MONASH

WP2/96

ISSN 1032-3813
ISBN 0 7326 0783 3

MONASH UNIVERSITY

ANCORA IMPARO

AUSTRALIA

PRINCIPAL COMPONENTS ANALYSIS OF COINTEGRATED TIME SERIES

c
David Harris

Working Paper 2/96
June 1996

DEPARTMENT OF ECONOMETRICS

Principal Components Analysis of Cointegrated Time Series

David Harris*

Department of Econometrics

Monash University

Clayton Victoria 3168

Australia

Phone: +61 3 9905 2453

Fax: +61 3 9905 5474

Email: David.Harris@BusEco.monash.edu.au

Abstract

This paper considers the analysis of cointegrated time series using principal components methods. These methods have the advantage of neither requiring the normalisation imposed by the triangular error correction model, nor the specification of a finite order vector autoregression. An asymptotically efficient estimator of the cointegrating vectors is given, along with tests for cointegration and tests of certain linear restrictions on the cointegrating vectors. An illustrative application is provided.

---

*I am grateful to Brett Inder for many helpful discussions, to Max King, Pentti Saikkonen and two anonymous referees for comments on a previous version of the paper, and to Janine Dixon for research assistance. A previous version of this paper was presented to the Econometric Society World Congress in Tokyo, 1995.

1 Introduction

It is convenient to consider estimators and tests for cointegrating vectors as falling into two categories. One category requires restrictions to be placed on the cointegrating vectors so that the individual elements of the vectors are identified. An obvious example of this is the use of a cointegrating regression, or, equivalently, a triangular error correction model. This topic has been studied extensively, with the properties of OLS regression being derived by Stock (1987), Phillips and Durlauf (1986) and Park and Phillips (1988). Modifications to OLS regression designed to give asymptotically efficient estimators were provided by Phillips and Hansen (1990), Park (1991), Saikkonen (1991), Engle and Yoo (1990), Phillips and Loretan (1992), Stock and Watson (1993) and Inder (1995), and maximum likelihood estimation of the triangular error correction model was considered by Phillips (1991a). Related estimators that do not refer explicitly to cointegrating regressions include Ahn and Reinsel (1990), Engle and Granger (1987) and Saikkonen (1992). The optimal properties of these estimators and resulting hypothesis tests are dependent on these identifying restrictions being valid. The advantage of most of the methods listed above is that they apply for a wide class of data generating processes, since it is often not necessary to fully parameterise the short run dynamic part of the model.

The second category of estimators of cointegrating vectors does not impose identifying restrictions. Instead, a basis for the space spanned by the cointegrating vectors is estimated, and any identifying restrictions can subsequently be tested. The most prominent of these methods is based on the reduced rank regression estimation of an error correction model, the theory for which was derived by Johansen (1988, 1991). The disadvantage of this method is that it requires the specification of a finite order VAR prior to estimation. Unlike many of the regression methods listed above, it is necessary to parameterise the short run dynamic part of the error correction model before estimation and inference can be carried out on the cointegrating vectors. In many applications, varying the length of the VAR can result in varying conclusions about the cointegrating vectors.

It is the purpose of this paper to provide an estimator of the cointegrating vectors that requires neither the prior imposition of identifying restrictions, nor the prior specification of a full model of the short run dynamics. This will be done by considering another method which falls in the second category of estimators, but which has received less attention in the literature. This is the principal components estimator, which was first used in the context of testing for common trends by Stock and Watson (1988). Its asymptotic properties have not been published (apart from Gonzalo (1994), who derived its asymptotic distribution for a particular data generating process), and its potential as an estimator of cointegrating vectors has not been fully explored. In section 2 of this paper we will show that the principal components estimator is consistent but asymptotically inefficient in general. We will then provide a modified principal components estimator that is asymptotically efficient for a wide class of data generating processes. This is done for three cases — where the model contains no deterministic components, where the model contains a level term, and where the model contains a linear trend. In sections 3 and 4 we consider some hypothesis testing problems which can be addressed using the modified principal components estimator. These are tests of certain linear restrictions on the cointegrating vectors and testing for cointegration. An illustrative application is provided in section 5.

2 Principal Components Estimation

In this section we will define the Principal Components (PC) estimator of the cointegrating vectors, derive its asymptotic distribution, and give a modified PC estimator that is asymptotically efficient in the sense of Saikkonen (1991). In the next subsection we consider the prototypical case of a cointegrated system with no deterministic terms, and we then allow for a level and a linear trend in the following subsection.

2.1 No Deterministic Terms

Suppose that we have a sample y_t, $(t = 1, \ldots, T)$, which is a p dimensional $I(1)$ cointegrated time series generated by

$$\beta' y_t = z_t, \tag{1}$$

$$\beta_\perp' \Delta y_t = w_t, \tag{2}$$

where β is a full rank $p \times r$ matrix of cointegrating vectors and $\beta_\perp$ is a full rank $p \times (p-r)$ matrix such that $\beta' \beta_\perp = 0$. Throughout we will assume that the cointegrating rank r is known, although in practice we could simply proceed conditional on the outcome of some statistical procedure to choose r. Further we assume that the $p \times 1$ random vector $\zeta_t = (z_t', w_t')'$ is a zero mean stationary time series satisfying the functional central limit theorem

$$T^{-1/2} \sum_{t=1}^{[Ts]} \zeta_t \xrightarrow{d} B_\zeta(s), \quad 0 < s \leq 1, \tag{3}$$

where B_ζ is a p dimensional Brownian motion with covariance matrix

$$\Omega_{\zeta\zeta} = \sum_{j=-\infty}^{\infty} E(\zeta_t \zeta_{t-j}'),$$

(see Phillips and Durlauf (1986) for more on the multivariate functional central limit theorem). We will also assume that the additional weak convergence result

$$T^{-1} \sum_{t=1}^{T} \left(\sum_{j=1}^{t} \zeta_j \right) \zeta_t' \xrightarrow{d} \int_0^1 B_\zeta(s)\, dB_\zeta(s) + \sum_{k=0}^{\infty} E\left(\zeta_{t-k} \zeta_t'\right) \tag{4}$$

holds (see Hansen (1992b) for appropriate conditions and a proof of this convergence). We then partition B_ζ and $\Omega_{\zeta\zeta}$ conformably with ζ_t as follows:

$$B_\zeta = \begin{pmatrix} B_z \\ B_w \end{pmatrix}, \qquad \Omega_{\zeta\zeta} = \begin{pmatrix} \Omega_{zz} & \Omega_{zw} \\ \Omega_{wz} & \Omega_{ww} \end{pmatrix}.$$

Note that the data generating process (1)–(2) is a natural extension of a cointegrating regression. However, one difference is that a cointegrating regression generally does not involve the arbitrary normalisation of $\beta' \beta_\perp = 0$. For example, if we let

$$y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}, \qquad \beta = \begin{pmatrix} I_r \\ -\beta_0 \end{pmatrix}, \qquad \beta_c = \begin{pmatrix} 0 \\ I_{p-r} \end{pmatrix},$$

where y_{1t} is $r \times 1$ and y_{2t} is $(p-r) \times 1$, then the data generating process

$$\beta' y_t = z_t, \tag{5}$$

$$\beta_c' \Delta y_t = w_t, \tag{6}$$

becomes

$$y_{1t} = \beta_0' y_{2t} + z_t, \tag{7}$$

$$\Delta y_{2t} = w_t, \tag{8}$$

which is the standard cointegrating regression data generating process. That is, if we want to use a cointegrating regression to obtain estimates of the r cointegrating vectors in β, we require the partitioning of y_t into y_{1t} and y_{2t} such that neither y_{1t} or y_{2t} are individually cointegrated. The properties of cointegrating regression estimators depend upon this identification being made correctly. We aim to use the principal components method to obtain an estimator of the space spanned by the columns of β that does not require the prior specification of these identifying restrictions. Our normalisation of $\beta'\beta_\perp = 0$ does not affect this space. Note also that we make no assumptions about the data generating process for ζ_t apart from it being stationary and satisfying the regularity conditions for (3) and (4) above.

The PC estimator of β in (1)–(2) is based on the sample cross-product matrix

$$S_{yy} = T^{-1} \sum_{t=1}^{T} y_t y_t'.$$

We define $\hat{\beta}$ to be the $p \times r$ matrix of eigenvectors corresponding to the smallest r eigenvalues satisfying

$$|\lambda I_p - S_{yy}| = 0, \tag{9}$$

normalised such that $\hat{\beta}'\hat{\beta} = I_r$. It follows that an estimator of $\beta_\perp$ (denoted $\hat{\beta}_\perp$) is the $p \times (p-r)$ matrix of eigenvectors corresponding to the largest $p-r$ eigenvalues. The intuition behind suggesting $\hat{\beta}$ is the same as that behind OLS in a cointegrating regression. Using OLS in a cointegrating regression such as (7) provides the linear combination of y_{1t} and y_{2t} with minimum sample variance, or, the "most stationary" linear combination. In (1)–(2), the $I(1)$ time series y_t has explosive variance, but the cointegrating vectors β give r linear combinations $z_t = \beta' y_t$ that have finite variance. We therefore choose the r linear combinations of y_t with minimum variance (i.e. the smallest r principal components) to be those corresponding to the estimated cointegrating vectors. On this basis, PC provides a natural extension of OLS estimation to models without the identifying restrictions necessary to write down a cointegrating regression.

Engle and Granger (1987) went some way towards this motivation of a PC approach in their discussion of possible estimators of cointegrating vectors. In particular, they demonstrated that the matrix $T^{-1}S_{yy}$ (M_T in their notation) converges to a singular matrix whose null space is the same as the space spanned by the cointegrating vectors. This fact lies at the heart of our proof of the consistency of the PC estimator given in Lemma 1 below. However, Engle and Granger (1987) proceeded to impose identifying restrictions on the cointegrating vectors before suggesting their least squares estimator, while we aim to avoid the prior imposition of these restrictions.

The PC estimator can also be given a reduced rank regression interpretation. It is well known (see Brillinger (1981) for example) that the principal components analysis of y_t is the same as a reduced rank regression of y_t on itself. In this case we choose the linear combinations of y_t that have minimum sample correlations with y_t, since they should correspond to the stationary linear combinations of y_t which will have zero correlation with the $I(1)$ time series y_t. With this interpretation, principal components has much in common with the reduced rank regression estimator of Yang (1994), which involves a reduced rank regression of (pre-whitened versions of) y_t on y_{t-1} and the choice of the linear combinations corresponding to the minimum correlations. The reduced rank regression estimator of Johansen (1988, 1991) uses a reduced rank regression of (pre-whitened versions of) Δy_t on y_{t-1} and the choice of the linear combinations corresponding to the maximum correlations.

We can now give the asymptotic properties of the PC estimator.

LEMMA 1 *If* y_t *is generated by (1)–(2) then* $\hat{\beta}$ *is super-consistent in the sense that*

$(I_p - \beta(\beta'\beta)^{-1}\beta')\,\hat{\beta} = O_p(T^{-1})$. *The asymptotic distribution of* $\hat{\beta}$ *is given by*

$$T\left(\hat{\beta}(\beta'\hat{\beta})^{-1}\beta'\beta - \beta\right) \xrightarrow{d} -\beta_{\perp}\left(\int B_w B_w'\right)^{-1}\left(\int B_w dB_z' + \Delta_{wz}\right), \qquad (10)$$

where $\Delta_{wz} = \sum_{j=0}^{\infty} E(w_{t-j}z_t')$. *Also,* $\hat{\beta}_{\perp}$ *is consistent in the sense that*

$$\left(I_p - \beta_{\perp}\left(\beta_{\perp}'\beta_{\perp}\right)^{-1}\beta_{\perp}'\right)\hat{\beta}_{\perp} = O_p\left(T^{-1}\right).$$

Note that we are estimating a basis for the space spanned by β, and not β itself, so we do not have a standard consistency result such as "$\hat{\beta} \xrightarrow{p} \beta$". Instead we have the result in Lemma 1 which shows that the columns of $\hat{\beta}$ asymptotically span the same space as the columns of β, and hence that a basis for the cointegrating space is consistently estimated. Similarly, the asymptotic distribution does not deal with the standardised error $T\left(\hat{\beta} - \beta\right)$, but rather with a re-arranged form of $T\left(\hat{\beta} - \beta\left(\beta'\beta\right)^{-1}\beta'\hat{\beta}\right)$. The distribution given in (10) is dependent on the nuisance parameters Ω_{wz} and Δ_{wz}. As a result, the distribution is not mixed normal (see Phillips (1991a)) and does not have a zero mean. If these nuisance parameters were zero, then the distribution would be normal with a zero mean (conditional on B_w), and, as shown by Saikkonen (1991), $\hat{\beta}$ would then be an asymptotically efficient estimator. These results correspond to those found by Park and Phillips (1988) for the OLS estimation of a cointegrating regression. As suggested by (7) and (8), z_t plays the role of the regression error term, while w_t drives the $I(1)$ part of the system. Both OLS and PC are asymptotically efficient if z_t and w_t are uncorrelated at all lags (or at least $\Omega_{wz} = \Delta_{wz} = 0$), but their distributions are dependent on nuisance parameters otherwise. Phillips and Hansen (1990) and Park (1992) have suggested semi-parametric modifications to OLS that always produce asymptotically efficient estimators, and we will now give an asymptotically efficient modified PC estimator analogous to Park's estimator.

As can be seen in the proof of the previous lemma, $\hat{\beta}$ can be given the representation

$$T\left(\hat{\beta}\left(\beta'\hat{\beta}\right)^{-1}\beta'\beta - \beta\right) = -\beta_{\perp}\left(T^{-1}\beta_{\perp}'S_{yy}\beta_{\perp}\right)^{-1}\beta_{\perp}'S_{yy}\beta + O_p\left(T^{-1}\right), \qquad (11)$$

and the nuisance parameter dependency of the asymptotic distribution arises from the intermediate result

$$\beta_{\perp}'S_{yy}\beta \xrightarrow{d} \int B_w dB_z' + \Delta_{wz}. \qquad (12)$$

In particular, it is the correlation between B_w and B_z (given by Ω_{wz}) and the presence of Δ_{wz} that causes the problems. We define a transformation of the data which directly addresses these two nuisance parameters. The transformation makes use of consistent (but not necessarily asymptotically efficient) estimators of β and $\beta_\perp$, and it follows from Lemma 1 that PC provides consistent estimators $\hat{\beta}$ and $\hat{\beta}_\perp$. We also make use of consistent estimators of the nuisance parameters which have the general form

$$\hat{\Omega}_{ab} = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{m}\right)\hat{\Gamma}_{ab}(j), \qquad \hat{\Delta}_{ab} = \sum_{j=0}^{T-1} k\left(\frac{j}{m}\right)\hat{\Gamma}_{ab}(j), \tag{13}$$

where

$$\hat{\Gamma}_{ab}(j) = T^{-1}\sum_{t=1}^{T} a_{t-j}b_t',$$

a_t and b_t are any time series, $k(.)$ is a lag window and m is a bandwidth parameter such that $m \to \infty$ and $m/T \to 0$ as $T \to \infty$. Andrews (1991) gives the optimal choice of $k(.)$ and m for estimating Ω_{ab} in the sense of the minimisation of the asymptotic mean squared error of the estimator. This corresponds to using the Quadratic Spectral lag window and an automatic data-dependent bandwidth parameter, formulae for which are given by Andrews (1991). Furthermore, Hansen (1992a, section 2.1) advocates the use of these optimal estimators when calculating the Fully Modified OLS estimator of Phillips and Hansen (1990), and provides additional exposition of their use in this context. Given that our aim is to suggest a modification to the PC estimator analogous to the Phillips and Hansen (1990) modification to the OLS estimator, these optimal estimators should be well suited. If we let

$$\hat{z}_t = \hat{\beta}' y_t, \qquad \hat{w}_t = \hat{\beta}_\perp' \Delta y_t, \qquad \hat{\zeta}_t = \begin{pmatrix} \hat{z}_t \\ \hat{w}_t \end{pmatrix},$$

then the transformation to y_t is

$$y_t^* = y_t - \hat{\beta}\left(\hat{\beta}'\hat{\beta}\right)^{-1}\hat{\Omega}_{zw}\hat{\Omega}_{ww}^{-1}\hat{w}_t - \hat{\beta}_\perp\left(\hat{\beta}_\perp'\hat{\beta}_\perp\right)^{-1}\hat{\Delta}_{w\zeta}\hat{S}_{\zeta\zeta}^{-1}\hat{\zeta}_t, \tag{14}$$

where $\hat{S}_{\zeta\zeta} = T^{-1}\sum\hat{\zeta}_t\hat{\zeta}_t$. The first modification in (14) is designed to remove the nuisance parameter Ω_{wz}, while the second modification removes Δ_{wz}. Notice that y_t^* is invariant to the normalisation of $\hat{\beta}$ and $\hat{\beta}_\perp$ in the sense that $\hat{\beta}$ (or $\hat{\beta}_\perp$) may be replaced by $\hat{\beta}\kappa$

$(\hat{\beta}_\perp \kappa)$, where κ is a full rank $r \times r$ $((p-r) \times (p-r))$ matrix, without affecting y_t^*. The modified PC estimator $\hat{\beta}^*$ is defined to be the $p \times r$ matrix of eigenvectors corresponding to the smallest r eigenvalues satisfying

$$|\lambda I_p - S_{yy}^*| = 0, \tag{15}$$

where $S_{yy}^* = T^{-1} \sum y_t^* y_t^{*\prime}$. That is, we simply apply the PC estimator to y_t^* instead of y_t. The asymptotic properties of this estimator are given by the following theorem.

THEOREM 1 *The modified PC estimator $\hat{\beta}^*$ is consistent, and its asymptotic distribution is given by*

$$T\left(\hat{\beta}^* \left(\beta'\hat{\beta}^*\right)^{-1} \beta'\beta - \beta\right) \xrightarrow{d} -\beta_\perp \left(\int B_w B_w'\right)^{-1} \int B_w dB_z^{*\prime}, \tag{16}$$

where $B_z^* = B_z - \Omega_{zw}\Omega_{ww}^{-1} B_w$ is independent of B_w.

Clearly the asymptotic distribution of $\hat{\beta}^*$ does not depend on the nuisance parameters Ω_{wz} and Δ_{wz} and it follows that $\hat{\beta}^*$ is asymptotically efficient. Conditional on B_w, the distribution specified by the right hand side of (16) is normal with mean 0 and variance matrix $\beta_\perp \left(\int B_w B_w'\right)^{-1} \beta_\perp' \otimes \Omega_{zz}$, a fact we will use in defining hypothesis tests for β below.

2.2 Deterministic Terms

We now consider the principal components estimation of the cointegrating vectors where deterministic terms may also enter the system. In particular, we consider the two cases of

$$y_t = \mu + x_t, \tag{17}$$

$$y_t = \mu + \delta t + x_t, \tag{18}$$

where x_t is an $I(1)$ time series generated by

$$\beta' x_t = z_t, \tag{19}$$

$$\beta_\perp' \Delta x_t = w_t, \tag{20}$$

and $\zeta_t = (z_t', w_t')'$ is a zero mean $I(0)$ time series that satisfies the functional central limit theorem, just as in the previous section. Then y_t in both (17) and (18) is cointegrated with cointegrating vectors given by the columns of β, since $\beta' y_t$ is stationary about a constant in (17), and stationary about a linear trend in (18).

When y_t is generated by (17) the PC estimator $\bar{\beta}$ is the $p \times r$ matrix of eigenvectors corresponding to the smallest r eigenvalues satisfying

$$|\lambda I_p - \bar{S}_{xx}| = 0, \tag{21}$$

where $\bar{S}_{xx} = T^{-1} \sum_{t=1}^{T} \bar{x}_t \bar{x}_t'$ and $\bar{x}_t = y_t - T^{-1} \sum_{t=1}^{T} y_t$. When y_t is generated by (18) the PC estimator $\tilde{\beta}$ is the $p \times r$ matrix of eigenvectors corresponding to the smallest r eigenvalues satisfying

$$|\lambda I_p - \tilde{S}_{xx}| = 0, \tag{22}$$

where $\tilde{S}_{xx} = T^{-1} \sum_{t=1}^{T} \tilde{x}_t \tilde{x}_t'$ and $\tilde{x}_t$ are the residuals from a de-trending regression of y_t on a constant and time trend.

It is shown in Theorem 2 below that both $\bar{\beta}$ and $\tilde{\beta}$ suffer from the same problems as $\hat{\beta}$ in the previous section. That is, although both are consistent, their asymptotic distributions are dependent on the nuisance parameters Ω_{wz} and Δ_{wz}, and they are neither asymptotically normal nor asymptotically efficient. This is to be expected since the source of these problems is the correlation between z_t and w_t in (19) and (20). However we can use the modified PC estimator from the previous section to provide asymptotically efficient alternatives to $\bar{\beta}$ and $\tilde{\beta}$.

In the case of $\bar{\beta}$, we define the variables

$$\bar{z}_t = \bar{\beta}' \bar{x}_t, \qquad \bar{w}_t = \bar{\beta}_\perp' \Delta \bar{x}_t, \qquad \bar{\zeta}_t = \begin{pmatrix} \bar{z}_t \\ \bar{w}_t \end{pmatrix},$$

and the data transformation

$$\bar{x}_t^* = \bar{x}_t - \bar{\beta} \left(\bar{\beta}' \bar{\beta}\right)^{-1} \bar{\Omega}_{zw} \bar{\Omega}_{ww}^{-1} \bar{w}_t - \bar{\beta}_\perp \left(\bar{\beta}_\perp' \bar{\beta}_\perp\right)^{-1} \bar{\Delta}_{w\zeta} \bar{S}_{\zeta\zeta}^{-1} \bar{\zeta}_t. \tag{23}$$

In (23) we make use of the consistent estimator $\bar{\beta}_\perp$, which is the estimator of $\beta_\perp$ obtained using the eigenvectors corresponding to the largest $p-r$ eigenvalues satisfying (21), and the nuisance parameter estimators $\bar{\Omega}_{zw}$, $\bar{\Omega}_{ww}$ and $\bar{\Delta}_{w\zeta}$ which are of the form (13) and calculated using $\bar{z}_t$ and $\bar{w}_t$. Then the modified PC estimator $\bar{\beta}^*$ is the $p \times r$ matrix of eigenvectors corresponding to the smallest r eigenvalues satisfying

$$|\lambda I_p - \bar{S}^*_{xx}| = 0, \tag{24}$$

where $\bar{S}^*_{xx} = T^{-1} \sum_{t=1}^{T} \bar{x}^*_t \bar{x}^{*\prime}_t$.

We proceed similarly to modify $\tilde{\beta}$. Define

$$\tilde{z}_t = \tilde{\beta}' \tilde{x}_t, \qquad \tilde{w}_t = \tilde{\beta}' \Delta \tilde{x}_t, \qquad \tilde{\zeta}_t = \begin{pmatrix} \tilde{z}_t \\ \tilde{w}_t \end{pmatrix},$$

and

$$\tilde{x}^*_t = \tilde{x}_t - \tilde{\beta} \left(\tilde{\beta}' \tilde{\beta} \right)^{-1} \tilde{\Omega}_{zw} \tilde{\Omega}^{-1}_{ww} \tilde{w}_t - \tilde{\beta}_\perp \left(\tilde{\beta}'_\perp \tilde{\beta}_\perp \right)^{-1} \tilde{\Delta}_{w\zeta} \tilde{S}^{-1}_{\zeta\zeta} \tilde{\zeta}_t. \tag{25}$$

The modified PC estimator $\tilde{\beta}^*$ is the $p \times r$ matrix of eigenvectors corresponding to the smallest r eigenvalues satisfying

$$|\lambda I_p - \tilde{S}^*_{xx}| = 0, \tag{26}$$

where $\tilde{S}^*_{xx} = T^{-1} \sum_{t=1}^{T} \tilde{x}^*_t \tilde{x}^{*\prime}_t$. The asymptotic properties of these estimators are given in the following theorem.

THEOREM 2 *(i) If y_t is generated by (17), (19)-(20) then $\bar{\beta}$ and $\bar{\beta}^*$ are consistent and have asymptotic distributions given by*

$$T \left(\bar{\beta} \left(\beta' \bar{\beta} \right)^{-1} \beta' \beta - \beta \right) \xrightarrow{d} -\beta_\perp \left(\int \bar{B}_w \bar{B}'_w \right)^{-1} \left(\int \bar{B}_w dB'_z + \Delta_{wz} \right), \tag{27}$$

$$T \left(\bar{\beta}^* \left(\beta' \bar{\beta}^* \right)^{-1} \beta' \beta - \beta \right) \xrightarrow{d} -\beta_\perp \left(\int \bar{B}_w \bar{B}'_w \right)^{-1} \int \bar{B}_w dB^{*\prime}_z, \tag{28}$$

where $\bar{B}_w(s) = B_w(s) - \int B_w$.
(ii) If y_t is generated by (18), (19)-(20) then $\tilde{\beta}$ and $\tilde{\beta}^*$ are consistent and have asymptotic distributions given by

$$T \left(\tilde{\beta} \left(\beta' \tilde{\beta} \right)^{-1} \beta' \beta - \beta \right) \xrightarrow{d} -\beta_\perp \left(\int \tilde{B}_w \tilde{B}'_w \right)^{-1} \left(\int \tilde{B}_w dB'_z + \Delta_{wz} \right), \tag{29}$$

$$T\left(\tilde{\beta}^{*}\left(\beta'\tilde{\beta}^{*}\right)^{-1}\beta'\beta-\beta\right)\xrightarrow{d}-\beta_{\perp}\left(\int\tilde{B}_{w}\tilde{B}_{w}'\right)^{-1}\int\tilde{B}_{w}dB_{z}^{*\prime},\tag{30}$$

where $\tilde{B}_w(s) = B_w(s) + 2(3s-2)\int B_w - 6(2s-1)\int rB_w$.

Equations (27) and (29) provide a natural extension of the results in Theorems 3.2(a) and 3.3(a) of Park and Phillips (1988) for the OLS estimation of cointegrating regressions to the case of PC estimation. Clearly both $\bar{\beta}$ and $\tilde{\beta}$ suffer from the same nuisance parameter dependencies as $\hat{\beta}$. Equations (28) and (30) show that the modified PC estimators $\bar{\beta}^*$ and $\tilde{\beta}^*$ are asymptotically efficient.

3 Tests of Linear Restrictions on the Cointegrating Vectors

We consider here two types of linear restrictions on the cointegrating vectors. Both types of restrictions have been used by Johansen and Juselius (1992) among others. The first hypothesis is that

$$H_0 : \beta = H_1\varphi,\tag{31}$$

where H_1 is a known $p \times s$ matrix, φ is an $s \times r$ matrix of unknown parameters and $r \leq s \leq p$. Equation (31) states that the cointegrating space lies within the space spanned by the columns of H_1. If $s = r$ then the null hypothesis fully specifies β, and if $s = p$ then no restriction is placed on β. To test (31) we introduce the $p \times (p-s)$ matrix J_1 chosen such that $J_1'H_1 = 0$. For example, J_1 can be obtained as the $p \times (p-s)$ matrix of eigenvectors corresponding to the zero eigenvalues of H_1H_1'. Then β can always be written

$$\beta = H_1\varphi + J_1\phi,\tag{32}$$

where ϕ is an unknown $(p-s) \times r$ matrix. Clearly, if (31) is true, then $\phi = 0$. Alternatively, $J_1'\beta = 0$ if (31) is true, and $J_1'\beta \neq 0$ if (31) is not true. We will therefore base our tests on the significance of the sample quantities $J_1'\hat{\beta}^*$, $J_1'\bar{\beta}^*$ and $J_1'\tilde{\beta}^*$. The test statistics and their asymptotic properties are given in the following theorem.

THEOREM 3 *(i) If y_t is generated by (1)–(2), then the test statistic for testing (31) is*

$$\hat{s}_1 = T\ \mathrm{tr}\hat{\Omega}_{zz}^{*-1}\hat{\beta}^{*\prime}J_1\left(J_1'S_{yy}^{-1}J_1\right)^{-1}J_1'\hat{\beta}^{*},$$

where $\hat{\Omega}^*_{zz} = \hat{\Omega}_{zz} - \hat{\Omega}_{zw}\hat{\Omega}^{-1}_{ww}\hat{\Omega}_{wz}$. *If* H_0 *is true then* $\hat{s}_1 \overset{A}{\sim} \chi^2_{r(p-s)}$. *If* H_0 *is false then* $\hat{s}_1 = O_p(T)$.

(ii) If y_t *is generated by (17), (19)-(20), then the test statistic for testing (31) is*

$$\bar{s}_1 = T \ \mathrm{tr}\bar{\Omega}^{*-1}_{zz}\bar{\beta}^{*\prime}J_1\left(J_1'\bar{S}^{-1}_{xx}J_1\right)^{-1}J_1'\bar{\beta}^*,$$

where $\bar{\Omega}^*_{zz} = \bar{\Omega}_{zz} - \bar{\Omega}_{zw}\bar{\Omega}^{-1}_{ww}\bar{\Omega}_{wz}$. *If* H_0 *is true then* $\bar{s}_1 \overset{A}{\sim} \chi^2_{r(p-s)}$. *If* H_0 *is false then* $\bar{s}_1 = O_p(T)$.

(iii) If y_t *is generated by (18), (19)-(20), then the test statistic for testing (31) is*

$$\tilde{s}_1 = T \ \mathrm{tr}\tilde{\Omega}^{*-1}_{zz}\tilde{\beta}^{*\prime}J_1\left(J_1'\tilde{S}^{-1}_{xx}J_1\right)^{-1}J_1'\tilde{\beta}^*,$$

where $\tilde{\Omega}^*_{zz} = \tilde{\Omega}_{zz} - \tilde{\Omega}_{zw}\tilde{\Omega}^{-1}_{ww}\tilde{\Omega}_{wz}$. *If* H_0 *is true then* $\tilde{s}_1 \overset{A}{\sim} \chi^2_{r(p-s)}$. *If* H_0 *is false then* $\tilde{s}_1 = O_p(T)$.

These three tests can be considered as Wald-type tests in the sense that each statistic tests the significance of the difference between an unrestricted estimate of β and the null hypothesis. The fact that each has an asymptotic χ^2 distribution under the null hypothesis is a direct consequence of $\hat{\beta}^*$, $\bar{\beta}^*$ and $\tilde{\beta}^*$ having asymptotically conditionally normal distributions. The use of asymptotically inefficient estimators such as $\hat{\beta}$, $\bar{\beta}$ and $\tilde{\beta}$ in the test statistics would not result in tests with asymptotic χ^2 distributions. The derivation of the asymptotic distributions for each test statistic is unaffected by the replacement of S_{yy}, $\bar{S}_{xx}$ and $\tilde{S}_{xx}$ by S^*_{yy}, $\bar{S}^*_{xx}$ and $\tilde{S}^*_{xx}$ respectively. The presence of these terms partially standardises the asymptotic normal distributions of the estimators, and the choice may potentially be made on the grounds of finite sample performance. Given that the test statistics diverge under the alternative hypothesis, rejecting the null hypothesis for values of the test statistics larger than the appropriate χ^2 critical values will provide consistent tests. Note that the test statistics are invariant to the normalisation of both J_1 and the estimator of β. This is appropriate since the null hypothesis is a test about the space spanned by β, and we would not want a different test from a re-normalised version of the null. Similarly, the test statistic is invariant to the method of constructing J_1 from H_1.

The second hypothesis we consider is

$$H_0 : \beta = (H_2, \theta), \tag{33}$$

where H_2 is a known $p \times r_1$ matrix, θ is a $p \times (r - r_1)$ matrix of unknown parameters, and $0 \leq r_1 \leq r$. The hypothesis specifies r_1 columns of β and leaves the other columns unspecified. If $r = r_1$ then β is completely specified and (33) is equivalent to (31) with $s = r$. Note that (33) is similar to (31) in the sense that in (31) β must lie in the space spanned by the columns of H_1, while in (33) H_2 must lie in the space spanned by the columns of β. Tests of the two hypotheses are therefore quite similar. If (33) is true then $\beta'_\perp H_2 = 0$ since $\beta'_\perp \beta = 0$. We therefore base our tests of (33) on the significance of the sample quantities $\hat{\beta}^{*\prime}_\perp H_2$, $\bar{\beta}^{*\prime}_\perp H_2$ and $\tilde{\beta}^{*\prime}_\perp H_2$, where $\hat{\beta}^{*}_\perp$, $\bar{\beta}^{*}_\perp$ and $\tilde{\beta}^{*}_\perp$ are calculated from the corresponding efficient estimators of the cointegrating vectors to satisfy $\hat{\beta}^{*\prime}_\perp \hat{\beta}^* = 0$, $\bar{\beta}^{*\prime}_\perp \bar{\beta}^* = 0$ and $\tilde{\beta}^{*\prime}_\perp \tilde{\beta}^* = 0$ respectively. The tests are given in the following theorem.

THEOREM 4 *(i) If y_t is generated by (1)–(2), then the test statistic for testing (33) is*

$$\hat{s}_2 = T \ \mathrm{tr} \hat{\Omega}^{-1}_{HH} H'_2 \hat{\beta}^*_\perp \left(\hat{\beta}^{*\prime}_\perp S^{-1}_{yy} \hat{\beta}^*_\perp \right)^{-1} \hat{\beta}'_\perp H_2,$$

where $\hat{\Omega}_{HH} = H'_2 \hat{\beta}^* \left(\hat{\beta}^{*\prime} \hat{\beta}^* \right)^{-1} \hat{\Omega}^*_{zz} \left(\hat{\beta}^{*\prime} \hat{\beta}^* \right)^{-1} \hat{\beta}^{*\prime} H_2$. *If H_0 is true then* $\hat{s}_2 \overset{A}{\sim} \chi^2_{r_1(p-r)}$. *If H_0 is false then* $\hat{s}_2 = O_p \left(T^2 \right)$.

(ii) If y_t is generated by (17), (19)–(20), then the test statistic for testing (33) is

$$\bar{s}_2 = T \ \mathrm{tr} \bar{\Omega}^{-1}_{HH} H'_2 \bar{\beta}^*_\perp \left(\bar{\beta}^{*\prime}_\perp \bar{S}^{-1}_{xx} \bar{\beta}^*_\perp \right)^{-1} \bar{\beta}'_\perp H_2,$$

where $\bar{\Omega}_{HH} = H'_2 \bar{\beta}^* \left(\bar{\beta}^{*\prime} \bar{\beta}^* \right)^{-1} \bar{\Omega}^*_{zz} \left(\bar{\beta}^{*\prime} \bar{\beta}^* \right)^{-1} \bar{\beta}^{*\prime} H_2$. *If H_0 is true then* $\bar{s}_2 \overset{A}{\sim} \chi^2_{r_1(p-r)}$. *If H_0 is false then* $\bar{s}_2 = O_p \left(T^2 \right)$.

(iii) If y_t is generated by (18), (19)–(20), then the test statistic for testing (33) is

$$\tilde{s}_2 = T \ \mathrm{tr} \tilde{\Omega}^{-1}_{HH} H'_2 \tilde{\beta}^*_\perp \left(\tilde{\beta}^{*\prime}_\perp \tilde{S}^{-1}_{xx} \tilde{\beta}^*_\perp \right)^{-1} \tilde{\beta}'_\perp H_2,$$

where $\tilde{\Omega}_{HH} = H'_2 \tilde{\beta}^* \left(\tilde{\beta}^{*\prime} \tilde{\beta}^* \right)^{-1} \tilde{\Omega}^*_{zz} \left(\tilde{\beta}^{*\prime} \tilde{\beta}^* \right)^{-1} \tilde{\beta}^{*\prime} H_2$. *If H_0 is true then* $\tilde{s}_2 \overset{A}{\sim} \chi^2_{r_1(p-r)}$. *If H_0 is false then* $\tilde{s}_2 = O_p \left(T^2 \right)$.

This theorem provides Wald-type tests of (33) that have asymptotic χ^2 distributions and that are consistent when we reject the null for large values of the test statistics.

The comments following Theorem 3 apply equally here. One difference between the tests of Theorems 3 and 4 is that the rates of divergence of the test statistics under the alternative hypotheses are $O_p(T^2)$ for Theorem 4 rather than $O_p(T)$ for Theorem 3. To explore this point more fully we give the asymptotic distributions of the $\hat{s}_1$ and $\hat{s}_2$ under sequences of local alternatives.

LEMMA 2 *(i) Under the sequence of local alternatives*

$$H_{AT} : J_1'\beta = T^{-1}A_1,$$

where A_1 is a fixed $p-s \times r$ matrix, $\hat{s}_1$ has a non-central $\chi^2_{r(p-s)}$ distribution conditional on B_w, with non-centrality parameter

$$\mathrm{tr}\Omega_{zz}^{-1}A_1'\left[J_1'\beta_\perp\left(\int B_wB_w'\right)^{-1}\beta_\perp'J_1\right]^{-1}A_1.$$

(ii) Under the sequence of local alternatives

$$H_{AT} : H_2'\beta_\perp = T^{-1}A_2,$$

where A_2 is a fixed $r_1 \times (p-r)$ matrix, $\hat{s}_2$ has a non-central $\chi^2_{r_1(p-r)}$ distribution conditional on B_w, with non-centrality parameter

$$\mathrm{tr}\Omega_{HH}^{-1}A_2'\left[\beta_\perp'\beta_\perp\left(\int B_wB_w'\right)^{-1}\beta_\perp'\beta_\perp\right]^{-1}A_2.$$

This result shows that both $\hat{s}_1$ and $\hat{s}_2$ have power against alternatives that approach the null at rate $O(T)$, and hence the faster rate of divergence of $\hat{s}_2$ under a fixed alternative does not lead to asymptotic power of a higher order. The conditional distributions are presented for convenience, and the calculation of the asymptotic local power functions would require the evaluation of the non-standard unconditional distributions. The unconditional distribution of $\hat{s}_1$ is a mixture of non-central $\chi^2_{r(p-s)}$ random variables, averaged over realisations of $\int B_wB_w'$, while $\hat{s}_2$ is similarly a mixture of non-central $\chi^2_{r_1(p-r)}$ random variables.

4 Testing for Cointegration

All estimators of the cointegrating vectors require prior knowledge of the cointegrating rank r. In this section we provide a test for cointegration based on the efficient estimators in section 2. This test may be used as a diagnostic to check if a predetermined r is appropriate, or it may be used to form a sequence of tests to select r if it is unknown. The test we suggest is an extension of the residual based tests of the null hypothesis of cointegration in cointegrating regressions developed by Harris and Inder (1994) and Shin (1994), which in turn extended tests of the null hypothesis of stationarity of a univariate time series given by Kwiatkowski, Phillips, Schmidt and Shin (1992). Choi and Ahn (1995) have also suggested tests for the null hypothesis of cointegration that are applicable in systems of cointegrating regressions. They extend Shin's (1994) tests to the case of a multivariate cointegrating regression, and do the same to the test of the null hypothesis of stationarity of Choi (1994). Of course, these tests require the imposition of identifying restrictions to specify the cointegrating regressions, while our tests suggested here using principal components analysis do not require these restrictions.

Suppose the null hypothesis is that y_t has r cointegrating vectors, and the alternative is that there are fewer than r cointegrating vectors. We note that this selection of hypotheses is the opposite of that used for tests for cointegration by Johansen (1988, 1991) in the sense that, in Johansen's tests, the alternative hypothesis is that there are greater than r cointegrating vectors. It should be relatively straightforward to derive tests against this alternative by adapting the residual based tests of Phillips and Ouliaris (1990) to principal components estimation. However, since the theory for such tests would not be derived from Theorems 1 and 2, we do not pursue them here.

We define the time series

$$\hat{z}_t^* = \hat{\beta}^{*\prime} y_t^*, \quad \bar{z}_t^* = \bar{\beta}^{*\prime} \bar{x}_t^*, \quad \tilde{z}_t^* = \tilde{\beta}^{*\prime} \tilde{x}_t^*.$$

Under the null hypothesis the relevant time series will be $I(0)$, while under the alternative it will have at least one $I(1)$ component. We therefore construct tests for the null hypothesis of the stationarity of $\hat{z}_t^*$, $\bar{z}_t^*$ and $\tilde{z}_t^*$. Let

$$\hat{S}_t = \sum_{j=1}^{t} \hat{z}_j^*, \quad \bar{S}_t = \sum_{j=1}^{t} \bar{z}_j^*, \quad \tilde{S}_t = \sum_{j=1}^{t} \tilde{z}_j^*,$$

and define the test statistics

$$\begin{aligned}
\hat{c} &= T^{-2} \sum_{t=1}^{T} \hat{S}_t' \hat{\Omega}_{zz}^{*-1} \hat{S}_t, \\
\bar{c} &= T^{-2} \sum_{t=1}^{T} \bar{S}_t' \bar{\Omega}_{zz}^{*-1} \bar{S}_t, \\
\tilde{c} &= T^{-2} \sum_{t=1}^{T} \tilde{S}_t' \tilde{\Omega}_{zz}^{*-1} \tilde{S}_t.
\end{aligned}$$

To present the asymptotic properties of these test statistics, we first define a $p \times 1$ standard Brownian motion W, partitioned as $(W_1', W_2')'$ conformably with B_ς. Now let

$$\bar{W}_2(s) = W_2(s) - \int W_2,$$

$$\tilde{W}_2(s) = W_2(s) + 2(3s-2) \int W_2 - 6(2s-1) \int rW_2,$$

be de-meaned and de-trended W_2 respectively, and define

$$V(s) = W_1(s) - \int dW_1 W_2' \left(\int W_2 W_2' \right)^{-1} \int_0^s W_2(r)\, dr,$$

$$\bar{V}(s) = W_1(s) - \int dW_1 \bar{W}_2' \left(\int \bar{W}_2 \bar{W}_2' \right)^{-1} \int_0^s \bar{W}_2(r)\, dr,$$

$$\tilde{V}(s) = W_1(s) - \int dW_1 \tilde{W}_2' \left(\int \tilde{W}_2 \tilde{W}_2' \right)^{-1} \int_0^s \tilde{W}_2(r)\, dr.$$

The asymptotic properties of the test statistics are now given by the following theorem.

THEOREM 5 *(i) If y_t is generated by (1)–(2) then, under the null hypothesis of r cointegrating vectors,*

$$\hat{c} \xrightarrow{d} \int_0^1 V(s)' V(s)\, ds.$$

(ii) If y_t is generated by (17), (19)–(20), then under the null

$$\bar{c} \xrightarrow{d} \int_0^1 \bar{V}(s)' \bar{V}(s)\, ds.$$

(iii) If y_t is generated by (18), (19)–(20), then under the null

$$\tilde{c} \xrightarrow{d} \int_0^1 \tilde{V}(s)' \tilde{V}(s)\, ds.$$

(iv) In each case, the test statistic is $O_p(T/m)$ under the alternative of fewer than r cointegrating vectors, where m is the bandwidth parameter used in calculating $\hat{\Omega}^*_{zz}$, $\bar{\Omega}^*_{zz}$ and $\tilde{\Omega}^*_{zz}$ respectively.

The null distributions found in this theorem are non-standard, and critical values are given in Tables 1a, 1b and 1c. In view of part (iv) of this theorem, a consistent test is obtained by rejecting the null hypothesis for values of the test statistics that are larger than the relevant critical values.

5 Application

In this section we provide an illustrative application of the methods suggested in the paper. King, Plosser, Stock and Watson (1991) give a simple real business cycle model involving private output, consumption and investment. An implication which can be drawn from this model is that the logs of output and consumption should be cointegrated with cointegrating vector $(1, -1)$, and similarly the logs of output and investment should be cointegrated with cointegrating vector $(1, -1)$. Therefore, considered as a vector, the time series $y_t = (C_t, I_t, Y_t)'$ should have a cointegrating rank of $r = 2$ with cointegrating vectors

$$\beta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}. \tag{34}$$

We use the Australian equivalent of the data analysed by King et al. That is, C_t, I_t and Y_t are logs of private final consumption, private investment and private output respectively. Each series is measured per capita and in constant 1989/90 dollars, and the observations are quarterly, seasonally adjusted and sampled from June 1971 to September 1994. Private output represents the output of the private sector only and is constructed by subtracting government expenditure from Australian gross domestic product. The data are graphed in Figure 1.

Since we have a hypothesis that there should be two cointegrating vectors, we set $r = 2$ and calculate the asymptotically efficient PC estimator from section 2:

$$\bar{\beta}^* = \begin{pmatrix} 0.62 & -0.48 \\ 0.18 & 0.86 \\ -0.76 & -0.19 \end{pmatrix}.$$

As mentioned in section 2.1, the nuisance parameter estimators in equation (13) were calculated using the Quadratic Spectral lag window and the optimal data-dependent bandwidth parameter based on univariate $AR(1)$ approximating equations. Note that $\bar{\beta}^*$ is an estimator of a basis for the cointegrating space and is not directly comparable with (34). However, re-normalising the estimate we find

$$\bar{\beta}^* = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -0.99 & -0.78 \end{pmatrix}$$

which is suggestive of (34). Before testing whether β in (34) and $\bar{\beta}^*$ span the same space, we test whether the linear combinations of y_t specified by $\bar{\beta}^*$ are $I(0)$ using the test in section 4. That is, we test the null hypothesis that $r = 2$ against the alternative that $r < 2$. In this sense, the cointegration test is a diagnostic test of our original belief that $r = 2$, which is essentially the same way that residual based tests for cointegration are used in regression based methods. The calculated test statistic is $\bar{c} = 0.27$, which when compared with the 5% critical value of 0.51, supports our assertion that $r = 2$.

We can now test some hypotheses about the cointegrating space. Our first hypothesis fully specifies the cointegrating space:

$$H_0 : \beta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}.$$

This is of the form of the second hypothesis considered in section 3 (see equation (33)), and we calculate the test statistic $\bar{s}_2 = 4.03$. Since $r_1 = 2$, there are 2 degrees of freedom for the test, and the χ^2 critical value is 5.99. We therefore accept that the cointegrating vectors have the structure given in (34). We can also individually test whether each column in (34) is a valid cointegrating vector. The calculated statistic for $(1, 0, -1)'$ is 0.018, and for $(0, 1, -1)'$ it is 3.27. Both tests have 1 degree of freedom, so in both cases we accept the hypothesized vector as a cointegrating vector. As an illustrative example of the first hypothesis considered in section 3, we can test whether I_t can be excluded from the cointegrating relations. The null hypothesis takes the form

$$H_0 : \beta = H_1\varphi, \quad H_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix},$$

where φ is a 2×2 matrix of unknown parameters. It follows that $J_1 = (0, 0, 1)'$ and the calculated test statistic is $\bar{s}_1 = 26.57$. The test has 2 degrees of freedom so we reject the hypothesis that I_t does not enter the cointegrating relations, as would be expected.

Table 1a. Critical Values for $\hat{c}$

Size = 10%		Number of Variables (p)					
		1	2	3	4	5	6
	1	1.17	0.86	0.62	0.47	0.38	0.31
Cointegrating	2		2.06	1.49	1.10	0.84	0.67
rank	3			2.82	2.06	1.54	1.19
(r)	4				3.51	2.64	1.95
	5					4.23	3.16
	6						4.93

Size = 5%		Number of Variables (p)					
		1	2	3	4	5	6
	1	1.64	1.21	0.89	0.68	0.54	0.44
Cointegrating	2		2.63	1.95	1.46	1.12	0.89
rank	3			3.45	2.56	1.97	1.52
(r)	4				4.21	3.23	2.42
	5					4.99	3.81
	6						5.71

Size = 1%		Number of Variables (p)					
		1	2	3	4	5	6
	1	2.75	2.09	1.61	1.25	0.99	0.80
Cointegrating	2		3.92	3.04	2.32	1.81	1.45
rank	3			4.78	3.79	2.97	2.36
(r)	4				5.73	4.62	3.54
	5					6.60	5.28
	6						7.47

Table 1b. Critical Values for $\bar{c}$

Size = 10%		Number of Variables (p)					
		1	2	3	4	5	6
	1	0.35	0.24	0.16	0.12	0.093	0.076
Cointegrating	2		0.61	0.42	0.30	0.22	0.17
rank	3			0.84	0.59	0.42	0.31
(r)	4				1.06	0.76	0.55
	5					1.28	0.93
	6						1.48

Size = 5%		Number of Variables (p)					
		1	2	3	4	5	6
	1	0.47	0.32	0.22	0.16	0.12	0.097
Cointegrating	2		0.74	0.53	0.37	0.27	0.21
rank	3			1.00	0.72	0.52	0.38
(r)	4				1.23	0.91	0.67
	5					1.47	1.09
	6						1.69

Size = 1%		Number of Variables (p)					
		1	2	3	4	5	6
	1	0.74	0.55	0.39	0.28	0.20	0.15
Cointegrating	2		1.07	0.81	0.60	0.43	0.32
rank	3			1.35	1.06	0.78	0.57
(r)	4				1.60	1.24	0.95
	5					1.89	1.46
	6						2.13

Table 1c. Critical Values for $\hat{c}$

Size = 10%		Number of Variables (p)					
		1	2	3	4	5	6
	1	0.12	0.098	0.082	0.069	0.059	0.051
Cointegrating	2		0.21	0.18	0.15	0.12	0.11
rank	3			0.30	0.25	0.21	0.18
(r)	4				0.38	0.32	0.27
	5					0.46	0.39
	6						0.54

Size = 5%		Number of Variables (p)					
		1	2	3	4	5	6
	1	0.15	0.12	0.10	0.085	0.072	0.062
Cointegrating	2		0.25	0.21	0.17	0.15	0.12
rank	3			0.34	0.28	0.24	0.20
(r)	4				0.42	0.36	0.30
	5					0.50	0.43
	6						0.59

Size = 1%		Number of Variables (p)					
		1	2	3	4	5	6
	1	0.22	0.18	0.15	0.13	0.11	0.092
Cointegrating	2		0.32	0.28	0.23	0.20	0.17
rank	3			0.43	0.37	0.31	0.26
(r)	4				0.52	0.45	0.38
	5					0.61	0.52
	6						0.70

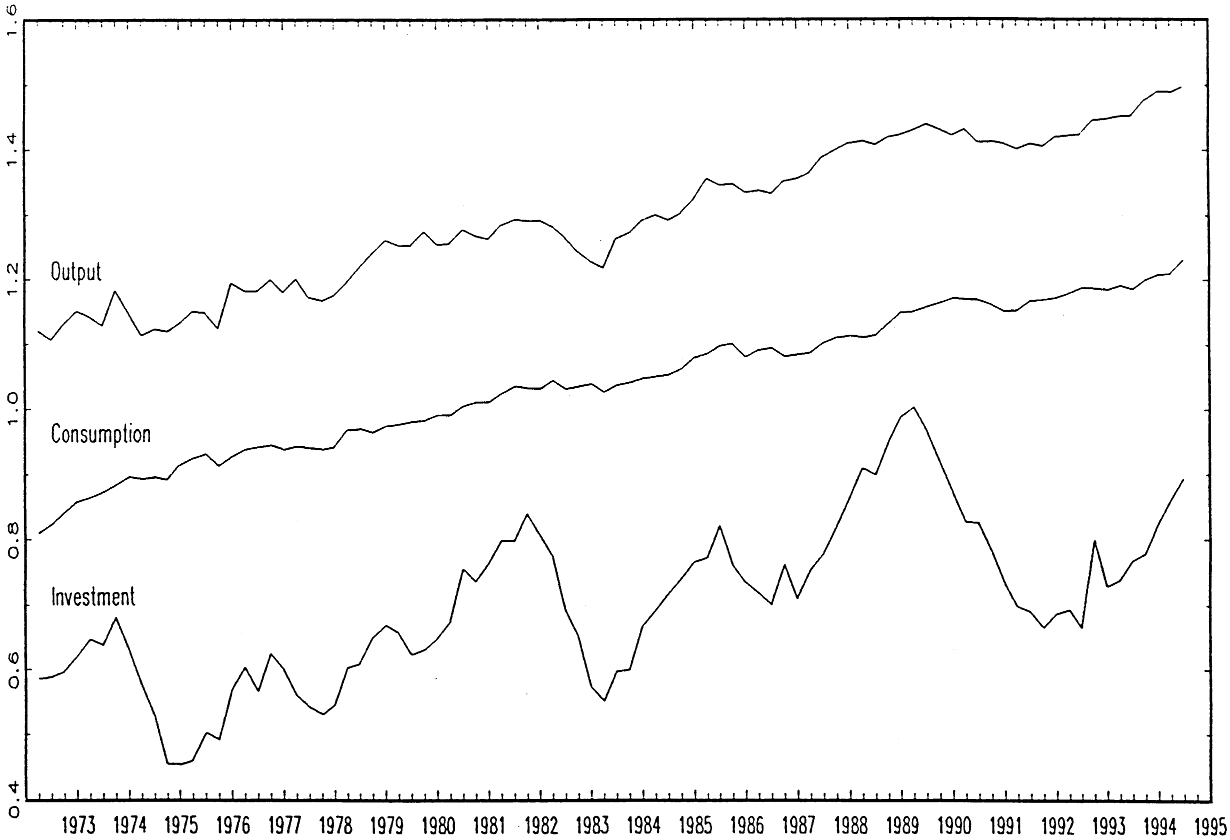

Figure 1. Logs of real private per capita Consumption, Investment and Output

Notes:

1. The data is obtained from the Australian dX database. All series are seasonally adjusted and measured in constant 1989/90 A$.

2. Private output is the difference between Gross Domestic Product (dX identifier: NPDQ.AK90GDP#E) and Government Expenditure, which is constructed as the sum of Government Final Consumption Expenditure (NADQ.AC#GG#99FCE), Public Enterprise Gross Fixed Capital Expenditure (NADQ.AC#GE#99GFC) and General Government Gross Fixed Capital Expenditure (NADQ.AC#GG#99GFC) each deflated by their respective Implicit Price Deflators (NPDQ.AD90GGC#, NPDQ.AD90PEK#, NPDQ.AD90GGK).

3. Real Private Consumption (NPDQ.AK90PRC#) and Private Gross Fixed Capital Expenditure (NPDQ.AK90PRK#) are used for consumption and investment respectively.

4. The Estimated Resident Population of Australia (DCRQ.UN71ERPAUSPER) is used to construct per capita series.

5. In Figure 1 only, 0.9 has been added to investment for readability.

6 Proofs

PROOF OF LEMMA 1. We will make use of the following intermediate results:

$$\beta' S_{yy}\beta \xrightarrow{p} \Sigma_{zz} = E\left(z_t z_t'\right), \tag{35}$$

$$T^{-1}\beta_\perp' S_{yy}\beta_\perp \xrightarrow{d} \int B_w B_w', \tag{36}$$

$$\beta_\perp' S_{yy}\beta \xrightarrow{d} \int B_w dB_z' + \Delta_{wz}, \tag{37}$$

where (36) and (37) follow from Park and Phillips (1988). We first demonstrate the behaviour of the eigenvalues satisfying (9). Pre-multiplying (9) by $|(\beta, T^{-1/2}\beta_\perp)'|$ and post-multiplying by $|(\beta, T^{-1/2}\beta_\perp)|$ gives

$$\left|\lambda\begin{pmatrix}\beta'\beta & 0\\ 0 & T^{-1}\beta_\perp'\beta_\perp\end{pmatrix} - \begin{pmatrix}\beta' S_{yy}\beta & T^{-1/2}\beta' S_{yy}\beta_\perp\\ T^{-1/2}\beta_\perp' S_{yy}\beta & T^{-1}\beta_\perp' S_{yy}\beta_\perp\end{pmatrix}\right| = 0,$$

which, using (35)–(37), shows that the r smallest eigenvalues converge to constants satisfying $|\lambda\beta'\beta - \Sigma_{zz}| = 0$, while the remaining $p-r$ eigenvalues diverge at rate $O_p(T)$. The estimator $\hat{\beta}$ satisfies

$$S_{yy}\hat{\beta} = \hat{\beta}\hat{\Lambda}. \tag{38}$$

We decompose $\hat{\beta}$ in the directions of β and $\beta_\perp$ as

$$\hat{\beta} = \beta(\beta'\beta)^{-1}\beta'\hat{\beta} + \beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'\hat{\beta}, \tag{39}$$

and substituting this into (38), pre-multiplying by $\beta_\perp$ and re-arranging gives

$$\beta_\perp' S_{yy}\beta_\perp\left(\beta_\perp'\beta_\perp\right)^{-1}\beta_\perp'\hat{\beta} - \beta_\perp'\beta_\perp\left(\beta_\perp'\beta_\perp\right)^{-1}\beta_\perp'\hat{\beta}\hat{\Lambda} = -\beta_\perp' S_{yy}\beta\left(\beta'\beta\right)^{-1}\beta'\hat{\beta}.$$

On vectorising this expression we find

$$\operatorname{vec}\left[\left(\beta_\perp'\beta_\perp\right)^{-1}\beta_\perp'\hat{\beta}\right] = \left[\left(\hat{\Lambda}\otimes\beta_\perp'\beta_\perp\right) - \left(I_r\otimes\beta_\perp' S_{yy}\beta_\perp\right)\right]^{-1}\operatorname{vec}\left[\beta_\perp' S_{yy}\beta\left(\beta'\beta\right)^{-1}\beta'\hat{\beta}\right]. \tag{40}$$

Since $\beta_\perp' S_{yy}\beta_\perp = O_p(T)$, it follows from (40) that $\left(\beta_\perp'\beta_\perp\right)^{-1}\beta_\perp'\hat{\beta} = O_p\left(T^{-1}\right)$. On rearranging (39), we find $\hat{\beta}-\beta(\beta'\beta)^{-1}\beta'\hat{\beta} = \beta_\perp\left(\beta_\perp'\beta_\perp\right)^{-1}\beta_\perp'\hat{\beta}$, and hence $\hat{\beta}-\beta(\beta'\beta)^{-1}\beta'\hat{\beta} =$

$O_p\left(T^{-1}\right)$. To find the asymptotic distribution of $\hat{\beta}$ we write (40) in un-vectorised form, pre-multiply by $T\beta_\perp$ and post-multiply by $\left(\beta'\hat{\beta}\right)^{-1}\beta'\beta$ to obtain

$$T\left(\hat{\beta}\left(\beta'\hat{\beta}\right)^{-1}\beta'\beta-\beta\right)=-\beta_\perp\left(T^{-1}\beta_\perp'S_{yy}\beta_\perp\right)^{-1}\beta_\perp'S_{yy}\beta+O_p\left(T^{-1}\right). \tag{41}$$

Applying (35)–(37) gives (10) as required. We note that $\beta'\hat{\beta}$ is invertible because β has full column rank (r) and the random matrix $\hat{\beta}$ also has full column rank by the construction of the eigenvectors in the principal components estimator. With respect to $\hat{\beta}_\perp$, we consider any estimator such that $\hat{\beta}'\hat{\beta}_\perp=0$. We write

$$\hat{\beta}_\perp=\beta\left(\beta'\beta\right)^{-1}\beta'\hat{\beta}_\perp+\beta_\perp\left(\beta_\perp'\beta_\perp\right)^{-1}\beta_\perp'\hat{\beta}_\perp,$$

so that

$$\hat{\beta}'\beta\left(\beta'\beta\right)^{-1}\beta'\hat{\beta}_\perp=-\hat{\beta}'\beta_\perp\left(\beta_\perp'\beta_\perp\right)^{-1}\beta_\perp'\hat{\beta}_\perp=O_p\left(T^{-1}\right)$$

because of the consistency of $\hat{\beta}$. Thus $\beta\left(\beta'\beta\right)^{-1}\beta'\hat{\beta}_\perp=\left(I_{p-r}-\beta_\perp\left(\beta_\perp'\beta_\perp\right)^{-1}\beta_\perp'\right)\hat{\beta}_\perp=O_p\left(T^{-1}\right)$. ■

PROOF OF THEOREM 1. We consider first the behaviour of the terms $\beta_\perp'S_{yy}^*\beta_\perp$ and $\beta_\perp'S_{yy}^*\beta$. Since the difference between $\beta_\perp'y_t^*$ and $\beta_\perp'y_t$ is $I(0)$, we find

$$\begin{aligned} T^{-1}\beta_\perp'S_{yy}^*\beta_\perp &= T^{-1}\beta_\perp'S_{yy}\beta_\perp+o_p(1) \\ &\xrightarrow{d} \int B_wB_w'. \end{aligned}$$

We can write $\beta_\perp'S_{yy}^*\beta$ as

$$\begin{aligned} \beta_\perp'S_{yy}^*\beta &= T^{-1}\sum_{t=1}^{T}\beta_\perp'y_t\left(z_t-\beta'\hat{\beta}\left(\hat{\beta}'\hat{\beta}\right)^{-1}\hat{\Omega}_{zw}\hat{\Omega}_{ww}^{-1}\hat{w}_t\right)' \\ &\quad -\beta_\perp'\hat{\beta}_\perp\left(\hat{\beta}_\perp'\hat{\beta}_\perp\right)^{-1}\hat{\Delta}_{w\zeta}\hat{S}_{\zeta\zeta}T^{-1}\sum_{t=1}^{T}\hat{\zeta}_t\left(z_t-\beta'\hat{\beta}\left(\hat{\beta}'\hat{\beta}\right)^{-1}\hat{\Omega}_{zw}\hat{\Omega}_{ww}^{-1}\hat{w}_t\right)' \\ &\quad -\beta_\perp'\hat{\beta}\left(\hat{\beta}'\hat{\beta}\right)^{-1}\hat{\Omega}_{zw}\hat{\Omega}_{ww}^{-1}T^{-1}\sum_{t=1}^{T}\hat{w}_ty_t^{*\prime}\beta \\ &\quad -T^{-1}\sum_{t=1}^{T}\beta_\perp'y_t\hat{\zeta}'\hat{S}_{\zeta\zeta}^{-1}\hat{\Delta}_{\zeta w}\left(\hat{\beta}_\perp'\hat{\beta}_\perp\right)^{-1}\hat{\beta}_\perp'\beta \\ &\quad +\beta_\perp'\hat{\beta}_\perp\left(\hat{\beta}_\perp'\hat{\beta}_\perp\right)^{-1}\hat{\Delta}_{w\zeta}\hat{S}_{\zeta\zeta}^{-1}T^{-1}\sum_{t=1}^{T}\hat{\zeta}_t\hat{\zeta}_t'\hat{S}_{\zeta\zeta}^{-1}\hat{\Delta}_{\zeta w}\left(\hat{\beta}_\perp'\hat{\beta}_\perp\right)^{-1}\hat{\beta}_\perp'\beta. \end{aligned}$$

The last three terms are $O_p(T^{-1})$ because $\hat{\beta}'\beta_\perp$ and $\hat{\beta}_\perp'\beta$ are $O_p(T^{-1})$. In the first two terms, we can replace $\hat{\beta}$ by $\hat{\beta}\left(\beta'\hat{\beta}\right)^{-1}\beta'\beta$ (including terms involving $\hat{z}_t$) because y_t^* has been defined to be invariant to the normalisation of $\hat{\beta}$. This can in turn be replaced by $\beta + O_p(T^{-1})$ because of the consistency of $\hat{\beta}$. Similarly, we can replace $\hat{\beta}_\perp$ by $\beta_\perp + O_p(T^{-1})$ since y_t^* is invariant to the normalisation of $\hat{\beta}_\perp$ and $\hat{\beta}_\perp$ is consistent. Therefore, the first two terms above are respectively:

$$
\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\beta_\perp' y_t\left(z_t - \beta'\hat{\beta}\left(\hat{\beta}'\hat{\beta}\right)^{-1}\hat{\Omega}_{zw}\hat{\Omega}_{ww}^{-1}\hat{w}_t\right)' \\
= \; & T^{-1}\sum_{t=1}^{T}\beta_\perp' y_t\left(z_t - \Omega_{zw}\Omega_{ww}^{-1}w_t\right)' + o_p(1) \\
\xrightarrow{d} \; & \int B_w dB_z^{*\prime} + \left(\Delta_{wz} - \Delta_{ww}\Omega_{ww}^{-1}\Omega_{wz}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
& \beta_\perp'\hat{\beta}_\perp\left(\hat{\beta}_\perp'\hat{\beta}_\perp\right)^{-1}\hat{\Delta}_{w\zeta}\hat{S}_{\zeta\zeta}T^{-1}\sum_{t=1}^{T}\hat{\zeta}_t\left(z_t - \beta'\hat{\beta}\left(\hat{\beta}'\hat{\beta}\right)^{-1}\hat{\Omega}_{zw}\hat{\Omega}_{ww}^{-1}\hat{w}_t\right)' \\
= \; & \Delta_{w\zeta}S_{\zeta\zeta}^{-1}T^{-1}\sum_{t=1}^{T}\zeta_t z_t' - \Delta_{w\zeta}S_{\zeta\zeta}^{-1}T^{-1}\sum_{t=1}^{T}\zeta_t w_t'\Omega_{ww}^{-1}\Omega_{wz} + o_p(1) \\
= \; & \Delta_{w\zeta}S_{\zeta\zeta}^{-1}T^{-1}\sum_{t=1}^{T}\zeta_t\zeta_t'(I_r, 0)' - \Delta_{w\zeta}S_{\zeta\zeta}^{-1}T^{-1}\sum_{t=1}^{T}\zeta_t\zeta_t'(0, I_{p-r})'\Omega_{ww}^{-1}\Omega_{wz} + o_p(1) \\
= \; & \Delta_{wz} - \Delta_{ww}\Omega_{ww}^{-1}\Omega_{wz} + o_p(1),
\end{aligned}
$$

since $\Delta_{w\zeta} = (\Delta_{wz}, \Delta_{ww})$. Thus

$$\beta_\perp' S_{yy}^*\beta \xrightarrow{d} \int B_w dB_z^{*\prime}.$$

It is similarly straightforward to show that $\beta' S_{yy}^*\beta = O_p(1)$. As in the proof of Lemma 1, we can then show that the smallest r eigenvalues satisfying (15) converge to positive constants, while the largest $p - r$ eigenvalues diverge at $O_p(T)$, and hence we can represent $\hat{\beta}^*$ as

$$T\left(\hat{\beta}^*\left(\beta'\hat{\beta}^*\right)^{-1}\beta'\beta - \beta\right) = -\beta_\perp\left(T^{-1}\beta_\perp' S_{yy}^*\beta_\perp\right)^{-1}\beta_\perp' S_{yy}^*\beta + O_p\left(T^{-1}\right). \quad (42)$$

Applying the above results proves the theorem. ■

PROOF OF THEOREM 2. *(i)* We first consider the properties of $\bar{x}_t$. We can write

$$\bar{x}_t = y_t - T^{-1}\sum_{t=1}^{T} y_t = x_t - T^{-1}\sum_{t=1}^{T} x_t.$$

Thus

$$\beta'\bar{x}_t = z_t - T^{-1}\sum_{t=1}^{T} z_t,$$

and

$$T^{-1/2}\beta_\perp'\bar{x}_t = T^{-1/2}\beta_\perp' x_t - T^{-3/2}\sum_{t=1}^{T}\beta_\perp' x_t \xrightarrow{d} B_w(s) - \int B_w = \bar{B}_w(s).$$

Using these results we find

$$\beta'\bar{S}_{xx}\beta \xrightarrow{p} \Sigma_{zz} = E\left(z_t z_t'\right),$$

$$T^{-1}\beta_\perp'\bar{S}_{xx}\beta_\perp = T^{-2}\sum_{t=1}^{T}\beta_\perp'\bar{x}_t\bar{x}_t'\beta_\perp \xrightarrow{d} \int \bar{B}_w\bar{B}_w',$$

and

$$\begin{aligned}
\beta_\perp'\bar{S}_{xx}\beta &= T^{-1}\sum_{t=1}^{T}\beta_\perp'\bar{x}_t\bar{x}_t'\beta \\
&= T^{-1}\sum_{t=1}^{T}\beta_\perp' x_t z_t' - T^{-3/2}\sum_{t=1}^{T}\beta_\perp' x_t \cdot T^{-1/2}\sum_{t=1}^{T} z_t' \\
&\xrightarrow{d} \int B_w dB_z' + \Delta_{wz} - \int B_w \cdot B_z(1)' \\
&= \int \bar{B}_w dB_z' + \Delta_{wz}.
\end{aligned}$$

Following the same arguments as in the proof of Lemma 1, we find that the smallest r eigenvalues satisfying $|\lambda I_p - \bar{S}_{xx}| = 0$ converge to finite constants, while the largest $p - r$ eigenvalues diverge at rate $O_p(T)$. As in Lemma 1, we can write

$$\mathrm{vec}\left[\left(\beta_\perp'\beta_\perp\right)^{-1}\beta_\perp'\hat{\beta}\right] = \left[\left(\bar{\Lambda}\otimes\beta_\perp'\beta_\perp\right) - \left(I_r\otimes\beta_\perp'\bar{S}_{xx}\beta_\perp\right)\right]^{-1}\mathrm{vec}\left[\beta_\perp'\bar{S}_{xx}\beta\left(\beta'\beta\right)^{-1}\beta'\bar{\beta}\right],$$

and show consistency by noting that $\beta_\perp'\bar{S}_{xx}\beta_\perp = O_p(T)$, and then

$$\begin{aligned}
T\left(\bar{\beta}\left(\beta'\bar{\beta}\right)^{-1}\beta'\beta - \beta\right) &= -\beta_\perp\left(T^{-1}\beta_\perp'\bar{S}_{xx}\beta_\perp\right)^{-1}\beta_\perp'\bar{S}_{xx}\beta + O_p\left(T^{-1}\right) \\
&\xrightarrow{d} -\beta_\perp\left(\int \bar{B}_w\bar{B}_w'\right)^{-1}\left(\int \bar{B}_w dB_z' + \Delta_{wz}\right),
\end{aligned}$$

to obtain the asymptotic distribution as required.

We now consider the modified estimator $\bar{\beta}^*$ based on $\bar{x}_t^*$ defined in (23). The method is similar to the proof of Theorem 1. First

$$T^{-1}\beta_\perp' \bar{S}_{xx}^* \beta_\perp = T^{-1}\beta_\perp' \bar{S}_{xx}\beta_\perp + O_p\left(T^{-1}\right) \xrightarrow{d} \int \bar{B}_w \bar{B}_w',$$

since $\bar{x}_t^*$ and $\bar{x}_t$ differ only by $I(0)$ terms. To find the behaviour of $\beta_\perp' \bar{S}_{xx}^* \beta$, note that we can write

$$\begin{aligned}
\beta_\perp' \bar{x}_t^* &= \beta_\perp' \bar{x}_t - \Delta_{w\zeta} S_{\zeta\zeta}^{*-1} \zeta_t^* + o_p(1), \\
\beta' \bar{x}_t^* &= z_t - T^{-1}\sum_{t=1}^T z_t - \Omega_{zw}\Omega_{ww}^{-1} w_t + o_p(1),
\end{aligned}$$

where $\zeta_t^* = \left(z_t' - T^{-1}\sum_{t=1}^T z_t', w_t'\right)'$ and $S_{\zeta\zeta}^* = T^{-1}\sum_{t=1}^T \zeta_t^* \zeta_t^{*\prime}$, since $\beta'\bar{x}_t = z_t - T^{-1}\sum_{t=1}^T z_t$ and $\beta_\perp' \Delta \bar{x}_t = \beta_\perp' \Delta x_t = w_t$. Then

$$\beta_\perp' \bar{S}_{xx}^* \beta = T^{-1}\sum_{t=1}^T \beta_\perp' \bar{x}_t \bar{x}_t^{*\prime}\beta - \Delta_{w\zeta} S_{\zeta\zeta}^{*-1} T^{-1}\sum_{t=1}^T \zeta_t^* \bar{x}_t^{*\prime}\beta + o_p(1).$$

The first term is

$$\begin{aligned}
& T^{-1}\sum_{t=1}^T \left(\beta_\perp' x_t - T^{-1}\sum_{t=1}^T \beta_\perp' x_t\right)\left(z_t - T^{-1}\sum_{t=1}^T z_t - \Omega_{zw}\Omega_{ww}^{-1} w_t\right)' \\
= \; & T^{-1}\sum_{t=1}^T \beta_\perp' x_t \left(z_t - \Omega_{zw}\Omega_{ww}^{-1} w_t\right)' - T^{-3/2}\sum_{t=1}^T \beta_\perp' x_t \cdot T^{-1/2}\sum_{t=1}^T \left(z_t - \Omega_{zw}\Omega_{ww}^{-1} w_t\right)' \\
\xrightarrow{d} \; & \int B_w dB_z^{*\prime} + \left(\Delta_{wz} - \Delta_{ww}\Omega_{ww}^{-1}\Omega_{wz}\right) - \int B_w \cdot B_z^*(1) \\
= \; & \int \bar{B}_w dB_z^{*\prime} + \left(\Delta_{wz} - \Delta_{ww}\Omega_{ww}^{-1}\Omega_{wz}\right),
\end{aligned}$$

and the second is

$$\begin{aligned}
& \Delta_{w\zeta} S_{\zeta\zeta}^{*-1} T^{-1}\sum_{t=1}^T \zeta_t^* \left(z_t - T^{-1}\sum_{t=1}^T z_t\right)' - \Delta_{w\zeta} S_{\zeta\zeta}^{*-1} T^{-1}\sum_{t=1}^T \zeta_t^* w_t' \Omega_{ww}^{-1}\Omega_{wz} \\
= \; & \Delta_{w\zeta} S_{\zeta\zeta}^{*-1} T^{-1}\sum_{t=1}^T \zeta_t^* \zeta_t^{*\prime}(I_r, 0) - \Delta_{w\zeta} S_{\zeta\zeta}^{*-1} T^{-1}\sum_{t=1}^T \zeta_t^* \zeta_t^{*\prime}(0, I_{p-r})\,\Omega_{ww}^{-1}\Omega_{wz} \\
= \; & \Delta_{wz} - \Delta_{ww}\Omega_{ww}^{-1}\Omega_{wz}.
\end{aligned}$$

Thus

$$\beta_\perp' \bar{S}_{xx}^* \beta \xrightarrow{d} \int \bar{B}_w dB_z^{*\prime}.$$

It then follows we can represent $\bar{\beta}^*$ as

$$\begin{aligned} T\left(\bar{\beta}^*\left(\beta'\bar{\beta}^*\right)^{-1}\beta'\beta-\beta\right) &= -\beta_\perp\left(T^{-1}\beta_\perp'\bar{S}_{xx}^*\beta_\perp\right)^{-1}\beta_\perp'\bar{S}_{xx}^*\beta+O_p\left(T^{-1}\right) \\ &\xrightarrow{d} -\beta_\perp\left(\int\bar{B}_w\bar{B}_w'\right)^{-1}\int\bar{B}_w dB_z^{*\prime}, \end{aligned}$$

as required.

(ii) We first consider $\tilde{x}_t$, which is the set of residuals

$$\begin{aligned} \tilde{x}_t &= y_t-\tilde{\mu}-\tilde{\delta}t \\ &= x_t-(\tilde{\mu}-\mu)-\left(\tilde{\delta}-\delta\right)t, \end{aligned}$$

where $\tilde{\mu}$ and $\tilde{\delta}$ are the OLS estimators of the de-trending regression:

$$\begin{aligned} \begin{pmatrix} \tilde{\mu}-\mu \\ \tilde{\delta}-\delta \end{pmatrix} &= \begin{pmatrix} T & \sum_{t=1}^T t \\ \sum_{t=1}^T t & \sum_{t=1}^T t^2 \end{pmatrix}^{-1}\begin{pmatrix} \sum_{t=1}^T x_t \\ \sum_{t=1}^T tx_t \end{pmatrix} \\ &= \frac{1}{T(T-1)}\begin{pmatrix} (4T+2)\sum_{t=1}^T x_t-6\sum_{t=1}^T tx_t \\ -6\sum_{t=1}^T x_t+12(T+1)^{-1}\sum_{t=1}^T tx_t \end{pmatrix}. \end{aligned}$$

We then have the following intermediate results:

$$\begin{aligned} T^{-1/2}\beta_\perp'(\tilde{\mu}-\mu) &= \frac{4T+2}{T-1}T^{-3/2}\sum_{t=1}^T\left(\sum_{j=1}^t w_j\right)-\frac{6T}{T-1}T^{-5/2}\sum_{t=1}^T t\left(\sum_{j=1}^t w_j\right) \\ &\xrightarrow{d} 4\int B_w-6\int rB_w, \\ T^{1/2}\beta'(\tilde{\mu}-\mu) &= \frac{4T+2}{T-1}T^{-1/2}\sum_{t=1}^T z_t-\frac{6T}{T-1}T^{-3/2}\sum_{t=1}^T tz_t \\ &\xrightarrow{d} 4B_z(1)-6\int rdB_z \\ &= -2B_z(1)+6\int B_z, \\ T^{1/2}\beta_\perp'\left(\tilde{\delta}-\delta\right) &= \frac{-6T}{T-1}T^{-3/2}\sum_{t=1}^T\left(\sum_{j=1}^t w_j\right)+\frac{12T^2}{(T-1)(T+1)}T^{-5/2}\sum_{t=1}^T t\left(\sum_{j=1}^t w_j\right) \\ &\xrightarrow{d} -6\int B_w+12\int rB_w, \\ T^{3/2}\beta'\left(\tilde{\delta}-\delta\right) &= \frac{-6T}{T-1}T^{-1/2}\sum_{t=1}^T z_t+\frac{12T^2}{(T-1)(T+1)}T^{-3/2}\sum_{t=1}^T tz_t \\ &\xrightarrow{d} -6B_z(1)+12\int rdB_z \\ &= 6B_z(1)-12\int B_z. \end{aligned}$$

Then

$$
\begin{aligned}
T^{-1/2}\beta_{\perp}'\tilde{x}_t &= T^{-1/2}\beta_{\perp}'x_t - T^{-1/2}\beta_{\perp}'(\tilde{\mu}-\mu) - T^{1/2}\beta_{\perp}'\left(\tilde{\delta}-\delta\right)\frac{t}{T} \\
&\xrightarrow{d} B_w(s) - \left(4\int B_w - 6\int rB_w\right) - s\left(-6\int B_w + 12\int rB_w\right) \\
&= B_w(s) + 2(3s-2)\int B_w - 6(2s-1)\int rB_w \\
&= \tilde{B}_w(s),
\end{aligned}
$$

so that

$$
T^{-1}\beta_{\perp}'\tilde{S}_{xx}\beta_{\perp} = T^{-2}\sum_{t=1}^{T}\beta_{\perp}'\tilde{x}_t\tilde{x}_t'\beta_{\perp} \xrightarrow{d} \int \tilde{B}_w\tilde{B}_w'.
$$

We now consider $\beta_{\perp}'\tilde{S}_{xx}\beta$:

$$
\begin{aligned}
\beta_{\perp}'\tilde{S}_{xx}\beta &= T^{-1}\sum_{t=1}^{T}\left(\sum_{j=1}^{t}w_j\right)z_t' - T^{-1}\sum_{t=1}^{T}\left(\sum_{j=1}^{t}w_j\right)(\tilde{\mu}-\mu)'\beta - \beta_{\perp}'(\tilde{\mu}-\mu)T^{-1}\sum_{t=1}^{T}z_t \\
&\quad -T^{-1}\sum_{t=1}^{T}t\left(\sum_{j=1}^{t}w_j\right)\left(\tilde{\delta}-\delta\right)'\beta + \beta_{\perp}'(\tilde{\mu}-\mu)(\tilde{\mu}-\mu)'\beta \\
&\quad -\beta_{\perp}'\left(\tilde{\delta}-\delta\right)T^{-1}\sum_{t=1}^{T}tz_t + \beta_{\perp}'\left(\tilde{\delta}-\delta\right)\left(\tilde{\delta}-\delta\right)'\beta T^{-1}\sum_{t=1}^{T}t^2 \\
&\quad +\beta_{\perp}'\left(\tilde{\delta}-\delta\right)(\tilde{\mu}-\mu)'\beta T^{-1}\sum_{t=1}^{T}t + \beta_{\perp}'(\tilde{\mu}-\mu)\left(\tilde{\delta}-\delta\right)'\beta T^{-1}\sum_{t=1}^{T}t.
\end{aligned}
$$

These terms have the following asymptotic behaviour:

$$
\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\left(\sum_{j=1}^{t}w_j\right)z_t' \xrightarrow{d} \int B_w dB_z' + \Delta_{wz}, \\
& T^{-1}\sum_{t=1}^{T}\left(\sum_{j=1}^{t}w_j\right)(\tilde{\mu}-\mu)'\beta \xrightarrow{d} \int B_w \cdot \left(-2B_z(1) + 6\int B_z\right)', \\
& \beta_{\perp}'(\tilde{\mu}-\mu)T^{-1}\sum_{t=1}^{T}z_t \xrightarrow{d} \left(4\int B_w - 6\int rB_w\right)\cdot B_z(1)', \\
& T^{-1}\sum_{t=1}^{T}t\left(\sum_{j=1}^{t}w_j\right)\left(\tilde{\delta}-\delta\right)'\beta \xrightarrow{d} \int rB_w \cdot \left(6B_z(1) - 12\int B_z\right)', \\
& \beta_{\perp}'(\tilde{\mu}-\mu)(\tilde{\mu}-\mu)'\beta \xrightarrow{d} \left(4\int B_w - 6\int rB_w\right)\left(-2B_z(1) + 6\int B_z\right)' \\
& \beta_{\perp}'\left(\tilde{\delta}-\delta\right)T^{-1}\sum_{t=1}^{T}tz_t \xrightarrow{d} \left(-6\int B_w + 12\int rB_w\right)\left(B_z(1) - \int B_z\right)',
\end{aligned}
$$

$$\beta'_{\perp}\left(\tilde{\delta}-\delta\right)(\tilde{\mu}-\mu)'\beta T^{-1}\sum_{t=1}^{T}t \xrightarrow{d} \frac{1}{2}\left(-6\int B_w + 12\int rB_w\right)\left(-2B_z(1)+6\int B_z\right)'$$

$$\beta'_{\perp}(\tilde{\mu}-\mu)\left(\tilde{\delta}-\delta\right)'\beta T^{-1}\sum_{t=1}^{T}t \xrightarrow{d} \frac{1}{2}\left(4\int B_w - 6\int rB_w\right)\left(6B_z(1)-12\int rdB_z\right)'$$

$$\beta'_{\perp}\left(\tilde{\delta}-\delta\right)\left(\tilde{\delta}-\delta\right)'\beta T^{-1}\sum_{t=1}^{T}t^2 \xrightarrow{d}$$

$$\frac{1}{3}\left(-6\int B_w + 12\int rB_w\right)\left(6B_z(1)-12\int rdB_z\right)'.$$

Combining these results gives

$$\begin{aligned}\beta'_{\perp}\tilde{S}_{xx}\beta \xrightarrow{d} & \int B_w dB'_z + 2\int B_w \cdot B_z(1)' - 6\int rB_w \cdot B_z(1)' \\ & +12\int rB_w \int B'_z - 6\int B_w \int B'_z + \Delta_{wz},\end{aligned}$$

which is simply a rearrangement of

$$\beta'_{\perp}\tilde{S}_{xx}\beta \xrightarrow{d} \int \tilde{B}_w dB'_z + \Delta_{wz}.$$

We again apply the arguments of the proof of Lemma 1 to show that the smallest r eigenvalues satisfying $\left|\lambda I_p - \tilde{S}_{xx}\right| = 0$ converge to positive constants, while the largest $p-r$ eigenvalues diverge at rate $O_p(T)$. It then follows that $\tilde{\beta}$ can be given the representation

$$T\left(\tilde{\beta}\left(\beta'\tilde{\beta}\right)^{-1}\beta'\beta-\beta\right) = -\beta_{\perp}\left(T^{-1}\beta'_{\perp}\tilde{S}_{xx}\beta_{\perp}\right)^{-1}\beta'_{\perp}\tilde{S}_{xx}\beta + O_p\left(T^{-1}\right),$$

which gives the distribution

$$T\left(\tilde{\beta}\left(\beta'\tilde{\beta}\right)^{-1}\beta'\beta-\beta\right) \xrightarrow{d} -\beta_{\perp}\left(\int \tilde{B}_w\tilde{B}'_w\right)^{-1}\left(\int \tilde{B}_w dB'_z + \Delta_{wz}\right),$$

as required.

To show the final result of the theorem, first note that

$$\begin{aligned}\beta'_{\perp}\Delta\tilde{x}_t &= \beta'_{\perp}\Delta x_t - \beta'_{\perp}\left(\tilde{\delta}-\delta\right) = w_t + O_p\left(T^{-1/2}\right), \\ \beta'\tilde{x}_t &= \beta' x_t - \beta'(\tilde{\mu}-\mu) - \beta'\left(\tilde{\delta}-\delta\right)t = z_t + O_p\left(T^{-1/2}\right).\end{aligned}$$

It then follows that

$$\begin{aligned}\beta'_{\perp}\tilde{x}^*_t &= \beta'_{\perp}\tilde{x}_t - \Delta_{w\zeta}S^{-1}_{\zeta\zeta}\zeta_t + o_p(1), \\ \beta'\tilde{x}^*_t &= z_t - \Omega_{zw}\Omega^{-1}_{ww}w_t + o_p(1),\end{aligned}$$

and so

$$\beta_\perp' \tilde{S}_{xx}^* \beta_\perp = \beta_\perp' \tilde{S}_{xx} \beta_\perp + O_p\left(T^{-1}\right) \xrightarrow{d} \int \tilde{B}_w \tilde{B}_w',$$

and

$$\begin{aligned}
\beta_\perp' \tilde{S}_{xz}^* \beta &= T^{-1} \sum_{t=1}^{T} \beta_\perp' \tilde{x}_t \left(z_t - \Omega_{zw}\Omega_{ww}^{-1} w_t\right)' - \Delta_{w\zeta} S_{\zeta\zeta}^{-1} T^{-1} \sum_{t=1}^{T} \zeta_t \left(z_t - \Omega_{zw}\Omega_{ww}^{-1} w_t\right)' + o_p(1) \\
&\xrightarrow{d} \int \tilde{B}_w dB_z^{*\prime} + \left(\Delta_{wz} - \Delta_{ww}\Omega_{ww}^{-1}\Omega_{wz}\right) - \left(\Delta_{wz} - \Delta_{ww}\Omega_{ww}^{-1}\Omega_{wz}\right) \\
&= \int \tilde{B}_w dB_z^{*\prime}.
\end{aligned}$$

The estimator $\tilde{\beta}^*$ can then be given the representation

$$T\left(\tilde{\beta}^* \left(\beta'\tilde{\beta}^*\right)^{-1} \beta'\beta - \beta\right) = -\beta_\perp \left(T^{-1}\beta_\perp' \tilde{S}_{xx}^* \beta_\perp\right)^{-1} \beta_\perp' \tilde{S}_{xz}^* \beta + O_p\left(T^{-1}\right),$$

which gives the distribution

$$T\left(\tilde{\beta}^* \left(\beta'\tilde{\beta}^*\right)^{-1} \beta'\beta - \beta\right) \xrightarrow{d} -\beta_\perp \left(\int \tilde{B}_w \tilde{B}_w'\right)^{-1} \int \tilde{B}_w dB_z^{*\prime},$$

as required. ■

Proof of Theorem 3. (i) Under H_0 we have $J_1'\beta = 0$, so we can write

$$\begin{aligned}
J_1'\hat{\beta}^* &= J_1' \left(\hat{\beta}^* - \beta\left(\beta'\beta\right)^{-1} \beta'\hat{\beta}^*\right) \\
&= J_1' \left(\hat{\beta}^* \left(\beta'\hat{\beta}^*\right)^{-1} \beta'\beta - \beta\right) \left(\beta'\beta\right)^{-1} \beta'\hat{\beta}^*,
\end{aligned}$$

and hence

$$T J_1'\hat{\beta}^* \left(\beta'\hat{\beta}^*\right)^{-1} \beta'\beta \xrightarrow{d} -J_1'\beta_\perp \left(\int B_w B_w'\right)^{-1} \int B_w dB_z^{*\prime},$$

by Theorem 1. We can also write

$$S_{yy}^{-1} = \begin{pmatrix} \beta & T^{-1/2}\beta_\perp \end{pmatrix} \begin{pmatrix} \beta' S_{yy} \beta & T^{-1/2}\beta' S_{yy} \beta_\perp \\ T^{-1/2}\beta_\perp' S_{yy} \beta & T^{-1}\beta_\perp' S_{yy} \beta_\perp \end{pmatrix}^{-1} \begin{pmatrix} \beta' \\ T^{-1/2}\beta_\perp' \end{pmatrix},$$

and

$$J_1' S_{yy}^{-1} J_1 = \begin{pmatrix} 0 & T^{-1/2} J_1'\beta_\perp \end{pmatrix} \begin{pmatrix} \beta' S_{yy} \beta & T^{-1/2}\beta' S_{yy} \beta_\perp \\ T^{-1/2}\beta_\perp' S_{yy} \beta & T^{-1}\beta_\perp' S_{yy} \beta_\perp \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ T^{-1/2}\beta_\perp' J_1 \end{pmatrix},$$

under H_0, so

$$TJ_1'S_{yy}^{-1}J_1 \xrightarrow{d} J_1'\beta_\perp\left(\int B_w B_w'\right)^{-1}\beta_\perp' J_1.$$

The long run covariance matrix estimator $\hat{\Omega}_{zz}^*$ satisfies

$$\beta'\beta\left(\hat{\beta}^{*\prime}\beta\right)^{-1}\hat{\Omega}_{zz}^*\left(\beta'\hat{\beta}^*\right)^{-1}\beta'\beta \xrightarrow{p} \Omega_{zz}^* = \Omega_{zz}-\Omega_{zw}\Omega_{ww}^{-1}\Omega_{wz},$$

since $\hat{\beta}^*\left(\beta'\hat{\beta}^*\right)^{-1}\beta'\beta \xrightarrow{p} \beta$. Note that Ω_{zz}^* is the variance of B_z^*. It then follows that

$$\begin{aligned}
\hat{s}_1 &= T\mathrm{tr}\hat{\Omega}_{zz}^{*-1}\hat{\beta}^{*\prime}J_1\left(J_1'S_{yy}^{-1}J_1\right)^{-1}J_1'\hat{\beta} \\
&\xrightarrow{d} \mathrm{tr}\Omega_{zz}^{*-1}\int dB_z^*B_w'\left(\int B_wB_w'\right)^{-1}\beta_\perp'J_1\left[J_1'\beta_\perp\left(\int B_wB_w'\right)^{-1}\beta_\perp'J_1\right]^{-1}\times \\
&\qquad\qquad J_1\beta_\perp\left(\int B_wB_w'\right)^{-1}\int B_w dB_z^{*\prime} \\
&= (\mathrm{vec}\, Z)'\,\mathrm{vec}\, Z,
\end{aligned}$$

where

$$Z=\left[J_1'\beta_\perp\left(\int B_wB_w'\right)^{-1}\beta_\perp'J_1\right]^{-1/2}J_1\beta_\perp\left(\int B_wB_w'\right)^{-1}\int B_w dV',$$

V is an r dimensional standard Brownian motion independent of B_w, and thus Z is a $(p-s)\times r$ random variable with the distribution

$$\mathrm{vec}\, Z \sim N\left(0, I_{p-s}\otimes I_r\right).$$

Hence $\hat{s}_1$ is asymptotically $\chi^2_{r(p-s)}$ under the null.

Under the alternative, $\beta' J_1 \neq 0$. From the consistency of $\hat{\beta}^*$ we have

$$\hat{\beta}^* = \beta\left(\beta'\beta\right)^{-1}\beta'\hat{\beta}^* + O_p\left(T^{-1}\right),$$

and hence

$$\hat{\beta}^{*\prime}J_1 = \hat{\beta}^{*\prime}\beta\left(\beta'\beta\right)^{-1}\beta'J_1 + O_p\left(T^{-1}\right).$$

That is, $\hat{\beta}^{*\prime}J_1$ is $O_p(1)$ under the alternative. We also have

$$\begin{aligned}
J_1'S_{yy}^{-1}J_1 &= \begin{pmatrix} J_1'\beta & T^{-\frac{1}{2}}J_1'\beta_\perp \end{pmatrix}\begin{pmatrix} \beta'S_{yy}\beta & T^{-1/2}\beta'S_{yy}\beta_\perp \\ T^{-1/2}\beta_\perp'S_{yy}\beta & T^{-1}\beta_\perp'S_{yy}\beta_\perp \end{pmatrix}^{-1}\begin{pmatrix} \beta'J_1 \\ T^{-\frac{1}{2}}\beta_\perp'J_1 \end{pmatrix}, \\
&= J_1'\beta\left(\beta'S_{yy}\beta\right)^{-1}\beta'J_1 + o_p(1).
\end{aligned}$$

Therefore each term inside the trace is $O_p(1)$ under the alternative, and hence $\hat{s}_1$ is $O_p(T)$.

Parts (ii) and (iii) follow in exactly the same way as part (i). ∎

PROOF OF THEOREM 4. (i) To prove this result we will show that

$$T\beta_\perp'\beta_\perp\left(\hat{\beta}_\perp^{*\prime}\beta_\perp\right)^{-1}\hat{\beta}_\perp^{*\prime}S_{yy}^{-1}\hat{\beta}_\perp^{*}\left(\beta_\perp'\hat{\beta}_\perp^{*}\right)^{-1}\beta_\perp'\beta_\perp \xrightarrow{d} \beta_\perp'\beta_\perp\left(\int B_w B_w'\right)^{-1}\beta_\perp'\beta_\perp, \tag{43}$$

$$T\beta_\perp'\beta_\perp\left(\hat{\beta}_\perp^{*\prime}\beta_\perp\right)^{-1}\hat{\beta}_\perp^{*\prime}H_2 \xrightarrow{d} -\beta_\perp'\beta_\perp\left(\int B_w B_w'\right)^{-1}\int B_w dB_H', \tag{44}$$

$$\hat{\Omega}_{HH} \xrightarrow{p} \Omega_{HH}, \tag{45}$$

where

$$B_H = H_2'\beta\left(\beta'\beta\right)^{-1}B_z^{*},$$

is an r_1 dimensional Brownian motion with covariance matrix

$$\Omega_{HH} = H_2'\beta\left(\beta'\beta\right)^{-1}\Omega_{zz}^{*}\left(\beta'\beta\right)^{-1}\beta'H_2.$$

We note that the test statistic $\hat{s}_2$ is invariant to the replacement of $\hat{\beta}_\perp^{*}$ by $\hat{\beta}_\perp^{*}\left(\beta_\perp'\hat{\beta}_\perp^{*}\right)\beta_\perp'\beta_\perp$, so we can then apply (43)–(45) to $\hat{s}_2$ to find

$$\begin{aligned}\hat{s}_2 &\xrightarrow{d} \operatorname{tr}\int dUW'\left(\int WW'\right)^{-1}\int WdU', \\ &= \left(\operatorname{vec}Z\right)'\left(\operatorname{vec}Z\right),\end{aligned}$$

where U and W are independent r_1 and $p-r$ dimensional standard Brownian motions, and

$$Z = \left(\int WW'\right)^{-1/2}\int WdU'$$

is a $(p-r)\times r_1$ dimensional random variable with distribution

$$\operatorname{vec}Z \sim N\left(0, I_{p-r}\otimes I_{r_1}\right).$$

Thus $\hat{s}_2$ has an asymptotic $\chi^2_{(p-r)r_1}$ distribution under the null.

To show (44), we note that $\hat{\beta}_\perp^{*}$ satisfies

$$\hat{\beta}_\perp^{*\prime}\hat{\beta}^{*} = 0.$$

We can decompose $\hat{\beta}_\perp^*$ as

$$\hat{\beta}_\perp^* = \beta\left(\beta'\beta\right)^{-1}\beta'\hat{\beta}_\perp^* + \beta_\perp\left(\beta_\perp'\beta_\perp\right)^{-1}\beta_\perp'\hat{\beta}_\perp^*,$$

so multiplying by $\hat{\beta}^{*\prime}$ gives

$$\hat{\beta}^{*\prime}\beta\left(\beta'\beta\right)^{-1}\beta'\hat{\beta}_\perp^* + \hat{\beta}^{*\prime}\beta_\perp\left(\beta_\perp'\beta_\perp\right)^{-1}\beta_\perp'\hat{\beta}_\perp^* = 0,$$

and

$$\begin{aligned}
\left(\beta'\beta\right)^{-1}\beta'\hat{\beta}_\perp^* &= -\left(\hat{\beta}^{*\prime}\beta\right)^{-1}\hat{\beta}^{*\prime}\beta_\perp\left(\beta_\perp'\beta_\perp\right)^{-1}\beta_\perp'\hat{\beta}_\perp^* \\
&= -\left(\hat{\beta}^{*\prime}\beta\right)^{-1}\left(\hat{\beta}^* - \beta\left(\beta'\beta\right)^{-1}\beta'\hat{\beta}^*\right)'\hat{\beta}_\perp^* \\
&= -\left(\beta'\beta\right)^{-1}\left(\hat{\beta}^*\left(\beta'\hat{\beta}^*\right)^{-1}\beta'\beta - \beta\right)'\hat{\beta}_\perp^*.
\end{aligned}$$

Since $\hat{\beta}_\perp^* - \beta_\perp\left(\beta_\perp'\beta_\perp\right)^{-1}\beta_\perp'\hat{\beta}_\perp^* = \beta\left(\beta'\beta\right)^{-1}\beta'\hat{\beta}_\perp^*$ we can write

$$\hat{\beta}_\perp^{*\prime} - \hat{\beta}_\perp^{*\prime}\beta_\perp\left(\beta_\perp'\beta_\perp\right)^{-1}\beta_\perp' = -\hat{\beta}_\perp^{*\prime}\left(\hat{\beta}^*\left(\beta'\hat{\beta}^*\right)^{-1}\beta'\beta - \beta\right)\left(\beta'\beta\right)^{-1}\beta',$$

and hence

$$\begin{aligned}
\beta_\perp'\beta_\perp\left(\hat{\beta}_\perp^{*\prime}\beta_\perp\right)^{-1}\hat{\beta}_\perp^{*\prime} - \beta_\perp' &= \beta_\perp'\beta_\perp\left(\hat{\beta}_\perp^{*\prime}\beta_\perp\right)^{-1}\hat{\beta}_\perp^{*\prime}\left(\hat{\beta}^*\left(\beta'\hat{\beta}^*\right)^{-1}\beta'\beta - \beta\right)\left(\beta'\beta\right)^{-1}\beta' \\
&= \beta_\perp'\left(\hat{\beta}^*\left(\beta'\hat{\beta}^*\right)^{-1}\beta'\beta - \beta\right)\left(\beta'\beta\right)^{-1}\beta' + O_p\left(T^{-1}\right).
\end{aligned}$$

Under the null hypothesis, $\beta_\perp'H_2 = 0$, so

$$\beta_\perp'\beta_\perp\left(\hat{\beta}_\perp^{*\prime}\beta_\perp\right)^{-1}\hat{\beta}_\perp^{*\prime}H_2 = \beta_\perp'\left(\hat{\beta}^*\left(\beta'\hat{\beta}^*\right)^{-1}\beta'\beta - \beta\right)\left(\beta'\beta\right)^{-1}\beta'H_2 + O_p\left(T^{-1}\right),$$

and using the asymptotic distribution of $\hat{\beta}^*$ from Theorem 1 gives (44).

To show (43) we write

$$S_{yy}^{-1} = \begin{pmatrix} \beta & T^{-1/2}\beta_\perp \end{pmatrix}\begin{pmatrix} \beta'S_{yy}\beta & T^{-1/2}\beta'S_{yy}\beta_\perp \\ T^{-1/2}\beta_\perp'S_{yy}\beta & T^{-1}\beta_\perp'S_{yy}\beta_\perp \end{pmatrix}^{-1}\begin{pmatrix} \beta' \\ T^{-1/2}\beta_\perp' \end{pmatrix}.$$

Then

$$\begin{aligned}
&T\beta_\perp'\beta_\perp\left(\hat{\beta}_\perp^{*\prime}\beta_\perp\right)^{-1}\hat{\beta}_\perp^{*\prime}S_{yy}^{-1}\hat{\beta}_\perp^*\left(\beta_\perp'\hat{\beta}_\perp^*\right)^{-1}\beta_\perp'\beta_\perp = \\
&\quad\begin{pmatrix} 0 & \beta_\perp'\beta_\perp \end{pmatrix}\begin{pmatrix} \beta'S_{yy}\beta & O_p\left(T^{-1/2}\right) \\ O_p\left(T^{-1/2}\right) & T^{-1}\beta_\perp'S_{yy}\beta_\perp \end{pmatrix}^{-1}\begin{pmatrix} 0 \\ \beta_\perp'\beta_\perp \end{pmatrix} + O_p\left(T^{-1}\right)
\end{aligned}$$

since $\hat{\beta}_\perp^* \left(\beta_\perp' \hat{\beta}_\perp^*\right)^{-1} \beta_\perp' \beta_\perp = \beta_\perp + O_p\left(T^{-1}\right)$ and $\beta_\perp' S_{yy} \beta = O_p(1)$. Letting $T \to \infty$ gives (43).

To show (45), note that $\hat{\Omega}_{HH}$ is invariant to the replacement of $\hat{\beta}^*$ by $\hat{\beta}^* \left(\beta' \hat{\beta}^*\right)^{-1} \beta' \beta = \beta + O_p\left(T^{-1}\right)$, so that

$$\hat{\Omega}_{HH} \xrightarrow{p} H_2' \beta \left(\beta'\beta\right)^{-1} \Omega_{zz} \left(\beta'\beta\right)^{-1} \beta' H_2,$$

which completes this part of the proof.

Under the alternative we have $H_2' \beta_\perp \neq 0$, so that (44) is replaced by

$$\beta_\perp' \beta_\perp \left(\hat{\beta}_\perp^{*\prime} \beta_\perp\right)^{-1} \hat{\beta}_\perp^{*\prime} H_2 = \beta_\perp' H_2 + O_p\left(T^{-1}\right).$$

Equations (43) and (45) are unchanged under the alternative, so we find that $\hat{s}_2 = O_p\left(T^2\right)$.

Parts (ii) and (iii) follow in the same way. ■

PROOF OF LEMMA 2. (i) This follows similar steps to the proof of Theorem 3(i). Under $H_{AT}: J_1'\beta = T^{-1} A_1$,

$$\begin{aligned} T J_1' \hat{\beta}^* \left(\beta' \hat{\beta}^*\right)^{-1} \beta' \beta &= T J_1' \left(\hat{\beta}^* \left(\beta' \hat{\beta}^*\right)^{-1} \beta' \beta - \beta\right) + T J_1' \beta \\ &\xrightarrow{d} -J_1' \beta_\perp \left(\int B_w B_w'\right)^{-1} \int B_w dB_z^{*\prime} + A_1, \end{aligned}$$

while

$$T J_1' S_{yy}^{-1} J_1 \xrightarrow{d} J_1' \beta_\perp \left(\int B_w B_w'\right)^{-1} \beta_\perp' J_1,$$

as in Theorem 3(i). Thus

$$\begin{aligned} \hat{s}_1 &\xrightarrow{d} \operatorname{tr} \Omega_{zz}^{*-1} \left(A_1' - \int dB_z^* B_w' \left(\int B_w B_w'\right)^{-1} \beta_\perp' J_1\right) \left[J_1' \beta_\perp \left(\int B_w B_w'\right)^{-1} \beta_\perp' J_1\right]^{-1} \times \\ &\qquad \left(A_1 - J_1' \beta_\perp \left(\int B_w B_w'\right)^{-1} \int B_w dB_z^{*\prime}\right) \\ &= (\operatorname{vec} Z_1)' \operatorname{vec} Z_1, \end{aligned}$$

where

$$Z_1 = \left[J_1' \beta_\perp \left(\int B_w B_w'\right)^{-1} \beta_\perp' J_1\right]^{-1/2} \left(A_1 - J_1' \beta_\perp \left(\int B_w B_w'\right)^{-1} \int B_w dB_z^{*\prime}\right) \Omega_{zz}^{*-1/2}.$$

Now

$$\operatorname{vec} Z_1 = \operatorname{vec}\left[J_1'\beta_\perp\left(\int B_w B_w'\right)^{-1}\beta_\perp' J_1\right]^{-1/2} A_1\Omega_{zz}^{*-1/2} + \operatorname{vec} Z,$$

where Z was defined in the proof of Theorem 3(i). Since $\operatorname{vec} Z \sim N(0, I_{p-s}\otimes I_r)$, it follows that, conditional on B_w,

$$\operatorname{vec} Z_1 \sim N\left(\operatorname{vec}\left[J_1'\beta_\perp\left(\int B_w B_w'\right)^{-1}\beta_\perp' J_1\right]^{-1/2} A_1\Omega_{zz}^{*-1/2}, I_{p-s}\otimes I_r\right),$$

and hence $(\operatorname{vec} Z_1)'\operatorname{vec} Z_1$ has a conditional non-central $\chi^2_{r(p-s)}$ distribution with non-centrality parameter

$$\operatorname{tr}\Omega_{zz}^{*-1}A_1'\left[J_1'\beta_\perp\left(\int B_w B_w'\right)^{-1}\beta_\perp' J_1\right]^{-1}A_1.$$

(ii) We proceed in a similar way to the proof of Lemma 2(i). Under $H_{AT}: H_2'\beta_\perp = T^{-1}A_2$, (44) becomes

$$\begin{aligned} T\left(\hat\beta_\perp^*\left(\beta_\perp'\hat\beta_\perp^*\right)^{-1}\beta_\perp'\beta_\perp\right)' H_2 &= T\beta_\perp' H_2 + T\left(\hat\beta_\perp^*\left(\beta_\perp'\hat\beta_\perp^*\right)^{-1}\beta_\perp'\beta_\perp - \beta_\perp\right)' H_2 \\ &= A_2 + T\beta_\perp'\left(\hat\beta^*\left(\beta'\hat\beta^*\right)^{-1}\beta'\beta - \beta\right)(\beta'\beta)^{-1}\beta' H_2 \\ &\xrightarrow{d} A_2 - \beta_\perp'\beta_\perp\left(\int B_w B_w'\right)^{-1}\int B_w dB_H', \end{aligned}$$

while (43) and (45) are unchanged. Thus

$$\begin{aligned} \hat s_2 &\xrightarrow{d} \operatorname{tr}\Omega_{HH}^{-1}\left(A_2' - \int dB_H B_w'\left(\int B_w B_w'\right)^{-1}\beta_\perp'\beta_\perp\right)\left[\beta_\perp'\beta_\perp\left(\int B_w B_w'\right)^{-1}\beta_\perp'\beta_\perp\right]^{-1}\times \\ &\qquad \left(A_2 - \beta_\perp'\beta_\perp\left(\int B_w B_w'\right)^{-1}\int B_w dB_H'\right) \\ &= (\operatorname{vec} Z_2)'\operatorname{vec} Z_2, \end{aligned}$$

where

$$Z_2 = \left[\beta_\perp'\beta_\perp\left(\int B_w B_w'\right)^{-1}\beta_\perp'\beta_\perp\right]^{-1/2}\left(A_2 - \beta_\perp'\beta_\perp\left(\int B_w B_w'\right)^{-1}\int B_w dB_H'\right)\Omega_{HH}^{-1/2}.$$

Then

$$\operatorname{vec} Z_2 = \operatorname{vec}\left[\beta_\perp'\beta_\perp\left(\int B_w B_w'\right)^{-1}\beta_\perp'\beta_\perp\right]^{-1/2} A_2\Omega_{HH}^{-1/2} + \operatorname{vec} Z,$$

where Z was defined in the proof of Lemma 2(i). Since $\operatorname{vec} Z \sim N\left(0, I_{p-r} \otimes I_{r_1}\right)$, we find

$$\operatorname{vec} Z_2 \sim N\left(\operatorname{vec}\left[\beta_\perp'\beta_\perp\left(\int B_w B_w'\right)^{-1}\beta_\perp'\beta_\perp\right]^{-1/2} A_2 \Omega_{HH}^{-1/2}, I_{p-r} \otimes I_{r_1}\right)$$

conditional on B_w. Thus $(\operatorname{vec} Z_2)' \operatorname{vec} Z_2$ has a non-central $\chi^2_{r_1(p-r)}$ distribution with non-centrality parameter

$$\operatorname{tr} \Omega_{HH}^{-1} A_2'\left[\beta_\perp'\beta_\perp\left(\int B_w B_w'\right)^{-1}\beta_\perp'\beta_\perp\right]^{-1} A_2.$$

■

Proof of Theorem 5. (i) We first write

$$\hat{S}_t'\hat{\Omega}_{zz}^{*-1}\hat{S}_t = \hat{S}_t'\left(\beta'\hat{\beta}^*\right)^{-1}\beta'\beta\left[\beta'\beta\left(\hat{\beta}^{*\prime}\beta\right)^{-1}\hat{\Omega}_{zz}^*\left(\beta'\hat{\beta}^*\right)^{-1}\beta'\beta\right]^{-1}\beta'\beta\left(\hat{\beta}^{*\prime}\beta\right)^{-1}\hat{S}_t,$$

which demonstrates the invariance of the test statistic to the normalisation of β. Then

$$\beta'\beta\left(\hat{\beta}^{*\prime}\beta\right)^{-1}\hat{\Omega}_{zz}^*\left(\beta'\hat{\beta}^*\right)^{-1}\beta'\beta \xrightarrow{p} \Omega_{zz}^*,$$

and

$$\begin{aligned}
\beta'\beta\left(\hat{\beta}^{*\prime}\beta\right)^{-1}T^{-1/2}\hat{S}_{[Ts]} &= \beta'\beta\left(\hat{\beta}^{*\prime}\beta\right)^{-1}T^{-1/2}\sum_{t=1}^{[Ts]}\hat{\beta}^{*\prime}y_t^* \\
&= T^{-1/2}\sum_{t=1}^{[Ts]}\beta' y_t^* + T\left(\hat{\beta}^{*\prime}\left(\beta'\hat{\beta}^*\right)^{-1}\beta'\beta - \beta\right)' T^{-3/2}\sum_{t=1}^{[Ts]} y_t^* \\
&\xrightarrow{d} B_z^*(s) - \int dB_z^* B_w'\left(\int B_w B_w'\right)^{-1}\int_0^s B_w(r)\,dr \\
&= B_z^*(s) - \int dB_z^* W_2'\left(\int W_2 W_2'\right)^{-1}\int_0^s W_2(r)\,dr.
\end{aligned}$$

Thus

$$\begin{aligned}
&T^{-1/2}\left[\beta'\beta\left(\hat{\beta}^{*\prime}\beta\right)^{-1}\hat{\Omega}_{zz}^*\left(\beta'\hat{\beta}^*\right)^{-1}\beta'\beta\right]^{-1/2}\beta'\beta\left(\hat{\beta}^{*\prime}\beta\right)^{-1}\hat{S}_{[Ts]} \\
&\xrightarrow{d} W_1(s) - \int dW_1 W_2'\left(\int W_2 W_2'\right)^{-1}\int_0^s W_2(r)\,dr = V(s)
\end{aligned}$$

and

$$T^{-2}\sum_{t=1}^{T}\hat{S}_t'\hat{\Omega}_{zz}^{*-1}\hat{S}_t \xrightarrow{d} \int V'V$$

as required.

Parts (ii) and (iii) follow in the same way with B_w, W_2 and V replaced by $\bar{B}_w$, $\bar{W}_2$ and $\bar{V}$ respectively in part (ii), and $\tilde{B}_w$, $\tilde{W}_2$ and $\tilde{V}$ in part (iii).

(iv) Under the alternative there are $s\,(< r)$ cointegrating vectors, so we can partition $\hat{\beta}^*$ as

$$\hat{\beta}^* = \left(\underset{(p\times s)}{\hat{\beta}^*_s} \quad \underset{(p\times(r-s))}{\hat{\beta}^*_{r-s}} \right),$$

so that $\hat{\beta}^*_s$ provides a consistent estimator of the $p \times s$ matrix of cointegrating vectors β, while $\hat{\beta}^*_{r-s}$ asymptotically lies only in the space spanned by $\beta_\perp$. Therefore, when we partition $\hat{z}^*_t$ as

$$\hat{z}^*_t = \begin{pmatrix} \hat{\beta}^{*\prime}_s y^*_t \\ \hat{\beta}^{*\prime}_{r-s} y^*_t \end{pmatrix},$$

the first s components are asymptotically $I\,(0)$, while the remaining $r-s$ components are asymptotically $I\,(1)$. Therefore, $T^{-1/2}S_t$ is $O_p\left(T^{1/2}\right)$, and $\hat{\Omega}^*_{zz}$ is $O_p\,(mT)$ when applied to $I\,(1)$ time series (see Phillips (1991b)). Combining these results gives that $\hat{c}$ is $O_p\,(T/m)$ under the alternative. That $\bar{c}$ and $\tilde{c}$ are also $O_p\,(T/m)$ follows in the same way. ■

7 References

Ahn, S.K. and G.C. Reinsel (1990) Estimation for Partially Nonstationary Multivariate Autoregressive Models. *Journal of the American Statistical Association* 85, 813–823.

Andrews, D.W.K. (1991) Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* 59, 817–858.

Brillinger, D.R. (1981) *Time Series: Data Analysis and Theory.* New York: McGraw-Hill.

Choi, I. and B.C. Ahn (1995) Testing for Cointegration in a System of Equations, *Econometric Theory* 11, 952–983.

Engle, R.F. and B.S. Yoo (1991) Cointegrated Time Series: An Overview with New Results. In R.F. Engle and C.W.J. Granger (eds.) *Long Run Economic Relationships:*

Readings in Cointegration, Oxford: Oxford University Press.

Gonzalo, J. (1994) Five Alternative Methods of Estimating Long-Run Equilibrium Relationships. *Journal of Econometrics* 60, 203–233.

Hansen, B.E. (1992a) Tests for Parameter Instability in Regressions with I(1) Processes, *Journal of Business and Economic Statistics* 10, 321–335.

Hansen, B.E. (1992b) Convergence to Stochastic Integrals for Dependent Heterogeneous Processes, *Econometric Theory* 8, 489–500.

Harris, D. and B. Inder (1994) A Test of the Null Hypothesis of Cointegration. In C. Hargreaves (ed.), *Nonstationary Time Series Analysis and Cointegration*, pp. 133–152. Oxford: Oxford University Press.

Inder, B. (1995) Finite Sample Arguments for Appropriate Estimation of Cointegrating Relationships. Working Paper, Monash University.

King, R.G., C.I. Plosser, J.H. Stock and M.W. Watson (1991) Stochastic Trends and Economic Fluctuations. *The American Economic Review* 81, 819–840.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992) Testing the Null of Stationarity against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root? *Journal of Econometrics* 54, 159–178.

Johansen, S. (1988) Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control* 12, 231–254.

Johansen, S. (1991) Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica* 59, 1551–1580.

Johansen, S. and K. Juselius (1992) Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK. *Journal of Econometrics* 53, 211–244.

Park, J.Y. (1992) Canonical Cointegrating Regressions. *Econometrica* 60, 119–143.

Park, J.Y. and P.C.B. Phillips (1988) Statistical Inference in Regressions with Integrated Processes: Part 1. *Econometric Theory* 4, 468–498.

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

Phillips, P.C.B. (1991b) Spectral Regression for Cointegrated Time Series. In W.A. Barnett, J.Powell and G.E. Tauchen (eds), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp413–435. Cambridge: Cambridge University Press.

Phillips, P.C.B. and S.N. Durlauf (1986) Multiple Time Series Regression with Integrated Processes. *Review of Economic Studies* 53, 473–495.

Phillips, P.C.B. and B.E. Hansen (1990) Statistical Inference in Instrumental Variables Regression with I(1) Processes. *Review of Economic Studies* 57, 99–125.

Phillips, P.C.B. and M. Loretan (1991) Estimating Long-run Economic Equilibria. *Review of Economic Studies* 58, 407–436.

Phillips, P.C.B. and S. Ouliaris (1990) Asymptotic Properties of Residual-Based Tests for Cointegration. *Econometrica* 58, 165–193.

Saikkonen, P. (1991) Asymptotically Efficient Estimation of Cointegration Regressions. *Econometric Theory* 7, 1–21.

Saikkonen, P. (1992) Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation. *Econometric Theory* 8, 1–27.

Shin, Y. (1994) A Residual-Based Test of the Null of Cointegration against the Alternative of No Cointegration. *Econometric Theory* 10, 91–115.

Stock, J.H. (1987) Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors. *Econometrica* 55, 1035–1056.

Stock, J.H. and M.W. Watson (1988) Testing for Common Trends. *Journal of the American Statistical Association* 83, 1097–1107.

Stock, J.H. and M.W. Watson (1993) A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. *Econometrica* 61, 783–820.

Yang, M. (1994) Canonical Correlation Analysis of Cointegrated Processes. University of New South Wales, School of Economics Discussion Paper 94/10.