbf{\Sigma}\mathbf{C}_1' + \mathbf{C}_2\mathbf{\Sigma}\mathbf{C}_2' + \mathbf{C}_1\mathbf{\Sigma}\mathbf{C}_0'\mathbf{B}_1' + \mathbf{B}_1\mathbf{C}_0\mathbf{\Sigma}\mathbf{C}_1' \\
&\quad + \mathbf{B}_1^2\mathbf{C}_0\mathbf{\Sigma}\mathbf{C}_2' + \mathbf{B}_1\mathbf{C}_1\mathbf{\Sigma}\mathbf{C}_2' + \mathbf{C}_2\mathbf{\Sigma}\mathbf{C}_0'\mathbf{B}_1^{2\prime} + \mathbf{C}_2\mathbf{\Sigma}\mathbf{C}_1'\mathbf{B}_1' \\
&= \left(1 - \beta^2\lambda^2\right)^{-1}\left(\mathbf{C}_0\mathbf{\Sigma}\mathbf{C}_0' + \mathbf{C}_1\mathbf{\Sigma}\mathbf{C}_1' + \mathbf{C}_2\mathbf{\Sigma}\mathbf{C}_2'\right) \\
&\quad + \beta\lambda\left(1 - \beta^2\lambda^2\right)^{-1}\left(\mathbf{C}_1\mathbf{\Sigma}\mathbf{C}_0' + \mathbf{C}_0\mathbf{\Sigma}\mathbf{C}_1' + \mathbf{C}_2\mathbf{\Sigma}\mathbf{C}_1' + \mathbf{C}_1\mathbf{\Sigma}\mathbf{C}_2'\right) \\
&\quad + \beta^2\lambda^2\left(1 - \beta^2\lambda^2\right)^{-1}\left(\mathbf{C}_2\mathbf{\Sigma}\mathbf{C}_0' + \mathbf{C}_0\mathbf{\Sigma}\mathbf{C}_2'\right)
\end{aligned} \tag{105}$$

In several special cases, such as $\lambda = 0$ or $\rho = 0$, (105) simplifies greatly.

ACKNOWLEDGEMENTS

This research was financed in part by grants R000233447 and R231184 from the UK Economic and Social Research Council. Helpful comments from Mike Clements, Juan Dolado, Andreas Fischer, Søren Johansen, Grayham Mizon, Tom Rothenberg, Neil Shephard and an anonymous referee of this Journal are gratefully acknowledged.

References

Banerjee, A., Dolado, J.J., Galbraith, J.W. and Hendry, D.F. (1993), *Cointegration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford University Press.

Banerjee, A. and Hendry, D.F. (1992), Testing Integration and Cointegration: An Overview, *Oxford Bulletin of Economics and Statistics*, **54**, 225-55.

Barndorff-Nielsen, O. (1978), *Information and Exponential Families in Statistical Theory*. Chichester: John Wiley.

Bewley, R.A. (1979), The Direct Estimation of the Equilibrium Response in a Linear Model, *Economics Letters*, **3**, 357-61.

Boswijk, H.P. (1992a), Efficient Inference on Cointegration Parameters in Structural Error Correction Models, University of Amsterdam, Econometrics Discussion Paper.

Boswijk, H.P. (1992b), *Cointegration, Identification and Exogeneity*. Amsterdam: Tinbergen Institute Research Series, Vol. 37.

Dickey, D.A. and Fuller, W.A. (1979) Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of American Statistical Association*, **74**, 427-31.

Dickey, D.A. and Fuller, W.A. (1981) Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, *Econometrica*, **49**, 1057-72.

Dolado J.J. (1992), A Note on Weak Exogeneity in VAR Cointegrated Systems, *Economic Letters*, **38**, 139-43.

Doornik, J.A. and Hendry, D.F. (1992), *Pc-Give 7: An Interactive Econometric Modelling System*. Oxford: Institute of Economics and Statistics, University of Oxford.

Engle, R.F. and Granger, C.W.J. (1987) Co-Integration and Error Correction: Representation, Estimation, and Testing, *Econometrica*, **55**, 251-76.

Engle, R.F., Hendry, D.F. and Richard, J-F. (1983) Exogeneity, *Econometrica*, **51**, 277-304.

Ericsson, N.R. (1992), Cointegration, Exogeneity and Policy Analysis: An Overview, *Journal of Policy Modeling*, **14**, 251-280.

Fuller, W.A. (1976) *Introduction to Statistical Time Series*. New York: John Wiley and Sons.

Gonzalo, J. (1989), Comparison of Five Alternative Methods of Estimating Long Run Equilibrium Relationships, Discussion Paper 89-55, University of California at San Diego.

Granger, C.W.J. (1969) Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, *Econometrica*, **37**, 424-38.

Hendry, D.F. (1990), Using PC-NAIVE in Teaching Econometrics, *Oxford Bulletin of Economics and Statistics*, **53**, 199-223.

Hendry, D.F. and Mizon, G.E. (1993), Evaluating Dynamic Econometric Models by Encompassing the VAR, p272-300 in *Models Methods and Applications of Econometrics*. P.C.B. Phillips (ed.), Oxford: Basil Blackwell.

Hendry, D.F. Neale, A.J. and Ericsson, N.R. (1991), *PCNAIVE: An Interactive Program for Monte Carlo Experimentation in Econometrics*. Version 6.0, Oxford: Institute of Economics and Statistics, University of Oxford.

Johansen, S. (1992a), Cointegration in Partial Systems and the Efficiency of Single Equation Analysis, *Journal of Econometrics*, **52**, 389-402.

Johansen, S. (1992b), Testing Weak Exogeneity and the Order of Cointegration in UK Money Demand, *Journal of Policy Modeling*, **14**, 313-34.

Johansen, S. and Juselius, K. (1990) Maximum Likelihood Estimation and Inference on Cointegration — with Application to the Demand for Money, *Oxford Bulletin of Economics and Statistics*, **52**, 169-210.

Kiviet, J.F. and Phillips, G.D.A. (1992), Exact Similar Tests for Unit Roots and Cointegration, *Oxford Bulletin of Economics and Statistics*, **54**, 349-67.

Koopmans, T.C. (1950), When is an Equation System Complete for Statistical Purposes?, in *Statistical Inference in Dynamic Economic Models*, Koopmans, T.C. (ed.), New York: John Wiley.

Maasoumi, E. (1978), A Modified Stein-like Estimator for the Reduced Form Coefficients of Simultaneous Equations, *Econometrica*, **46**, 695-704.

Park, J.Y. and Phillips, P.C.B. (1988), Statistical Inference in Regressions with Integrated Processes. Part 1, *Econometric Theory*, **4**, 468-97.

Park, J.Y. and Phillips, P.C.B. (1989), Statistical Inference in Regressions with Integrated Processes. Part 2, *Econometric Theory*, **5**, 95-131.

Phillips, P.C.B. (1986) Understanding Spurious Regressions in Econometrics, *Journal of Econometrics*, **33**, 311-40.

Phillips, P.C.B. (1987) Time Series Regression with a Unit Root, *Econometrica*, **55**, 2, 277-301.

Phillips, P.C.B. (1988) Regression Theory for Near-integrated Time Series, *Econometrica*, **56**, 5, 1021-43.

Phillips, P.C.B. (1991), Optimal Inference in Cointegrated Systems, *Econometrica*, **59**, 283-306.

Phillips, P.C.B. and Durlauf, S.N. (1986), Multiple Time Series Regression with Integrated Processes, *Review of Economic Studies*, **53**, 473-95.

Phillips, P.C.B. and Loretan, M. (1991), Estimating Long-Run Economic Equilibria, *Review of Economic Studies*, **58**, 407-36.

Richard, J-F. (1980), Models with Several Regimes and Changes in Exogeneity, *Review of Economic Studies*, **47**, 1-20.

Sargan, J.D. (1982), On Monte Carlo Estimates of Moments that are Infinite, *Advances in Econometrics*, **1**, 267-99.

Stock, J.H. (1987) Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors, *Econometrica*, **55**, 5, 1035-56.

Urbain J-P. (1992), On Weak Exogeneity in Error Correction Models, *Oxford Bulletin of Economics and Statistics*, **54**, 187-207.