barcelona gse
graduate school of economics

Regression-based Analysis of Cointegration Systems

Javier Gomez-Biscarri
Javier Hualde

September 2014

Barcelona GSE Working Paper Series

Working Paper nº 780

Regression-based analysis of cointegration systems

Javier Gomez-Biscarri
Universitat Pompeu Fabra and Barcelona GSE

Javier Hualde*
Universidad Pública de Navarra

September 10, 2014

Abstract

Two estimation procedures dominate the cointegration literature: Johansen's maximum likelihood inference on vector autoregressive error correction models and estimation of Phillips' triangular forms. This latter methodology is essentially semiparametric, focusing on estimating long run parameters by means of cointegrating regressions, but it is less used in practice than Johansen's approach since its implementation requires prior knowledge of features, such as the cointegrating rank and an appropriate set of non-cointegrated regressors. In this paper we develop a simple and automatic procedure (based on unit root and regression-based cointegration testing) which, without imposing a parametric specification for the short run components of the model, provides an estimator of the cointegrating rank and data-based just-identifying conditions for the cointegrating parameters (leading to a Phillips' triangular form). A Monte Carlo experiment and an empirical example are also provided.

JEL Classification: C32.

Keywords. Cointegrating space; Phillips' triangular form; Johansen's methodology; regression-based cointegration testing.

*Corresponding author. Department of Economics, Universidad Pública de Navarra. Campus de Arrosadía. Edificio Los Madroños 31006 Pamplona, Spain. Phone: +34 948 169674. Fax: +34 948 169721. Email: javier.hualde@unavarra.es

1 Introduction

Cointegration has been one of the main workhorses of time series econometrics in the last two decades and, even if the literature is somewhat mature, it still attracts substantial attention from both theoretical and empirical perspectives (see, e.g., Hoover et al., 2008, or Johansen, 2010). Two approaches to the estimation of cointegration systems appear to be dominant. The first, developed by Johansen (1988, 1991) and Ahn and Reinsel (1990), focuses on maximum likelihood inference on vector autoregressive (VAR) error correction models. This approach has been the most popular mainly because it both provides an estimator of the cointegrating rank and leads to empirical (data-based) just-identifying restrictions from which estimators of the cointegrating vectors can be easily obtained. Additionally, it offers estimators of the short run parameters and a neat hypothesis testing procedure, where given economic theories can be checked. The second dominant strategy focuses on estimation of the so-called Phillips' triangular form (Phillips, 1991a), which consists of specifying the cointegrating relations by a set of reduced form regression equations from which estimation of structural equations (those with economic meaning) can be derived (see Saikkonen, 1993). This approach relates directly to the simultaneous equations models methodology, with long tradition in econometrics. Within this setting, different estimation methods have been proposed and, noticeably, it has been shown that parametric assumptions on the short run components do not lead to gains in asymptotic efficiency in the estimation of cointegrating vectors (see Phillips and Hansen, 1990, Phillips, 1991b). Pesaran and Shin (2002) provide a comparison of both methods.

In contrast to Johansen's approach, Phillips' methodology is essentially semiparametric, focusing on the long run components of the model and taking an agnostic approach about the short run dynamics, which, in any case, once parameterized, can be subsequently estimated if desired. While this appears to be an attractive feature (compared to a fully parametric approach), there are several limitations associated to Phillips' methodology. First, the procedure takes the cointegrating rank as given. Moreover, based on this rank, the vector of observables is decomposed into two subvectors corresponding to dependent variables and regressors in particular cointegrating regressions. Specifically, the number of dependent variables should be the same as the cointegrating rank and the rest of the variables (regressors) must not be cointegrated, otherwise they would be perfectly correlated asymptotically. In short, Phillips' approach imposes a priori identification conditions on the cointegrating parameters, and this leads to uncertainty on how to act in practice. As a result, in comparison to Johansen's approach, this methodology has been hardly used in applied work, except

in the single-equation framework, where an extensive literature on testing for cointegration exists (see, e.g., Haug, 1996, for a review). Note, however, that when facing multivariate systems the possibility that various cointegrating relations occur needs to be allowed for, so, at first sight, single-equation regression approaches appear not to be useful. Single-equation tests can be extended to systems of equations (see Ahn and Choi, 1995), but this would still require imposing identifying restrictions to design cointegrating regressions, so the application of such methods in practice remains speculative.

In this paper we develop a procedure to infer the cointegrating rank and to design a set of regressions from which the cointegrating vectors in system frameworks can be estimated. Our analysis focuses on $I(1)$ systems (where, after differencing and possibly demeaning, the vector of observables is covariance stationary with nonzero and bounded spectral density), but our method allows for simple extensions to higher order settings (although we do not pursue this here). Given the extensive literature on cointegration, it is warranted that we highlight the extent of our contribution. First, our proposal requires neither the imposition of a priori identifying conditions nor the specification of a parametric model for the short run components. There are procedures in the literature which achieve a similar goal, like the principal components approach (see Stock and Watson, 1988, Harris, 1997, Snell, 1999), the nonparametric method of Bierens (1997), or the test of common stochastic trends of Nyblom and Harvey (2000). However, while the main focus of these proposals is to test for a particular cointegrating rank r, they do not provide a formal discussion of estimation of r (with the exception of Bierens, 1997), and their estimators of the cointegrating vectors are based on eigenvalues routines and orthogonality restrictions, which might be difficult to interpret. On the contrary, we offer a formal discussion of the properties of our estimator of r and, in addition, this estimator is based on extremely simple techniques (like unit root testing) which belong to the standard time series toolkit. Second, once the cointegrating rank is determined, our method provides data-based just-identifying restrictions leading to a Phillips' triangular form. In particular, our proposal identifies automatically the set of regressors from which a Phillips' triangular form can be straightforwardly estimated without imposing any a priori identification conditions. Thus, in practice, we provide a method which makes the application of Phillips' approach feasible, hence offering a valid alternative to Johansen's methodology. We believe there are relevant contexts where our proposal might indeed enjoy clear advantages over Johansen's. In particular, avoiding parametric assumptions on the short-run dynamics makes the method more robust to misspecification and it also

seems preferable in high dimensional models. In the latter, parametric prescriptions would lead to estimating a very large number of parameters, thus possibly inducing small sample problems. Additionally, Gonzalo and Lee (1998) showed that residual-based cointegration tests are more robust than Johansen's likelihood ratio (LR) type of tests to empirically relevant departures from the model, such as autoregressive processes with roots (marginally) larger than unity or stochastic roots, mistaken order of integration of the system ($I(2)$ taken as $I(1)$ with drift), wrong choice of deterministic components or fractional processes. Finally, our proposal sheds light on the delicate issue of choosing appropriately the regressors in cointegrating regressions. Here, our results appear to be useful even in bivariate settings, where residual-based cointegration testing is routinely applied by practitioners, but where a wrong design of the possible cointegrating regression (due to cointegrated regressors) might lead to erroneous conclusions.

The paper is organized as follows. In Section 2 we introduce some preliminary concepts and a result on which our methodology is based. In Section 3 we present a method to select common trends which, as will be seen below, is an essential component of our procedure. In Section 4 we introduce our estimator of the cointegrating rank. Next, in Section 5, we compare the finite sample performance of our procedure with that of Johansen's trace test (see, e.g., Johansen, 1995). In Section 6 we present an empirical analysis of the term structure of US interest rates and, finally, in Section 7, we conclude. Proofs of theorems are relegated to the Appendix.

2 Preliminary concepts and results

First, we introduce some definitions. We say that a scalar or vector process ξ_t is integrated of order zero ($\xi_t \sim I(0)$) if $\xi_t - E(\xi_t)$ is covariance stationary with nonzero and bounded spectral density at all frequencies. Then, a scalar or vector ζ_t is integrated of order one ($\zeta_t \sim I(1)$), if $\Delta\zeta_t$ is $I(0)$, where $\Delta = 1 - L$, L being the lag operator. Note that if a vector ζ_t is $I(1)$, our definition (which is almost identical to that of Johansen, 1995) implies that at least one of the individual components of ζ_t is $I(1)$, but, in general, an $I(1)$ vector is allowed to have individual components with distinct integration orders.

Next, we define cointegration for $I(1)$ processes. Given a $p \times 1$ process $z_t \sim I(1)$, z_t is cointegrated if there exists a $p \times 1$ vector $\gamma \neq 0$ such that $\gamma' z_t - E(\gamma' z_t)$ (prime denoting transposition) can be made covariance stationary by a suitable choice of initial values. Hereafter, a process which can be made covariance stationary (or $I(0)$)

by a suitable choice of initial values will be just denoted as stationary (or $I(0)$). Again, this definition is almost identical to that of Johansen (1995) and it is significantly more general than the standard notion of Engle and Granger (1987), where all observables are required to have identical integration orders. Note that, according to our definition, some of the cointegrating vectors might be trivial, just indicating that a particular (possibly demeaned) observable is stationary. Also, note that $\gamma' z_t$ need not be $I(0)$ (e.g. if $\gamma' z_t$ is noninvertible), although $I(0)$ cointegrating linear combinations of the observables might be the most common situation encountered in practice. As usual, the cointegrating rank among the elements of z_t (denoted as r) is the number of linearly independent cointegrating vectors, and the space generated by these vectors (whose dimension is r) will be denoted as contegrating space.

A very general model which generates a possibly $I(1)$ and cointegrated $p \times 1$ vector of observables z_t is

$$\Upsilon \Delta (z_t - \mu_t) = u_t, \tag{1}$$

where Υ is a $p \times p$ nonsingular matrix, μ_t is a deterministic component and u_t is a zero-mean $p \times 1$ covariance stationary process which satisfies one (and only one) of the following conditions:

(i) u_t is $I(0)$ with nonsingular spectral density matrix at all frequencies;

(ii) some components of u_t form a subvector which is $I(0)$ with nonsingular spectral density matrix at all frequencies, the rest of the components forming another subvector which is the first difference of a zero-mean stationary process with bounded spectral density matrix at all frequencies;

(iii) u_t is the first difference of a zero-mean stationary process with bounded spectral density matrix at all frequencies.

We also set $E(z_0) = \mu_0$, so (1) immediately implies that $E(z_t) = \mu_t$, $t \geq 1$. In (1), the integration and cointegration properties of z_t depend both on Υ and u_t. For example, under (i) or (ii), $u_t \sim I(0)$, so $\Upsilon^{-1} u_t \sim I(0)$, and it immediately follows that $z_t \sim I(1)$. Also, under (i) all individual components of z_t are $I(1)$ and they do not cointegrate, so $r = 0$. On the other hand, under (iii), $u_t = v_t - v_{t-1}$ where v_t is a stationary process with bounded spectral density matrix at all frequencies. Then

$$z_t - \mu_t = z_0 - \mu_0 - \Upsilon^{-1} v_0 + \Upsilon^{-1} v_t, \tag{2}$$

so $z_t - \mu_t$ can be made stationary by setting $z_0 - \mu_0 = \Upsilon^{-1} v_0$. This represents an extreme case of cointegration where the demeaned vector of observables is itself

stationary. In this case we will say that the cointegrating rank is full (or $r = p$). Under (ii), if the individual components of $\Upsilon^{-1}u_t$ are $I(0)$, then all components of z_t are $I(1)$, but some (but not all) observables are allowed to be stationary. Also, (ii) permits other cointegration possibilities with $0 < r < p$, noting that

$$\Upsilon(z_t - \mu_t) = \Upsilon(z_0 - \mu_0) + \sum_{j=1}^{t} u_j. \tag{3}$$

In fact, $p - r$ is the number of $I(0)$ components in u_t. Although more general structures could be allowed for, as in Johansen (1995), (ii), (iii) imply that the vector of cointegrating errors is stationary with bounded spectral density. This vector might be $I(0)$ (as in Johansen, 1995), but our assumptions allow for more general possibilities (e.g., all cointegrating errors are noninvertible).

Next, we present a simple result which gives necessary and sufficient conditions for the existence of a particular cointegrating rank r. Below, we trivially refer to a scalar process being cointegrated if it is stationary.

Theorem 1. Let z_t be generated by (1). Then the cointegrating rank among the elements of z_t is $r \in \{1, ..., p\}$ if and only if a. and b. hold, where

a. There exists a subvector of z_t of dimension $p - r$ (say $z_{(b)t}$) whose individual components are $I(1)$ and do not cointegrate;

b. All subvectors of z_t of dimension larger than $p - r$ containing $z_{(b)t}$ cointegrate.

Remark 1. If $r \in \{1, ..., p-1\}$, by Theorem 1 there exist $p - r$ variables in z_t which are individually $I(1)$ and not cointegrated. Say these variables are Qz_t, where Q is a $(p-r) \times p$ selection matrix. The rest of the variables are Pz_t, where P is a corresponding $r \times p$ selection matrix. Again, by Theorem 1, any set of $p-r+1$ variables formed by any of the variables in Pz_t and all those in Qz_t is always cointegrated, so there exists an $r \times (p-r)$ matrix B such that

$$P(z_t - \mu_t) = BQ(z_t - \mu_t) + w_t, \tag{4}$$

where the r-dimensional zero-mean process w_t is stationary (the connection between w_t and u_t in (1) is explored below). There is no loss of generality in (4) (which implies a normalization of the cointegration vectors) because the variables in Pz_t need to be present with nonzero coefficients in the cointegrating relations. (4) implies that there is always a particular ordering of the observables leading to a Phillips' triangular form ((4) being the first block of this form). Note that B could take arbitrary values, even

$B = 0_{r\times(p-r)}$ if, after demeaning, all the variables in Pz_t are stationary. Related to (1), (4) leads to restrictions on Υ and on the structure of u_t. In particular, (4) and the $I(1)$ and noncointegrated condition of the individual components of Qz_t is captured by (1) letting

$$\Upsilon = \begin{pmatrix} P - BQ \\ Q \end{pmatrix}, \tag{5}$$

$w_t = (P - BQ)(z_0 - \mu_0) + \sum_{j=1}^{t} u_{(a)j}$, where $u_t = (u'_{(a)t}, u'_{(b)t})'$, $u_{(a)t}$, $u_{(b)t}$ being r and $(p-r)$-dimensional subvectors of u_t, respectively, such that $u_{(a)t}$ is the first difference of a zero-mean r-dimensional stationary process with bounded spectral density and $u_{(b)t}$ is $I(0)$ with nonsingular spectral density at all frequencies. Also, noting that $Q(z_t - \mu_t) = Q(z_0 - \mu_0) + \sum_{j=1}^{t} u_{(b)j}$, all individual components of Qz_t are $I(1)$ and they do not cointegrate. Given r, any $(p-r)$-dimensional set of individually $I(1)$ and noncointegrated variables (like Qz_t) will be denoted as a set of common trends (and also, a variable in such set will be referred to as a common trend).

Remark 2. Theorem 1 motivates a step-wise method to infer the rank. Specifically, the key is to find sequentially variables which might be common trends. The heuristic idea is to search initially for at least one $I(1)$ observable. If we suspect that there is none, then there is evidence in favour of $r = p$. If, on the contrary, we suspect that (at least) a particular observable is $I(1)$, then there is evidence in favour of $r < p$. In our sequential method this will be denoted as Step 1. In that latter case, we select a candidate for common trend, and check whether all pairs containing that candidate are cointegrated. If they are, then there is evidence in favour of $r = p-1$. Otherwise, we suspect that $r < p-1$, because there is (at least) a pair of observables which appear not to be cointegrated (Step 2). We then select an additional candidate for common trend, and test for cointegration within sets of three observables (Step 3) and so on. In Section 3 we explain how the choice of common trends is done at every step of the procedure and justify the properties of the choice rule. Then, in Section 4 we present our formal procedure to estimate the rank and justify the properties of our proposed rank estimator.

Finally, we introduce some additional notation. Denote by $a_{j,t}$ the jth element of a vector a_t. Given l distinct natural numbers k_j, $j = 1, ..., l$, $l \in \{1, ..., p\}$ such that $1 \leq k_1, k_2, ..., k_l \leq p$, we define the null and alternative hypotheses:

$$H_{k_1,k_2,...,k_l} : \text{the observables } z_{k_1,t}, z_{k_2,t}, ..., z_{k_l,t} \text{ are not cointegrated,} \tag{6}$$

$$\overline{H}_{k_1,k_2,...,k_l} : \text{the observables } z_{k_1,t}, z_{k_2,t}, ..., z_{k_l,t} \text{ are cointegrated,} \tag{7}$$

where if $l = 1$, H_{k_1} is equivalent to $z_{k_1,t} \sim I(1)$. Our procedure will be essentially based on consistent tests for $H_{k_1,k_2,...,k_l}$ and related test statistics $\widehat{\tau}_{k_1,k_2,...,k_l}$. Although other options are possible, for reasons pinpointed in the Introduction we advocate here for the use of traditional unit root and residual-based cointegration tests. These are simple to compute and also widely available to practitioners. Typical examples are the Augmented Dickey-Fuller (ADF) or Phillips-Perron (PP) methods (along with their residual-based versions). Note that the residual-based testing procedures rely heavily on the specification of a possible cointegrating regression where the regressors need to be carefully chosen. Section 3 below will shed light on this issue. Notationally, if $l > 1$, $\widehat{\tau}_{k_1,k_2,...,k_l}$ corresponds to a residual-based test statistic where $z_{k_l,t}$ is the left hand side variable in the regression from which residuals are derived, whereas if $l = 1$, $\widehat{\tau}_{k_1}$ is a standard unit root test statistic. Note that these statistics depend on user-chosen deterministic structures (typically a constant or a constant plus a time trend). Although allowing for more generality is possible, we impose throughout that for each l, the deterministic structure used when calculating $\widehat{\tau}_{k_1,k_2,...,k_l}$ is identical irrespective of the particular variables involved in the statistic. This is not restrictive, as it simply amounts to assume, e.g., that μ_t is a polynomial of time, say $\mu_t = \sum_{j=0}^{d} \phi_j t^j$, where d is not underspecified (typically, $d = 0, 1, 2$), being also relatively unspecific about zero restrictions on vectors ϕ_j and $\Upsilon\phi_j$. In some cases this might lead to test statistics based on regressions with irrelevant deterministic components (which might cause power loses), but this appears to be a second order problem compared to the underspecification of the deterministic structure.

We characterize further the behaviour of $\widehat{\tau}_{k_1,k_2,...,k_l}$. For a certain continuous random variable ζ_l, let

$$\widehat{\tau}_{k_1,k_2,...,k_l} \rightarrow_d \zeta_l, \text{ under } H_{k_1,k_2,...,k_l}, \tag{8}$$

where "$\rightarrow_d$" denotes convergence in distribution. Note that (8) implies that for any $\alpha \in (0,1)$, there exist real numbers $\nu_l(\alpha)$, $l = 1, ..., p$, such that

$$\lim_{n\to\infty} \Pr\left(\widehat{\tau}_{k_1,k_2,...,k_l} < \nu_l(\alpha)\right) = \alpha, \text{ under } H_{k_1,k_2,...,k_l}, \tag{9}$$

where n denotes sample size. In addition, if $z_{k_1,t}, ..., z_{k_{l-1},t}$ are common trends (obviously for $l = 1$ this requirement does not apply), we assume that for any arbitrarily large positive K,

$$\lim_{n\to\infty} \Pr\left(\widehat{\tau}_{k_1,k_2,...,k_l} < -K\right) = 1 \text{ under } \overline{H}_{k_1,k_2,...,k_l}. \tag{10}$$

It can be readily shown that, under very general conditions, the ADF and PP

procedures (and their residual-based versions) satisfy (9), (10). Note also that if $z_{k_1,t}, ..., z_{k_{l-1},t}$ are common trends, these tests are consistent even in the case where $z_{k_l,t}$ (possibly demeaned) is stationary, so the possibility of a trivial cointegrating relation with a unit cointegrating vector will be detected. Alternatively, a residual-based unit root test based on a regression where the regressors are cointegrated (that is, we are under the alternative) is not, in general, consistent. This is the reason why the regression needs to be very carefully designed. Throughout, we assume that the tests for $H_{k_1,k_2,...,k_l}$ are performed with asymptotic level α, so, noting (9), (10), this implies that for any $k_1, k_2, ..., k_l$, $l \geq 1$, the hypothesis $H_{k_1,k_2,...,k_l}$ is rejected if $\widehat{\tau}_{k_1,k_2,...,k_l} < \nu_l(\alpha)$.

3 The choice of common trends

As pinpointed in Remark 2, one of the key aspects of our procedure is the sequential choice of variables which act as common trends. In fact, this issue is related to a typical problem (somehow overlooked by the literature) encountered when using residual-based cointegration testing techniques, which is the choice of the dependent variable in the possible cointegrating regression from which residuals are derived. In particular, a residual-based unit root test based on a regression where the regressors are cointegrated is not, in general, consistent, so the choice of left-hand side variable and regressors is critical. We illustrate this problem in an $I(1)$ setting by means of a simple bivariate example. Let $z_{1,t}$, $z_{2,t}$ be two observables such that $z_{1,t} \sim I(1)$, $z_{2,t} \sim I(1)$ and which do not cointegrate. Suppose we want to test whether these variables are cointegrated by means of a residual-based method. Here, the key issue is to choose the regressor in the possible cointegrating regression from which the residuals are derived. Noting (9), (10), a sensible choice to favour the possibility that the regressor is $I(1)$ (which ensures that the test is consistent under cointegration), might be to select as regressor the observable which shows the highest evidence of being $I(1)$ based on a certain criterion. Thus, a possible strategy might be to choose as regressor the observable for which $\widehat{\tau}_i$ is the highest. This choice will be denoted as the "naive" choice, and it implies that the residual based statistic which depends on such stochastic choice of regressor is

$$\widehat{\tau} = \widehat{\tau}_{12} 1(\widehat{\tau}_1 > \widehat{\tau}_2) + \widehat{\tau}_{21} 1(\widehat{\tau}_1 \leq \widehat{\tau}_2). \tag{11}$$

The challenge is to derive the null limiting distribution of $\widehat{\tau}$, noting that in our setting both observables are $I(1)$, so $\widehat{\tau}_{12}$, $\widehat{\tau}_{21}$, are well constructed statistics such that

$$\lim_{n\to\infty} \Pr\left(\widehat{\tau}_{12} < \nu_2(\alpha)\right) = \alpha, \qquad \lim_{n\to\infty} \Pr\left(\widehat{\tau}_{21} < \nu_2(\alpha)\right) = \alpha. \tag{12}$$

However, in view of (11), the derivation of the null limiting distribution of $\widehat{\tau}$ might be very complicated, as it depends on the joint behaviour of $\widehat{\tau}_1$, $\widehat{\tau}_2$, $\widehat{\tau}_{12}$, $\widehat{\tau}_{21}$. Thus, it is likely that this distribution is nuisance-parameter dependent, so, using standard critical values $\nu_2(\alpha)$, whether $\lim_{n\to\infty} \Pr\left(\widehat{\tau} < \nu_2(\alpha)\right) \leq \alpha$ or not remains speculative (unreported Monte Carlo results suggest that $\widehat{\tau}$ is oversized). Under the alternative, that is, when $z_{1,t}$, $z_{2,t}$, are cointegrated, the "naive choice" works well. First, if both $z_{1,t}$ and $z_{2,t}$ are $I(1)$, both statistics $\widehat{\tau}_{12}$ and $\widehat{\tau}_{21}$ diverge to $-\infty$, hence $\widehat{\tau}$ also does it and the test is consistent. If, possibly after demeaning, one of the observables is stationary (say $z_{1,t}$), the "naive" choice selects as common trend the $I(1)$ variable ($z_{2,t}$) with probability approaching one (wpa1). Then, noting (11), $1\left(\widehat{\tau}_1 > \widehat{\tau}_2\right) \to_p 0$, $1\left(\widehat{\tau}_1 \leq \widehat{\tau}_2\right) \to_p 1$, $\widehat{\tau}_{12} = O_p(1)$ (note that this statistic is constructed from the residuals of a regression of $z_{2,t}$, which is $I(1)$, on $z_{1,t}$, which is stationary, so even if $z_{1,t}$, $z_{2,t}$, are cointegrated, those residuals behave like an $I(1)$ variable) whereas $\widehat{\tau}_{21}$ diverges to $-\infty$, so the test is consistent.

A simple approach to the problem of calculating the null limiting distribution of the test statistic is to rely on a different choice rule which, even if both observables are $I(1)$, selects one (but just one) of the $I(1)$ variables as regressor wpa1. An example of such choice rule is the following. Let $\widehat{\kappa}_2 > 0$ be a statistic such that, as $n \to \infty$,

$$\widehat{\kappa}_2 = o_p(1) \text{ under } H_2, \tag{13}$$

and for a real sequence $g_n > 0$ such that either $g_n = 1$ or $g_n \to 0$, and a random variable $\xi_2 > 0$, a.s., let

$$g_n \widehat{\kappa}_2 \to_d \xi_2 \text{ under } \overline{H}_2. \tag{14}$$

A very convenient example of such $\widehat{\kappa}_2$ is the inverse of the usual variance ratio statistic (see Tanaka, 1990, Kwiatkowski et al., 1992, Breitung, 2002) applied to process $z_{2,t}$ (demeaned if needed). In particular

$$\widehat{\kappa}_2 = \frac{n \sum_{t=1}^{n} \left(z_{2,t}^*\right)^2}{\sum_{t=1}^{n} \left(\sum_{s=1}^{t} z_{2,s}^*\right)^2}, \tag{15}$$

where $z_{2,t}^*$ equals the observable $z_{2,t}$ or residuals arising from the regression of $z_{2,t}$

on deterministic components (typically a drift and/or a time trend). It is relatively cumbersome (but simple) to show that, under standard conditions, such $\widehat{\kappa}_2$ satisfies (13), (14), with g_n depending on the memory properties of $z_{2,t}$: if $z_{2,t} \sim I(0)$, $g_n = 1$, whereas if $z_{2,t}$ has negative memory, g_n converges to zero at a rate which depends on the degree of memory (e.g., $g_n = n^{-1}$ if $z_{2,t}$ is noninvertible). Then, the rule we propose is as follows: choose $z_{1,t}$ as common trend if $\widehat{\tau}_1^2 < \widehat{\tau}_2^2 \min\{1, \widehat{\kappa}_2\}$; choose $z_{2,t}$, otherwise. The intuition behind this rule is the following. If $\widehat{\kappa}_2$ is large (in particular larger than one), which may suggest that, after demeaning, $z_{2,t}$ is stationary, we let $\widehat{\tau}_1^2$ and $\widehat{\tau}_2^2$ compete, so the regressor is chosen on the basis of the evidence of unit root statistics. Alternatively, if $\widehat{\kappa}_2$ is small (in particular smaller than one), thus suggesting that $z_{2,t}$ might be $I(1)$, we choose $z_{2,t}$ unless the evidence based on the unit root statistics is very strong in favour of $z_{1,t}$ (because $\widehat{\tau}_1^2$ is much smaller than $\widehat{\tau}_2^2$). This procedure leads to choosing $z_{2,t}$ as the common trend unless either there is clear evidence that, after demeaning, $z_{2,t}$ is stationary or an overwhelming evidence of $I(1)$ behaviour in favour of $z_{1,t}$ (in practice, the value of $\widehat{\kappa}_2$ when $z_{2,t} \sim I(1)$ is usually very small, so this latter possibility is not expected to happen very often).

As shown in Theorem 2 below, this type of rule implies that a specific true common trend is chosen wpa1. In our present example, given that both $z_{1,t}$ and $z_{2,t}$ are $I(1)$, the rule selects wpa1 $z_{2,t}$. Thus, noting that

$$\widehat{\tau} = \widehat{\tau}_{12} 1(z_{1,t} \text{ is chosen as common trend}) + \widehat{\tau}_{21} 1(z_{2,t} \text{ is chosen as common trend}),$$

$\widehat{\tau} - \widehat{\tau}_{21} = o_p(1)$, so the null limiting distribution of $\widehat{\tau}$ is that of $\widehat{\tau}_{21}$.

There is, however, an important issue to be noted. Our choice rule (and therefore our statistic $\widehat{\tau}$) is sensitive to the ordering of the observables in the vector z_t. Specifically, if the second observable in the vector $(z_{2,t})$ is $I(1)$ it is chosen as common trend wpa1. However an alternative ordering of the observables might lead to a different statistic. In our particular bivariate example there are two different orderings (say $o = 1$, $o = 2$), leading therefore to two different statistics $\widehat{\tau}_{(1)}$, $\widehat{\tau}_{(2)}$, respectively. The statistic which captures the possibility that both orderings might be chosen is $\widetilde{\tau} = \widehat{\tau}_{(1)} 1(o = 1) + \widehat{\tau}_{(2)} 1(o = 2)$, and the challenge is to derive the null limiting distribution of $\widetilde{\tau}$. Clearly

$$\begin{aligned} \Pr(\widetilde{\tau} < \nu_2(\alpha)) &= \Pr\left(\widehat{\tau}_{(1)} < \nu_2(\alpha) \middle| o = 1\right) \Pr(o = 1) \\ &\quad + \Pr\left(\widehat{\tau}_{(2)} < \nu_2(\alpha) \middle| o = 2\right) \Pr(o = 2). \end{aligned} \tag{16}$$

Let o be independent from the observables, which is innocuous if the chosen or-

dering is arbitrary. Then, by independence, the conditional probabilities are identical to the unconditional ones and noting that with the previous choice of common trend $\lim_{n\to\infty}\Pr\left(\widehat{\tau}_{(1)}<\nu_2\left(\alpha\right)\right)=\lim_{n\to\infty}\Pr\left(\widehat{\tau}_{(2)}<\nu_2\left(\alpha\right)\right)=\alpha$, it follows that $\lim_{n\to\infty}\Pr\left(\widetilde{\tau}<\nu_2\left(\alpha\right)\right)=\alpha$.

As explained in Remark 2, our general procedure to estimate the rank requires a sequential choice of possible common trends, and generalizing the idea posed in the previous bivariate example leads to the following method. Let $\widehat{\theta}_{k_1,k_2,...,k_l}=\widehat{\tau}^2_{k_1,k_2,...,k_l}$, $\widehat{\theta}^*_{k_1,k_2,...,k_l}=\widehat{\tau}^2_{k_1,k_2,...,k_l}\min\left\{1,\widehat{\kappa}_{k_1,k_2,...,k_l}\right\}$, where $\widehat{\kappa}_{k_1,k_2,...,k_l}$ is the inverse of the variance ratio statistic (see (15)) applied either to $z^*_{k_1,t}$ if $l=1$ or to residuals from the regression of $z_{k_l,t}$ on $z_{k_1,t},...,z_{k_{l-1},t}$ and deterministic components if $l>1$. As before, it can be shown that

$$\widehat{\kappa}_{k_1,k_2,...,k_l}=o_p\left(1\right)\text{ under }H_{k_1,k_2,...,k_l}, \tag{17}$$

and that if $z_{k_1,t},...,z_{k_{l-1},t}$ are common trends (obviously for $l=1$ this requirement does not apply), for a real sequence $g_n>0$ such that either $g_n=1$ or $g_n\to 0$ and a random variable $\xi_{k_1,k_2,...,k_l}>0$, a.s.,

$$g_n\widehat{\kappa}_{k_1,k_2,...,k_l}\to_d\xi_{k_1,k_2,...,k_l}\text{ under }\overline{H}_{k_1,k_2,...,k_l}. \tag{18}$$

As anticipated in Remark 2 (and formally discussed in Section 4 below), the first possible common trend variable, say $z_{c_1,t}$, $c_1\in\{1,...,p\}$, is chosen in Step 2 of the procedure. In particular, we select $c_1=1$ if $\widehat{\theta}_1<\min_{j>1}\widehat{\theta}^*_j$; otherwise, $c_1=2$ if $\widehat{\theta}_2<\min_{j>2}\widehat{\theta}^*_j$; proceeding in this fashion, if $c_1\neq 1,2,...,p-2$, then $c_1=p-1$ if $\widehat{\theta}_{p-1}<\widehat{\theta}^*_p$; otherwise, $c_1=p$. This rule is a straightforward generalization to the p-dimensional case of the rule presented for the bivariate case.

The second possible common trend, say $z_{c_2,t}$, $c_2\in\{1,...,p\}$, $c_2\neq c_1$, is chosen in Step 3, and the idea is identical to that in Step 2 (although notationally more involved). In particular, given $\widehat{\tau}_{c_1,i}$, $\widehat{\kappa}_{c_1,i}$, $i\in\{1,...,p\}$, $i\neq c_1$, we choose $c_2=l$, such that $l\in\{1,...,p\}$, $l\neq c_1$, where l is the smallest $i\in\{1,...,p\}$, $i\neq c_1$ such that $\widehat{\theta}_{c_1,i}<\min_{j>i,j\neq c_1}\widehat{\theta}^*_{c_1,j}$; alternatively, if there is no i such that the previous condition holds, then $l=\max\{1,...,p\}$ such that $l\neq c_1$.

In general, in Step k, noting that in previous steps $z_{c_1,t},...,z_{c_{k-2},t}$ have been chosen as possible common trends, we choose the possible $(k-1)$th common trend, say $z_{c_{k-1},t}$, $c_{k-1}\in\{1,...,p\}$, $c_{k-1}\neq c_1,...,c_{k-2}$. Given $\widehat{\tau}_{c_1,...,c_{k-2},i}$, $\widehat{\kappa}_{c_1,...,c_{k-2},i}$, $i\in\{1,...,p\}$, $i\neq c_1,...,c_{k-2}$, we choose $c_{k-1}=l\in\{1,...,p\}$, $l\neq c_1,...,c_{k-2}$, where l is the smallest $i\in\{1,...,p\}$, $i\neq c_1,...,c_{k-2}$, such that $\widehat{\theta}_{c_1,...,c_{k-2},i}<\min_{j>i,j\neq c_1,...,c_{k-2}}\widehat{\theta}^*_{c_1,...,c_{k-2},j}$; alter-

natively, if there is no i such that the previous condition holds, then $l = \max\{1, ..., p\}$ such that $l \neq c_1, ..., c_{k-2}$.

Finally we describe the properties of our method of choosing common trends. As clarified before, this choice depends on the ordering of the observables in the vector z_t (with p observables there are $p!$ such alternative orderings). Let $r \in \{1, ..., p-1\}$ be the cointegrating rank (so $p - r$ common trends exist). Given an ordering of the variables (say $o = g$) independent from the observables, let i_1 be the largest number in $\{1, ..., p\}$ such that $z_{i_1,t} \sim I(1)$. Similarly, let i_2 be the largest number in $\{1, ..., p\}$, $i_2 \neq i_1$ such that $z_{i_1,t}$ and $z_{i_2,t}$ are not cointegrated. In general, for $m = 2, ..., p-r$, let i_m be the largest number in $\{1, ..., p\}$, $i_m \neq i_1, ..., i_{m-1}$ such that $z_{i_1,t}, ..., z_{i_{m-1},t}$ and $z_{i_m,t}$ are not cointegrated. Note that for any $l > 1$, i_l depends on $i_1, ..., i_{l-1}$, although we do not reflect this in the notation for simplicity. Also, all the i_l's depend on the ordering $o = g$, and again for simplicity, this is not reflected in the notation. Also, the existence of the i_l's depends on r. In particular, if $r = p - 1$, just i_1 exists; if $r = p-2$, just i_1 and i_2 exist; in general, for $r \in \{1, ..., p-1\}$, just $i_1, i_2, ..., i_{p-r}$ exist.

Theorem 2. Let $r \in \{1, ..., p-1\}$ be the cointegrating rank. Then, for any $1 \leq l \leq p - r$ and $g \in \{1, ..., p!\}$

$$\Pr(c_1 = i_1, ..., c_l = i_l | \, o = g) \to 1, \text{ as } n \to \infty. \tag{19}$$

Remark 3. Theorem 2 implies that, whenever there are common trends, we choose in every step of the procedure a unique valid common trend wpa1. In particular, the first common trend is the $I(1)$ variable labelled with the highest index in the chosen ordering (wpa1). Next, the second common trend is the variable not cointegrated with $z_{c_1,t}$ labelled with the highest index (wpa1), and so on. This result allows us to control the size of our sequential method and derive the neat results in Theorem 3 below. Note also that Theorem 2 is silent about the behaviour of our choice of common trends when there are no such common trends. However, this is not an issue of overriding concern, because, as it will become evident when justifying the properties of the estimator of the rank given in Theorem 3, the behaviour of our choice of common trends when such common trends do not exist (that is when $l > p - r$) is irrelevant for the properties of our estimator of r.

4 Estimation of the cointegrating rank

We present a procedure that leads to an estimator of the cointegrating rank of a vector of possibly cointegrated observable series z_t, and that provides a set of common trends from which a Phillips' triangular system can be immediately designed, therefore leading to straightforward estimation of the cointegrating space. As in Johansen's methodology, we impose that z_t is at most $I(1)$, although extensions to higher orders are possible by generalizing appropriately the methodology presented below. The choice for the maximum integration order of the vector of observables can be justified on a priori knowledge about the series, or on a preliminary analysis of the individual integration orders (by means of standard test procedures, like ADF or PP tests; see, e.g., Phillips and Ouliaris, 1990). As discussed in the previous section, the particular ordering (o) of the observables in the vector z_t plays a role in our procedure. First, choose a particular ordering, say $o = g$, $g \in \{1, 2, ..., p!\}$, where o is independent from the observables. Given this, our procedure is as follows:

Step 1. Define

$$H(1) : \bigcup_{i=1}^{p} H_i, \qquad \overline{H}(1) : \bigcap_{i=1}^{p} \overline{H}_i, \tag{20}$$

so $H(1)$ is rejected if all the individual hypotheses $H_i : \{z_{i,t} \sim I(1)\}$ are rejected, noting that, by Theorem 1, $H(1)$, $\overline{H}(1)$ are equivalent to $r < p$, $r = p$, respectively. This prompts estimation of the cointegrating rank r by

$$\left\{\widehat{r}_{(g)} = p\right\} = \{H(1) \text{ is rejected}\}, \tag{21}$$

where subindex (g) stresses the dependence of $\widehat{r}_{(g)}$ on the chosen ordering. If $H(1)$ is rejected (because, based on the $\widehat{\tau}_i$'s, as a result of performing p hypothesis tests, all H_i are rejected), the procedure finalizes because there is statistical evidence in favour of $r = p$. If $H(1)$ is not rejected, we proceed to the next step.

Step 2. Choose a possible common trend variable $z_{c_1,t}$, $c_1 \in \{1, ..., p\}$ by the rule discussed in Section 3. As explained there, this requires computing statistics $\widehat{\tau}_i$ (which have been already used in Step 1) and $\widehat{\kappa}_i$. Note that $z_{c_1,t}$ is the variable which shows the least evidence of being stationary (possibly after demeaning) by our criterion. Then, given the possible common trend $z_{c_1,t}$, define

$$H(2) : \bigcup_{\substack{i=1 \\ i \neq c_1}}^{p} H_{c_1,i}, \qquad \overline{H}(2) : \bigcap_{\substack{i=1 \\ i \neq c_1}}^{p} \overline{H}_{c_1,i}, \tag{22}$$

where, again, $H(2)$ is rejected if all $H_{c_1,i}$: $\{z_{c_1,t},\, z_{i,t}$ are not cointegrated$\}$, are rejected, i.e., $\widehat{\tau}_{c_1,i} < \nu_2(\alpha)$ for all $i \neq c_1$ (which requires performing $p-1$ hypothesis tests). By Theorem 1, $H(1) \cap H(2)$, $H(1) \cap \overline{H}(2)$, are equivalent to $r < p-1$, $r = p-1$, respectively, which prompts estimating r by

$$\left\{\widehat{r}_{(g)} = p-1\right\} = \{H(1) \text{ is not rejected and } H(2) \text{ is rejected}\}. \tag{23}$$

If $H(2)$ is not rejected, we proceed to the next step.

Step 3. Choose $z_{c_2,t}$, $c_2 \in \{1,...,p\}$, $c_2 \neq c_1$, as in Section 3, so $z_{c_1,t}$, $z_{c_2,t}$ are possible common trends. Again, this choice is based on statistics $\widehat{\tau}_{c_1,i}$ (computed in Step 2) and $\widehat{\kappa}_{c_1,i}$. Heuristically, $z_{c_2,t}$ is the variable which presents the least evidence of being cointegrated with $z_{c_1,t}$ on the basis of the criterion outlined in Section 3. Then, given the possible common trends $z_{c_1,t}$, $z_{c_2,t}$, define

$$H(3): \bigcup_{\substack{i=1 \\ i\neq c_1,c_2}}^{p} H_{c_1,c_2,i}, \qquad \overline{H}(3): \bigcap_{\substack{i=1 \\ i\neq c_1,c_2}}^{p} \overline{H}_{c_1,c_2,i}, \tag{24}$$

and estimate the rank by

$$\left\{\widehat{r}_{(g)} = p-2\right\} = \{H(1), H(2), \text{ are not rejected and } H(3) \text{ is rejected}\}, \tag{25}$$

noting that if $z_{c_1,t}$, $z_{c_2,t}$ are true common trends, then, by Theorem 1, $H(2) \cap H(3)$, $H(2) \cap \overline{H}(3)$, are equivalent to $r < p-2$, $r = p-2$, respectively, which justifies $\widehat{r}_{(g)}$. Again, note that Step 3 requires performing $p-2$ hypothesis tests

In general, for $k = 3,...,p$, we have

Step k. If $H(k-1)$ has not been rejected, noting that in previous steps $z_{c_1,t},...,z_{c_{k-2},t}$ have been chosen as possible common trends, choose $z_{c_{k-1},t}$, $c_{k-1} \in \{1,...,p\}$, $c_{k-1} \neq c_1,...,c_{k-2}$, so $z_{c_1,t},...,z_{c_{k-2},t},z_{c_{k-1},t}$, are possible common trends. Again, this choice is based on statistics $\widehat{\tau}_{c_1,...,c_{k-2},i}$ (computed in Step $k-1$) and $\widehat{\kappa}_{c_1,...,c_{k-2},i}$. Also, as before, $z_{c_{k-1},t}$ is the observable which presents the least evidence of being cointegrated with $z_{c_1,t},...,z_{c_{k-2},t}$. Then define

$$H(k): \bigcup_{\substack{i=1 \\ i\neq c_1,...,c_{k-1}}}^{p} H_{c_1,...,c_{k-1},i}, \qquad \overline{H}(k): \bigcap_{\substack{i=1 \\ i\neq c_1,...,c_{k-1}}}^{p} \overline{H}_{c_1,...,c_{k-1},i}, \tag{26}$$

and set

$$\left\{\widehat{r}_{(g)} = p-k+1\right\} = \left\{H(i),\, i=1,...,k-1, \text{ are not rejected and } H(k) \text{ is rejected}\right\}, \tag{27}$$

and for the step $k=p$, also $\left\{\widehat{r}_{(g)}=0\right\} = \left\{H(i),\, i=1,...,p, \text{ are not rejected}\right\}$. Step k requires performing $p-k+1$ hypothesis tests.

So far, for a given ordering $o=g$, we have derived a procedure to compute $\widehat{r}_{(g)}$ based on a particular choice rule for common trends. Then, given that any ordering could be chosen, our estimator of the rank is given by $\widehat{r} = \sum_{g=1}^{p!} \widehat{r}_{(g)} 1(o=g)$. Note that in practice this estimator is easy to compute, because it amounts to assuming an arbitrary ordering (which cannot depend on characteristics of the observables) and for that ordering (say $o=g$), computing $\widehat{r}_{(g)}$. The properties of $\widehat{r}$ are given in the next theorem.

Theorem 3. Let $r \in \{0,...,p\}$ be the cointegrating rank in z_t. Then the estimator $\widehat{r}$ of r has the property

$$\lim_{n\to\infty} \Pr(\widehat{r}=r) \geq 1-(p-r)\,\alpha, \tag{28}$$

$$\lim_{n\to\infty} \Pr(\widehat{r}=k) \leq \alpha, \quad k=r+1,...,p, \quad r<p, \tag{29}$$

$$\lim_{n\to\infty} \Pr(\widehat{r}=k) = 0, \quad k=0,...,r-1, \quad r>0. \tag{30}$$

Remark 4. Results in Theorem 3 are comparable to those of Theorem 12.3 of Johansen (1995), the only difference being (28), where for $r \leq p-2$ we obtain a smaller lower bound than Johansen (1995). Note however that this bound might not be strict and fits naturally with the upper bound given in (29). Also, when $r=p$, (28) implies that $\lim_{n\to\infty} \Pr(\widehat{r}=r) = 1$. As it is evident from the proof of the theorem, this result holds due to consistency of the tests, noting that when $r=p$ the alternative hypotheses $\overline{H}_i$, $i=1,...,p$, are true. Overestimation of the rank (as in (29)) occurs when in a certain step all tests for no cointegration (or $I(1)$ in the case of Step 1) are rejected, but at least one of the corresponding null hypotheses is true. Then, the probability of the intersection of these events (rejection of all corresponding nulls) is bounded above by the probability of a single event characterized by the rejection of the null when it is true. Asymptotically, this probability is bounded by α, so the probability of the intersection is also bounded by α (as $n\to\infty$), although this bound might not be tight (see, e.g., (49), (50)). Finally, (30) indicates that, asymptotically, our estimator never underestimates the rank. Again, this property rests on the consistency of the tests. As the proof of Theorem 3 shows, to underestimate the rank at least one of the

tests needs not to reject under the alternative and, due to consistency, the probability of this event vanishes asymptotically.

Remark 5. Nicely, $\widehat{r}$ does not impose that the chosen common trends are true ones. The choice of common trends is stochastic, and this is taken into account when analyzing the properties of $\widehat{r}$. However, note that the results in Theorem 3 are identical to those corresponding to an alternative (infeasible) estimator of r which bases every step of the procedure on true common trends. The reason for this appealing result is that, in every step, if there are valid common trends to be chosen, our method leads to a choice of valid common trends wpa1.

Remark 6. Our procedure is sequential, and the main challenge in the justification of Theorem 3 is to deal appropriately with this sequentiality without imposing infeasible assumptions about true common trends in the different steps. Note that this sequentiality is, in fact, also an issue in Johansen's procedure, where the hypotheses

$$F(j) : \{\text{the cointegrating rank is smaller or equal than } j\}, \; j = 0, ..., p, \tag{31}$$

are tested sequentially starting by $F(0)$. If $F(0)$ is rejected, $F(1)$ is carried out, if $F(1)$ is rejected, $F(2)$ is carried out and so on. Thus, for example if $F(0)$, $F(1)$, are rejected but $F(2)$ is not, the rank is estimated as $\widehat{r} = 2$. Here, the sequentiality problem is the following. Suppose that $F(0)$ is rejected, so $F(1)$ is tested. The asymptotic distribution on which the test for $F(1)$ is based, assumes that $F(0)$ is false, so under $F(1)$, $r = 1$ if $F(0)$ is false. However, it might be that the true state of nature is $r = 0$. Note that $F(0)$ is rejected with asymptotic probability α even if $F(0)$ is true, so, when performing the test for $F(1)$, one might be using a limiting distribution which is not the true one under the true state of nature ($r = 0$). This issue, of course, arises in every step of the procedure. Johansen (1995) indicates that simulations show that the distributions for smaller ranks than the maximum allowed in $F(j)$ are shifted towards smaller values (than if $r = j$), and, hence, they are not relevant for calculating the appropriate p-values. However, this is not formal evidence. In contrast, our method guarantees the use of a correct limiting distribution under the null in every step, but, of course, there is a price to pay for this: given that we conclude the procedure when one of $H(1)$, $H(2)$,..., is rejected, there is an α probability at each step of making the wrong decision, and these α's accumulate, as we show in (28). Thus, we provide a solution to the sequentiality problem at cost of providing a smaller lower bound for $\Pr(\widehat{r} = r)$ than Johansen $(1 - (p - r)\alpha$ instead of $1 - \alpha)$.

Remark 7. Theorem 3 indicates that, asymptotically, a very small α would be

desirable. In fact, it can be shown that letting $\alpha = \alpha_n \to 0$ with an adequate rate (accommodating α_n to the divergence rate of the corresponding test statistic under the alternative), then $\Pr(\widehat{r} = r) \to 1$ for any r. However, in practice, one needs to choose critical values, and choosing α_n such that $\alpha_n \to 0$ does not help much in finite samples. It might be a nice theoretical artifact (leading asymptotically to the best result), but the finite sample properties of such approach will be uncertain. Intuitively, the method might benefit in finite samples from using large (small) α's when r is large (small). However, a formal analysis of the linkage between the choice of α and the finite sample performance of $\widehat{r}$ appears to be very involved. We have provided some Monte Carlo evidence in Section 5 which nevertheless exemplifies the previous intuition.

Remark 8. Once the cointegrating rank has been estimated, our procedure leads to a set of (cointegrating) regressions from which estimation of the cointegrating space would be straightforward. Letting $r \in \{1, ..., p-1\}$, noting (4), (5), define $\beta' = P - BQ$, so the r columns of β span the cointegrating space. Then this space can be estimated by simple methods such as ordinary least squares (OLS), fully-modified OLS or Dynamic OLS applied to (4), choosing Pz_t (Qz_t), as left (right)-hand side variables, obtaining $\widehat{B}$, and $\widehat{\beta}' = P - \widehat{B}Q$. In practice, if in Step $p - r + 1$ we determine that $\widehat{r} = r$, $r = 1, ..., p-1$, our method will also necessarily determine one set of $p - r$ variables in z_t which, in view of the corresponding test statistic, appear not to be cointegrated. Thus, the choice of Qz_t is stochastic (data-based), but simple and automatic. Note that in the particular case where $r = 1$, the unique cointegrating vector can be estimated by a single regression equation (special case of a triangular form), where the $p - 1$ right-hand side variables should be those chosen as common trends in Step p. Note that for a given set of p observables it is only in this final step where a cointegrating regression with $p-1$ regressors might be well specified (in the sense that the regressors are not cointegrated), although it seems common among practitioners to initiate an analysis of cointegration by a regression of this type.

Our procedure, therefore, leads to an estimate of the rank r and a choice of $p - r$ common trends from which a Phillips triangular form can then be estimated. This implies imposing just-identifying restrictions on the cointegration vectors (see ex. 2 in Boswijk, 1996, or the discussion in Luukkonen et al., 1999, which both mention the identification scheme implicit in a triangular form), noting that our particular identification scheme is valid if we estimate correctly the true r and our chosen common trends are true ones (which is the case wpa1 if $\widehat{r} = r$). The validity of our identification restrictions can be checked by the procedures of Boswijk (1996) or

Luukkonen et al. (1999) (which provide tests for the validity of identifying restrictions of a more general form), or by the tests for the rank of a cointegration submatrix of Kurozumi (2005) and Paruolo (2006). Note, however, that the validity of all these procedures relies on the knowledge of the cointegrating rank.

Remark 9. An important issue is whether our approach is useful to test structural restrictions in the sense of Johansen and Juselius (1992). These restrictions describe long-run relations in which economic structural hypotheses are usually formulated, and take the form of linear restrictions on the cointegrating matrix β. For the sake of simplicity let $x_t = (P', Q')\, z_t$, so x_t is the reordered vector of observables where the last $p - r$ components of x_t are the common trends. Related to this reordering, the corresponding cointegrating matrix is $\theta = (A, B)'$, where A and B (defined in (4)) are $r \times r$ and $r \times (p - r)$ matrices, respectively, so that θ is a reordered version of β. In terms of θ, our just-identifying restrictions are $A = I_r$ (the r-rowed identity matrix).

We particularize our discussion to the $\mathcal{H}_4$ type of hypothesis of Johansen and Juselius (1992), although other types could be also considered. This hypothesis is the following:

$$\mathcal{H}_4 : \theta = H_4 \varphi, \tag{32}$$

where H_4 is a $p \times s$ known matrix, φ is a $s \times r$ unrestricted matrix and $r \leq s \leq p$. This hypothesis imposes the same $p - s$ restrictions on all r cointegrating vectors, so it could have been equally expressed as $R\theta = 0$, where R is a corresponding full rank $(p - s) \times p$ matrix. Partition R according to θ, that is $R = (R_A, R_B)$, where R_A, R_B are $(p - s) \times r$, $(p - s) \times (p - r)$, respectively. First, we show that restrictions for which $rank\,(R_B) < p - s$ are not compatible with our just-identifying restrictions. In this respect note that, as implied by Theorem 2, if the chosen rank is the true one, our procedure determines a correct choice of common trends wpa1. Thus if our choice of common trends is correct, our just-identifying restrictions are valid, so restrictions for which $rank\,(R_B) < p - s$ are of little interest because they cannot hold. The particular sense in which such restrictions are not compatible with the just-identifying ones is that if $rank\,(R_B) < p - s$, the observables in Qz_t are cointegrated (so Qz_t cannot be common trends). Let us justify this. If $rank\,(R_B) < p - s$, there exists a $(p - s) \times 1$ vector $\varsigma \neq 0$ such that $\varsigma' R_B = 0$, noting also that $\varsigma' R_A \neq 0$ (otherwise R would not be full rank). If $R\theta = 0$, then

$$R_A A' + R_B B' = 0 \Rightarrow \varsigma' R_A A' = 0, \tag{33}$$

so A would be singular and by Johansen (2005, p.98) the observables in Qz_t are

cointegrated.

Thus, if we focus on hypotheses for which $rank\,(R_B) = p - s$, imposing $A = I_r$, then

$$R_A A' + R_B B' = 0 \Rightarrow R_B B' = -R_A \tag{34}$$

so the structural restrictions $R\theta = 0$ can be straightforwardly formulated as overidentifying restrictions on B, which once B is estimated can be easily tested by well-known methods (e.g., Wald test), see, e.g. Saikkonen (1993).

5 Monte Carlo evidence

We investigate the finite sample performance of our procedure (denoted GBH) and compare it to that of Johansen's methodology. Our main analysis is based on generating systems of four observable series $z_{i,t}$, $i = 1, ..., 4$, of lengths $n = 50$, 100, 200, 500, 1000. First, for $j = 1, ..., 4,$, we generate an $I\,(0)$ vector v_t corresponding to two different data generating processes (DGP) given by

$$\text{DGP1: } v_{j,t} = \varepsilon_{j,t}; \text{ DGP2: } v_{j,t} = 0.8 v_{j,t-1} + \varepsilon_{j,t}; \tag{35}$$

where ε_t is a normal white noise innovation with $Var(\varepsilon_t) = I_4$. Then, for cointegrating ranks $r = 1, 2, 3$, the observables are generated as

$$z_{j,t} = \sum_{k=j+1}^{4} z_{k,t} + v_{j,t}, \quad j = 1, ..., r, \qquad \Delta z_{j,t} = v_{j,t}, \quad j = r+1, ..., 4. \tag{36}$$

We also cover cases $r = 0$, where $\Delta z_{j,t} = v_{j,t}$, $j = 1, 2, 3, 4$, and $r = 4$, where $z_t = v_t$. Once the observables are generated, they are randomly reordered and both our procedure and Johansen's are applied to this reordered vector.

We present in Tables 1-4 results for the estimated ranks provided by our method (Tables 1 and 3, where all individual tests in the $H(j)$ hypotheses are ADF) and by the trace test proposed by Johansen (Tables 2 and 4; see, e.g., Johansen, 1995) for different sample sizes, significance levels and values of r (results for the maximum eigenvalue test of Johansen, 1995a, are available from the authors upon request. They are not presented here because in all cases this procedure is outperformed by the trace test). The number of lags in the ADF tests and in Johansen's VAR are chosen automatically using the BIC. The tables show the proportion of the 10,000 replications leading to each estimated rank, so that the blocks in the main diagonals represent the percentage of "correct answers", whereas the rest of blocks indicate

proportions of "mistakes".

The results for DGP1 are quite straightforward. The performance of our procedure matches the theoretical properties in Theorem 3. First, when $r = p = 4$ the test correctly estimates the rank in 100% cases (for $n \geq 200$), showing evidence of (28). As said before, this is a consequence of consistency of the ADF tests. (i.e., when $p = 4$ an estimated rank of $\widehat{r} = 4$ comes from rejection of $H(1)$, which involves rejecting the four unit root tests on the four observables). In cases where $\widehat{r} < p$ (so some $H(i)$ hypotheses were not rejected), then the rank is correctly estimated with frequencies larger than the bound given in (28), the worst results corresponding to the $r = 0$ case, as might have been expected in view of (28). In any case, the results suggest that the bound in (28) is conservative. In Table 2 we present the corresponding results from Johansen's trace test procedure. Compared to our method, Johansen's procedure performs better in almost all cases, as it was to be expected: it is based on the true likelihood of the data and, even though the lag length is chosen via an information criterion, it can never be underestimated. Thus, Johansen's procedure here is based on a correctly specified model or, in the worst scenario, on a model which is only slightly overparameterized.

The results for DGP2 (Tables 3 and 4) are more interesting. The main challenge of DGP2 is that the stationary roots are now close to unity. We expect that the first step of our procedure (single unit root tests) may have difficulties, given the well known low power of unit root tests against "near unit roots." This is indeed the case: when $r = 4$ our procedure needs many observations ($n \geq 200$) to produce appropriate results (as a result of rejecting the unit root hypothesis in all four observables and estimate the correct rank). On the other hand, when the true rank is not full ($r < 4$) it is less likely that the procedure incorrectly rejects $H(1)$ and selects $\widehat{r} = 4$ (compare the case $r = 3$, $\widehat{r} = 4$ in Tables 1 and 3). In a sense, the near unit root variables help in keeping with the unit root null in the first steps of the procedure. The drawback is that the near unit roots tend to prevent rejection in further steps so that our procedure underestimates the cointegrating rank more frequently. These results are also related to the choice of significance level α (see Remark 7), which generates a trade-off between underestimation and overestimation of the rank: a low α leads to less frequent rejections of the null hypotheses of the ADF tests and thus to lower frequencies of rank overestimation. Alternatively, the frequencies of rank underestimation, coming from not rejecting the null hypotheses too often, are much larger. In any case, overall, in terms of proportions of correct answers, our procedure performs strictly better (worse) than Johansen's in 61 (7) out of 75 cases, which

is an encouraging result. In particular, for systems with high cointegrating rank, our procedure tends to outperform Johansen's for all sample sizes, but compared to DGP1, our method performs now slightly better even in the cases of intermediate and low true cointegration ranks. Overall, Johansen's method performs well, as it should be, given that it is, again, based on a correct likelihood, although for small n its performance is poor given that underspecification is possible now.

Next, we repeat the above simulations allowing for the innovations to be correlated (see Toda, 1995, for a discussion of how such correlation may affect the performance of LR cointegration tests). Tables 5, 6 show the frequencies with which the two procedures estimate the correct rank ($\widehat{r} = r$) for different values of the covariance of the innovations. In particular, we let $Cov(\varepsilon_{j,t},\varepsilon_{i,t})$, $j \neq i$, be 0.4 (Panel A), 0.8 (Panel B) and -0.3 (Panel C). The performance of both procedures worsens slightly in the white noise case (DGP 1), but, still, correct rank detection approaches the theoretical properties of the estimators, and Johansen's procedure tends to outperform our proposal. For the autoregressive case (DGP 2), our proposal behaves similarly to the uncorrelated case for all values of correlation. However, in line with the results in Toda (1995), the performance of Johansen's procedure worsens quite significantly in the high and negative correlation case for low values of r (so the procedure overestimates the rank). This suggests that in the case of "near unit roots" the correlation between innovations makes Johansen's procedure overestimate the extent of cointegration more frequently than our procedure does.

We also used an alternative DGP with moving average innovations (DGP3: $v_{j,t} = \varepsilon_{j,t} - 0.5\varepsilon_{j,t-1}$). We do not present these results (which are available upon request): the two procedures work reasonably well, although the performance of both is worse than for DGP1 and DGP2 (note that the error dynamics cannot be represented as a simple finite-length autoregresive process). In the uncorrelated case and with negative correlation (-0.3), our procedure outperforms Johansen's for small sample sizes and for systems with high cointegration rank (except when $r = p$, where both procedures consistently estimate the true rank). With positive correlation, our procedure overestimates the rank more often than Johansen's.

Overall, our method is comparable to Johansen's in systems with high r and it may be superior in some cases when the DGP departs from pure white noise and, therefore, lag lengths can be misspecified. Given that this is a relevant empirical situation, the above results suggest that our proposal may be a useful alternative to Johansen's in contexts with expected high cointegration ranks compared to the number of variables. This is the case, for example, of the solutions of some macroeconomic models that are

a function of a few integrated shocks or factors (our empirical application below is an example of such a setting). Additionally, in high-dymensional systems the number of parameters in the short-run dynamics may be quite large, and there may be some advantage to a procedure, such as ours, that does not require to specify these short-run dynamics. Note also that even in large systems, if the cointegrating rank is large, our method is based on tests that involve a small number of variables, which might simplify matters. In order to provide some evidence of this latter situation, we compare in Table 7-Panel A the performance of the two procedures when facing high-dimensional systems. In particular, we generate systems of ten variables ($p = 10$) which contain a mixture of $I(1)$ and $I(0)$ variables. The possible settings here are numerous, so we show the results of only a few combinations of p_1 (number of $I(1)$ variables), r_1 (the cointegration rank among the $I(1)$ variables) and p_0 (number of $I(0)$ variables). The cointegration rank of each system is $r = r_1 + p_0$ (so $r_1 = r - (p - p_1)$). The $I(1)$ observables are constructed as before using DGP2, and the $I(0)$ variables are generated as $z_{j,t} = v_{j,t} = 0.8v_{j,t-1} + \varepsilon_{j,t}$. In all cases $Var(\varepsilon_t) = I_{10}$. We generate samples of $n = 500$, 1000 observations and show the results (for tests with $\alpha = 0.05$) of correct rank estimation (i.e., we have only tabulated the frequencies for which $\widehat{r} = r$). As seen from the table, our procedure outperforms Johansen's in almost all cases, which is quite remarkable, noting that in our procedure the number of lags in the ADF tests is chosen using BIC, whereas for Johansen's method we have used the true lag of one. This is a consequence of a moderate sample size compared to the number of parameters to be estimated, but note that the employed sample sizes are larger than those in typical macroeconomic settings.

Finally, we compare our procedure with that of Johansen in two situations where, as described by Gonzalo and Lee (1998), regression-based methods might be more robust to misspecification, since the LR type tests tend to conclude in favour of cointegration even when the variables are not cointegrated ("spurious cointegration"). The first situation corresponds to processes with autoregressive roots marginally larger than one. We generate four observable series $z_{j,t} = 1.01z_{j,t-1} + \varepsilon_{j,t}$, $j = 1, 2, 3, 4$, where ε_t is generated as before with $Var(\varepsilon_t) = I_4$. The observables are independent, and the system, therefore, is close to one with rank $r = 0$. The second misspecification concerns fractional processes. We generate four observables $z_{j,t} = (1 - L)^{-1.4} \{\varepsilon_{j,t} 1(t > 0)\}$, $j = 1, 2, 3, 4$, where $1(\cdot)$ is the indicator function and ε_t is generated as before. Again, the true cointegration rank is $r = 0$. Results are reported in Table 7 (Panel B.1: roots close to unity; Panel B.2: fractional processes), where we only tabulate frequencies for which $\widehat{r} = 0$ (correct rank estimation) and

$\widehat{r} = 1$ ("spurious cointegration"). As anticipated by Gonzalo and Lee (1998), our method tends to indicate that $r = 0$ in both scenarios, whereas the outcome of Johansen's trace test is unclear, with a substantial number of cases indicating "spurious cointegration".

6 Empirical example

We illustrate our procedure with two examples in the context of the US term structure, using data which have already been analyzed in, for example, Diebold and Li (2006), Giese (2008) and, originally, in Hall et al. (1992). We call $i_{m,t}$ the yield of a zero-coupon bond of maturity m (hereafter "the interest rate of maturity m"), and define the term premium for that maturity as L_t^m. Under a no-arbitrage condition (see, e.g., Giese, 2008, for a derivation) the spread between a rate of any maturity and the short-term rate $i_{1,t}$ must be

$$i_{m,t} - i_{1,t} = \frac{1}{m} \sum_{j=1}^{m-1} (m-j) E_t \left(\Delta i_{1,t+j} \right) + L_t^m. \tag{37}$$

Since $i_{m,t}$ are typically considered to be $I(1)$ variables, the first term in the right hand side is stationary and if the term premium L_t^m is stationary, then interest rates of different maturities should cointegrate with the short-term rate, with cointegration vector $(1, -1)'$. Hence, all pairs of rates would be cointegrated (all term spreads would be stationary) and one interest rate (for example, $i_{1,t}$) could be taken as the common trend or, in the terminology used in the literature, as a common factor that would drive the level of the term structure. If, however, there are other common trends present (i.e., there are additional common factors in the term structure), then some term premia L_t^m may be $I(1)$ and not all term spreads would be stationary. The results in Diebold and Li (2006) and Giese (2008) suggest that short-term rates mostly depend on one factor (level) and would therefore be cointegrated with a particular short-rate that would act as the common trend (so short-term spreads are stationary); however, if we include longer maturities in the analysis, then additional factors (slope, curvature) could be present and further common trends would be needed. In our terminology, the first situation would imply that in a set of p interest rates of short maturities there should be $p-1$ cointegrating vectors and one common trend, whereas if we mix short and long maturities we could observe r cointegrating vectors and $p - r = 2, ..$ common trends. We use two subsets of interest rates from the database of different maturities constructed by Diebold and Li (2006) from US Treasuries.

Example 1 (Ex1 hereafter) focuses only on short rates and uses five series ($p = 5$) in the short end of the yield curve (maturities $m = 1, 3, 9, 12, 15$ months). Example 2 (Ex2 hereafter) selects a mixture of maturities ($m = 1, 3, 18, 48, 120$ months). In both cases we use monthly observations from January 1985 to December 2000 for a total of 192 observations. Evidence of structural breaks in the data prior to 1985 has been documented (see, e.g., Diebold and Li, 2006), so we opted for keeping the sample used by Diebold and Li (2006) and Giese (2008).

We examine cointegration among these two sets of interest rates by applying our procedure and Johansen's methodology. The five series in each example are ordered by maturity. We assume that the maximum integration order of the system is one (based on ADF tests for the first-differenced series, which provide very strong evidence against $I(2)$) and start our procedure by Step 1. Table 8 presents the sequences of tests involved. Based on the initial ADF tests $\widehat{\tau}_m$, $H(1)$ cannot be rejected in either example, so there is evidence of $p < r$ in both systems (because at least one interest rate appears to be $I(1)$). We then select the first common trend of the systems. As it is evident from Table 8, none of the conditions $\widehat{\theta}_1 < \min_{j>1} \widehat{\theta}_j^*$, $\widehat{\theta}_2 < \min_{j>2} \widehat{\theta}_j^*$, $\widehat{\theta}_3 < \min_{j>3} \widehat{\theta}_j^*$, $\widehat{\theta}_4 < \widehat{\theta}_5^*$ hold in either system, so our procedure chooses $z_{c_1,t} = i_{15,t}$ in Ex1 and $z_{c_1,t} = i_{120,t}$ in Ex2 as the first common trend. We next check whether all pairs of rates ($i_{m,t}$,$i_{15,t}$) $m = 1, 3, 9, 12$, (Ex1) and all pairs of rates ($i_{m,t}, i_{120,t}$) $m = 1, 3, 18, 48$ (Ex2) are cointegrated. This is done via ADF residual-based tests for cointegration (reported in Table 8 as $\widehat{\tau}_{c_1,m}$), where, following previous studies, we include a constant but no deterministic trend in the cointegrating regressions. Strong evidence of cointegration between all pairs of rates is detected in Ex1 but not in Ex2, where none of the four pairs shows evidence of cointegration. We therefore reject $H(2)$ for Ex1 and stop the testing procedure, concluding that $\widehat{r} = 4$. For Ex2, $H(2)$ is not rejected, so we move on to Step 3, where we choose a second common trend and then test $H(3)$, i.e., test whether groups of three interest rates are cointegrated. Given $\widehat{\theta}_{c_1,m}$, $\widehat{\theta}_{c_1,m}^*$, the choice for the second common trend is $i_{3,t}$. The test of $H(3)$ is, therefore, based on residual-based ADF statistics $\widehat{\tau}_{c_1,c_2,m}$ (last line of Table 8): all tests strongly reject the null of no cointegration, so $H(3)$ is rejected and we conclude the procedure with estimated rank $\widehat{r} = 3$ for Ex2.

Once the ranks have been estimated, cointegrating vectors can be easily estimated by regression-based methods where we choose $i_{15,t}$ as regressor (Ex1) and $i_{120,t}$ and $i_{3,t}$ as regressors (Ex2), the rest being left hand side variables. We estimate the systems

by OLS

$$\begin{aligned}(i_{1,t}, i_{3,t}, i_{9,t}, i_{12,t})' &= A_1 + B_1 i_{15,t} + w_{1,t} \quad \text{(Ex1)}, \\ (i_{1,t}, i_{18,t}, i_{48,t})' &= A_2 + B_2 (i_{3,t}, i_{120,t})' + w_{2,t} \quad \text{(Ex2)}, \end{aligned} \tag{38}$$

where $A_1 = (a_1, a_3, a_9, a_{12})'$, $B_1 = (b_1, b_3, b_9, b_{12})'$, $A_2 = (a_1, a_{18}, a_{48})'$, B_2 is a (3×2) matrix of coefficients and $w_{1,t}$ and $w_{2,t}$ are generic stationary processes. The results are shown in Table 9. We also report results of Dynamic-OLS (Stock and Watson, 1993), where four lags and leads of the first differences of the common trends are included in the regressions (other choices of lags and leads yield very similar results). These latter estimators of B_1 and B_2 have a mixed-normal asymptotic distribution, so standard inference rules can be applied. The results of OLS and DOLS are similar, and the estimated cointegrating vectors are quite intuitive: in Ex1 the cointegrating coefficients b_m increase with m; in Ex2 the longer m, the more the interest rate loads on the long-term rate ($i_{120,t}$). An alternative specification for Ex2, more in the line of the structure in Giese (2008), defines the common trends as $i_{120,t}$ (which would account for the level of the term structure) and $i_{120,t} - i_{3,t}$ (which would account for the slope). This specification would be equivalent to a reparameterization of our results.

Johansen's methodology applied to both examples yields the same estimates of the rank, $\widehat{r} = 4$ in Ex1 and $\widehat{r} = 3$ in Ex2 (the test procedure is well known, so we omit these results and offer them upon request), and the estimated coefficients are very similar to DOLS (we have included them in Table 9, normalized to be directly comparable to those of our cointegrating regressions).

7 Conclusions and final comments

We have presented a simple procedure to infer the cointegrating rank and, subsequently, estimate the cointegrating vectors in system frameworks. This method requires neither the imposition of a priori identifying conditions nor a parametric specification for the short run components of the model. Our proposal appears to be a simple alternative to Johansen's likelihood-based methodology, since only standard regression-based techniques are necessary, and it leads to feasible estimation of Phillips' triangular representations of cointegrated systems, providing guidance for the correct choice of left and right hand side variables. In terms of finite sample performance, our results are satisfactory and comparable to those provided by alter-

native procedures, like Johansen's maximum eigenvalue and trace tests. In addition, the outcomes of our procedure appear to be robust to some possible misspecifications and the procedure seems to work quite satisfactorily for high dimensional systems.

Our procedure places emphasis on the long-run structure of the data. If one were specifically interested in the short-run dynamics of model (1), these could be subsequently estimated by imposing parametric assumptions on u_t. The unknown cointegrating parameters in Υ would have to be replaced by the corresponding estimates and therefore the estimation of the short run structure of u_t would be based on cointegrating residuals and first differences of the common trends. It can easily be shown that the first order asymptotic properties of the estimates of the short run parameters are unaffected by using either the true cointegrated errors or the corresponding residuals.

Finally, the procedure can be generalized to $I(2)$ systems. The extension to cover more general forms of integration and cointegration (like fractional ones) is also possible, but this is rather more involved and will be the object of future research.

APPENDIX. PROOFS OF THEOREMS

Proof of Theorem 1. First, suppose that the cointegrating rank is r, so there exists a full rank $p \times r$ matrix β such that $\beta' z_t$ is stationary. Without loss of generality, we assume that the elements of z_t are ordered in such a way that $\beta = (\beta'_A, \beta'_B)'$, where β_A, β_B are $r \times r$, $(p-r) \times r$ matrices respectively, and β_A is nonsingular. Partition $z_t = (z'_{(A)t}, z'_{(B)t})'$ accordingly, so $z_{(A)t}$, $z_{(B)t}$ are $r \times 1$, $(p-r) \times 1$ vectors respectively. Then, we know that the elements of $z_{(B)t}$ are individually $I(1)$ and not cointegrated (see, e.g., Johansen, 2005, p.98), so a. holds.

Next, given that linear combinations of cointegrating vectors are also cointegrating, the columns of $\beta\beta_A^{-1} = \left(I_r, \left(\beta_A^{-1}\right)' \beta'_B\right)'$ are cointegrating vectors, so b. follows.

On the other hand, suppose that a. and b. hold. First, setting $z_t = (z'_{(a)t}, z'_{(b)t})'$, where $z_{(a)t}$ collects the individual components of z_t not collected by $z_{(b)t}$, there exists a $(p-r) \times r$ matrix A such that the columns of $\beta = (I_r, A')'$ are cointegrating vectors, so the cointegrating rank is larger or equal than r. If the rank is larger than r, there exists a vector $\theta = (\theta'_a, \theta'_b)'$, where θ_a is $r \times 1$ and θ_b is $(p-r) \times 1$, such that the columns of the $p \times (r+1)$ matrix (β, θ) are linearly independent and $\theta_a \neq 0$ (otherwise a. would not hold). However the vector

$$\begin{pmatrix} I_r \\ A \end{pmatrix} (-\theta_a) + \begin{pmatrix} \theta_a \\ \theta_b \end{pmatrix},$$

is nonzero by linear independence, and it is also cointegrating, which contradicts a. Thus, the rank must be r, to conclude the proof of the theorem.

Proof of Theorem 2. Let $\Pr(\cdot|\, o=g) = \Pr_g(\cdot)$. Then the proof follows by induction by showing: (i) $\Pr_g(c_1 = i_1) \to 1$ as $n \to \infty$; (ii) If for $r < p-1$ and $l = 2, ..., p-r$, $\Pr_g\left(\cap_{k=1}^{l-1}\{c_k = i_k\}\right) \to 1$ as $n \to \infty$, then $\Pr_g\left(\cap_{k=1}^{l}\{c_k = i_k\}\right) \to 1$ as $n \to \infty$.

First, we show (i). Let $i_1 = 1$. Then

$$\Pr_g(c_1 = 1) = \Pr_g\left(\bigcap_{j=2}^{p}\left\{\widehat{\theta}_1 < \widehat{\theta}_j^*\right\}\right) \geq \sum_{j=2}^{p}\Pr_g\left(\widehat{\theta}_1 < \widehat{\theta}_j^*\right) - (p-2).$$

Then the required result follows on showing that for all $j = 2, ..., p$,

$$\Pr_g\left(\widehat{\theta}_1 < \widehat{\theta}_j^*\right) \to 1 \text{ as } n \to \infty. \tag{39}$$

Clearly, $\Pr_g\left(\widehat{\theta}_1 < \widehat{\theta}_j^*\right) = \Pr_g\left(\widehat{\tau}_1^2 < \widehat{\tau}_j^2 \min\{1, \widehat{\kappa}_j\}\right)$, so (39) holds because by (8) $\widehat{\tau}_1^2 = O_p(1)$, whereas by (10), (18), $\widehat{\tau}_j^2 \min\{1, \widehat{\kappa}_j\}$ diverges to ∞, noting also that these statistics are independent from o.

Next, let $1 < i_1 < p$. Then

$$\begin{aligned}\Pr_g(c_1 = i_1) &= \Pr_g\left(\bigcap_{j=1}^{i_1-1}\{c_1 \neq j\}, \bigcap_{j=i_1+1}^{p}\left\{\widehat{\theta}_{i_1} < \widehat{\theta}_j^*\right\}\right) \geq \sum_{j=1}^{i_1-1}\Pr_g(c_1 \neq j) \\ &\quad + \sum_{j=i_1+1}^{p}\Pr_g\left(\widehat{\theta}_{i_1} < \widehat{\theta}_j^*\right) - (p-2).\end{aligned}$$

Then, the required result follows on showing

$$\Pr_g(c_1 \neq j) \to 1 \text{ as } n \to \infty, \text{ for all } j = 1, ..., i_1 - 1, \tag{40}$$

$$\Pr_g\left(\widehat{\theta}_{i_1} < \widehat{\theta}_j^*\right) \to 1 \text{ as } n \to \infty, \text{ for all } j = i_1 + 1, ..., p. \tag{41}$$

First, for $j = 1, ..., i_1 - 1$,

$$\Pr_g(c_1 \neq j) = \Pr_g\left(\bigcup_{k=j+1}^{p}\left\{\widehat{\theta}_j \geq \widehat{\theta}_k^*\right\}\right) \geq \Pr_g\left(\widehat{\theta}_j \geq \widehat{\theta}_{i_1}^*\right) = \Pr_g\left(\widehat{\tau}_j^2 \geq \widehat{\tau}_{i_1}^2 \min\{1, \widehat{\kappa}_{i_1}\}\right) \to 1,$$

as $n \to \infty$, because by (8) $\widehat{\tau}_{i_1}^2 = O_p(1)$, by (17) $\widehat{\kappa}_{i_1} = o_p(1)$, whereas by (8) or (10) $\widehat{\tau}_j^2$ either has an exact rate $O_p(1)$ or it diverges to ∞.

Next, for $j = i_1+1, ..., p$, $\Pr_g\left(\widehat{\theta}_{i_1} < \widehat{\theta}^*_j\right) = \Pr_g\left(\widehat{\tau}^2_{i_1} < \widehat{\tau}^2_j \min\left\{1, \widehat{\kappa}_j\right\}\right)$, so (41) holds because by (8) $\widehat{\tau}^2_{i_1} = O_p(1)$, whereas by (10), (18), $\widehat{\tau}^2_j \min\left\{1, \widehat{\kappa}_j\right\}$ diverges to ∞.

Finally, let $i_1 = p$. Then

$$\Pr_g(c_1 = p) = \Pr_g\left(\bigcap_{j=1}^{p-1}\{c_1 \neq j\}\right) \geq \sum_{j=1}^{p-1}\Pr_g(c_1 \neq j) - (p-2),$$

so the result follows by showing

$$\Pr_g(c_1 \neq j) \to 1 \text{ as } n \to \infty, \text{ for all } j = 1, ..., p-1,$$

which holds by identical reasons to (40), to conclude the proof of (i).

Next we show (ii). Let $\underline{a}_l$ ($\overline{a}_l$) be the smallest (largest) number within $\{1, ..., p\}$ such that $\underline{a}_l, \overline{a}_l \neq i_1, ..., i_{l-1}$. First, let $i_l = \underline{a}_l$. Then

$$\begin{aligned}
\Pr_g\left(\bigcap_{k=1}^{l}\{c_k = i_k\}\right) &= \Pr_g\left(\bigcap_{k=1}^{l-1}\{c_k = i_k\}, \bigcap_{\substack{j=i_l+1 \\ j\neq i_1,...,i_{l-1}}}^{\overline{a}_l}\left\{\widehat{\theta}_{i_1,...,i_{l-1},i_l} < \widehat{\theta}^*_{i_1,...,i_{l-1},j}\right\}\right) \\
&\geq \Pr_g\left(\bigcap_{k=1}^{l-1}\{c_k = i_k\}\right) + \Pr_g\left(\bigcap_{\substack{j=i_l+1 \\ j\neq i_1,...,i_{l-1}}}^{\overline{a}_l}\left\{\widehat{\theta}_{i_1,...,i_{l-1},i_l} < \widehat{\theta}^*_{i_1,...,i_{l-1},j}\right\}\right) - 1
\end{aligned}$$

so (ii) follows by showing that for any $j > i_l$, $j \neq i_1, ..., i_{l-1}$, $\Pr_g\left(\widehat{\theta}_{i_1,...,i_{l-1},i_l} < \widehat{\theta}^*_{i_1,...,i_{l-1},j}\right) \to 1$ as $n \to \infty$. Clearly

$$\Pr_g\left(\widehat{\theta}_{i_1,...,i_{l-1},i_l} < \widehat{\theta}^*_{i_1,...,i_{l-1},j}\right) = \Pr_g\left(\widehat{\tau}^2_{i_1,...,i_{l-1},i_l} < \widehat{\tau}^2_{i_1,...,i_{l-1},j} \min\left\{1, \widehat{\kappa}_{i_1,...,i_{l-1},j}\right\}\right),$$

so, noting that $z_{i_1,t}, ..., z_{i_{l-1},t}, z_{j,t}$, are cointegrated, the result holds because by (8) $\widehat{\tau}^2_{i_1,...,i_{l-1},i_l} = O_p(1)$, whereas by (10), (18), $\widehat{\tau}^2_{i_1,...,i_{l-1},j} \min\left\{1, \widehat{\kappa}_{i_1,...,i_{l-1},j}\right\}$ diverges to ∞.

Next, if $\underline{a}_l < i_l < \overline{a}_l$,

$$\begin{aligned}
&\Pr_g\left(\bigcap_{k=1}^{l}\{c_k = i_k\}\right) \\
= &\Pr_g\left(\bigcap_{k=1}^{l-1}\{c_k = i_k\}, \bigcap_{\substack{j=\underline{a}_l \\ j\neq i_1,...,i_{l-1}}}^{i_l-1}\{c_l \neq j\}, \bigcap_{\substack{j=i_l+1 \\ j\neq i_1,...,i_{l-1}}}^{\overline{a}_l}\left\{\widehat{\theta}_{i_1,...,i_{l-1},i_l} < \widehat{\theta}^*_{i_1,...,i_{l-1},j}\right\}\right) \\
\geq &\Pr_g\left(\bigcap_{k=1}^{l-1}\{c_k = i_k\}\right) + \Pr_g\left(\bigcap_{\substack{j=\underline{a}_l \\ j\neq i_1,...,i_{l-1}}}^{i_l-1}\{c_l \neq j\}\right) \\
&+ \Pr_g\left(\bigcap_{\substack{j=i_l+1 \\ j\neq i_1,...,i_{l-1}}}^{\overline{a}_l}\left\{\widehat{\theta}_{i_1,...,i_{l-1},i_l} < \widehat{\theta}^*_{i_1,...,i_{l-1},j}\right\}\right) - 2. \qquad (42)
\end{aligned}$$

First, we show that the second term on the right hand side of (42) tends to one as $n \to \infty$, which would hold if for any $j \in \{\underline{a}_l, ..., i_l - 1\}$, $j \neq i_1, ..., i_{l-1}$, $\Pr_g(c_l \neq j) \to 1$. Clearly

$$\begin{aligned}
\Pr_g(c_l \neq j) &= \Pr_g\left(\bigcup_{\substack{k=j+1 \\ k\neq i_1,...,i_{l-1}}}^{\overline{a}_l}\left\{\widehat{\theta}_{i_1,...,i_{l-1},j} \geq \widehat{\theta}^*_{i_1,...,i_{l-1},k}\right\}\right) \geq \Pr_g\left(\widehat{\theta}_{i_1,...,i_{l-1},j} \geq \widehat{\theta}^*_{i_1,...,i_{l-1},i_l}\right) \\
&= \Pr_g\left(\widehat{\tau}^2_{i_1,...,i_{l-1},j} \geq \widehat{\tau}^2_{i_1,...,i_{l-1},i_l}\min\left\{1, \widehat{\kappa}_{i_1,...,i_{l-1},i_l}\right\}\right) \to 1 \text{ as } n \to \infty, \qquad (43)
\end{aligned}$$

because by (8), (10), $\widehat{\tau}^2_{i_1,...,i_{l-1},j}$ either has an exact rate $O_p(1)$ or diverges, by (9) $\widehat{\tau}^2_{i_1,...,i_{l-1},i_l} = O_p(1)$, whereas by (17) $\widehat{\kappa}_{i_1,...,i_{l-1},i_l} = o_p(1)$. Next, noting that for $j > i_l$, $j \neq i_1, ..., i_{l-1}$, $z_{i_1,t}, ..., z_{i_{l-1},t}, z_{j,t}$, are cointegrated, it can be readily shown that the third term on the right hand side of (42) tends to one as $n \to \infty$, to conclude the proof for the $\underline{a}_l < i_l < \overline{a}_l$ case.

Finally, if $i_l = \overline{a}_l$

$$\Pr_g\left(\bigcap_{k=1}^{l}\{c_k = i_k\}\right) = \Pr_g\left(\bigcap_{k=1}^{l-1}\{c_k = i_k\}, \bigcap_{\substack{j=\underline{a}_l \\ j\neq i_1,...,i_{l-1}}}^{\overline{a}_l-1}\{c_l \neq j\}\right) \to 1 \text{ as } n \to \infty,$$

by identical arguments to those in the proof of (43), to conclude the proof of the theorem.

Proof of Theorem 3. First, we show that the results given in the theorem apply also to conditional probabilities (given $o = g$), that is

$$\lim_{n\to\infty} \Pr_g\left(\widehat{r}_{(g)} = r\right) \geq 1 - (p-r)\,\alpha, \quad r < p, \tag{44}$$

$$\lim_{n\to\infty} \Pr_g\left(\widehat{r}_{(g)} = k\right) \leq \alpha, \quad k = r+1, ..., p, \quad r < p, \tag{45}$$

$$\lim_{n\to\infty} \Pr_g\left(\widehat{r}_{(g)} = k\right) = 0, \quad k = 0, ..., r-1, \quad r > 0. \tag{46}$$

First, we show (44) for $r = p$. For any null hypothesis H_0, denote RH_0, AH_0, if H_0 is rejected or not rejected, respectively. Then

$$\begin{aligned}\Pr_g\left(\widehat{r}_{(g)} = p\right) &= \Pr_g\left(\bigcap_{j_1=1}^{p} RH_{j_1}\right) \geq \sum_{j_1=1}^{p} \Pr_g\left(RH_{j_1}\right) - (p-1) \\ &= \sum_{j_1=1}^{p} \Pr_g\left(\widehat{\tau}_{j_1} < \nu_1(\alpha)\right) - (p-1),\end{aligned}$$

so the results holds immediately by (10) because if $r = p$, the $\overline{H}_i$'s are true for all $i = 1, ..., p$, and, under $\overline{H}_{j_1}$, $\Pr_g\left(\widehat{\tau}_{j_1} < \nu_1(\alpha)\right) \to 1$ as $n \to \infty$, noting that the $\widehat{\tau}_{j_1}$'s are independent of o.

Next, we show (45) for $k = p$ (which requires $r < p$). First, given that $r < p$, by Theorem 1 there exists (at least) a particular l such that H_l is true. Then

$$\Pr_g\left(\widehat{r}_{(g)} = p\right) = \Pr_g\left(\bigcap_{j_1=1}^{p} RH_{j_1}\right) \leq \Pr_g\left(RH_l\right) = \Pr_g\left(\widehat{\tau}_l < \nu_1(\alpha)\right), \tag{47}$$

so (45) for $k = p$ is justified by (9) taking limits on both sides of (47). Next, we show (45) for $k = p-1$ (which requires $r < p-1$). By the law of total probability

$$\begin{aligned}\Pr_g\left(\widehat{r}_{(g)} = p-1\right) &= \sum_{k_1=1}^{p} \Pr_g\left(\widehat{r}_{(g)} = p-1, c_1 = k_1\right) \\ &= \Pr_g\left(\widehat{r}_{(g)} = p-1, c_1 = i_1\right) \\ &\quad + \sum_{k_1 \neq i_1} \Pr_g\left(\widehat{r}_{(g)} = p-1, c_1 = k_1\right),\end{aligned} \tag{48}$$

noting that the first term in (48) corresponds to the particular correct choice of common trend defined in Theorem 2. The idea of the proof is to derive a bound for the first term of (48) (noting that the test statistics involved there are well constructed), and also show that the second term of (48) (which is potentially problematic because

the properties of the tests statistics involved are unknown) is $o(1)$. First, the first term on the right hand side of (48) is

$$\Pr_g\left(\bigcup_{j_1=1}^{p} AH_{j_1}, \bigcap_{\substack{j_2=1 \\ j_2\neq i_1}}^{p} RH_{i_1,j_2}, c_1 = i_1\right) \leq \Pr_g\left(\bigcap_{\substack{j_2=1 \\ j_2\neq i_1}}^{p} \left\{\widehat{\tau}_{i_1,j_2} < \nu_2(\alpha)\right\}, c_1 = i_1\right), \quad (49)$$

noting that the residual-based test statistics $\widehat{\tau}_{i_1,j_2}$ are well constructed because $z_{i_1,t}$ is a true common trend. Then, by Theorem 1, if $r < p-1$ there exists (al least) a particular $l(i_1)$ such that $H_{i_1,l(i_1)}$ is true. Then, the right hand side of (49) is bounded by

$$\Pr_g\left(\left\{\widehat{\tau}_{i_1,l(i_1)} < \nu_2(\alpha)\right\}, c_1 = i_1\right) \leq \Pr_g\left(\widehat{\tau}_{i_1,l(i_1)} < \nu_2(\alpha)\right), \quad (50)$$

so, by (9), $\lim_{n\to\infty} \Pr_g\left(\widehat{r}_{(g)} = p-1, c_1 = i_1\right) \leq \alpha$. Then (45) follows because the second term on the right hand side of (48) is clearly $o(1)$ by Theorem 2.

Next, we show (45) for $k = p-2$ (which requires $r < p-2$). By the law of total probability

$$\begin{aligned}
\Pr_g\left(\widehat{r}_{(g)} = p-2\right) &= \sum_{k_1=1}^{p}\sum_{\substack{k_2=1 \\ k_2\neq k_1}}^{p} \Pr_g\left(\widehat{r}_{(g)} = p-2, c_1 = k_1, c_2 = k_2\right) \quad (51)\\
&= \Pr_g\left(\widehat{r}_{(g)} = p-2, c_1 = i_1, c_2 = i_2\right)\\
&\quad + \sum_{(k_1,k_2)\neq(i_1,i_2)}\sum \Pr_g\left(\widehat{r}_{(g)} = p-2, c_1 = k_1, c_2 = k_2\right). \quad (52)
\end{aligned}$$

The first term on the right hand side of (52) is

$$\begin{aligned}
&\Pr_g\left(\bigcup_{j_1=1}^{p} AH_{j_1}, \bigcup_{\substack{j_2=1 \\ j_2\neq i_1}}^{p} AH_{i_1,j_2}, \bigcap_{\substack{j_3=1 \\ j_3\neq i_1,i_2}}^{p} RH_{i_1,i_2,j_3}, c_1 = i_1, c_2 = i_2\right)\\
\leq\ &\Pr_g\left(\bigcap_{\substack{j_3=1 \\ j_3\neq i_1,i_2}}^{p} \left\{\widehat{\tau}_{i_1,i_2,j_3} < \nu_3(\alpha)\right\}, c_1 = i_1, c_2 = i_2\right), \quad (53)
\end{aligned}$$

noting again that the residual-based test statistics $\widehat{\tau}_{i_1,i_2,j_3}$ are well constructed because $z_{i_1,t}$, $z_{i_2,t}$, are true common trends. Then, by Theorem 1, if $r < p-2$ there exists (at least) a particular $l(i_1,i_2)$ such that $H_{i_1,i_2,l(i_1,i_2)}$ is true. Then, the right hand side of

(53) is bounded by

$$\Pr_g\left(\left\{\widehat{\tau}_{i_1,i_2,l(i_1,i_2)} < \nu_3(\alpha)\right\}, c_1 = i_1, c_2 = i_2\right) \leq \Pr_g\left(\widehat{\tau}_{i_1,i_2,l(i_1,i_2)} < \nu_3(\alpha)\right), \tag{54}$$

so the result holds by (9) by taking limits on the right hand side of (54) and noting that the second term on the right hand side of (52) is clearly $o(1)$ by Theorem 2.

The proof of (45) for a general $k = r+1, ..., p-1$, is as follows. By the law of total probability,

$$\begin{aligned}
\Pr_g\left(\widehat{r}_{(g)} = k\right) &= \sum_{k_1=1}^{p} \sum_{\substack{k_2=1 \\ k_2 \neq k_1}}^{p} ... \sum_{\substack{k_{p-k}=1 \\ k_{p-k} \neq k_1,...,k_{p-k-1}}}^{p} \Pr_g\left(\widehat{r}_{(g)} = k, c_1 = k_1, c_2 = k_2, ..., c_{p-k} = k_{p-k}\right) \\
&= \Pr_g\left(\widehat{r}_{(g)} = k, \bigcap_{h=1}^{p-k} \{c_h = i_h\}\right) \\
&\quad + \sum_{\left(k_1,...,k_{p-k}\right) \neq \left(i_1,...,i_{p-k}\right)} .... \sum \Pr_g\left(\widehat{r}_{(g)} = k, \bigcap_{h=1}^{p-k} \{c_h = k_h\}\right). \qquad (55)
\end{aligned}$$

The first term on the right hand side of (55) equals

$$\begin{aligned}
&\Pr_g\left(\bigcup_{j_1=1}^{p} AH_{j_1}, ..., \bigcup_{\substack{j_{p-k}=1 \\ j_{p-k} \neq i_1,...,i_{p-k-1}}}^{p} AH_{i_1,...,i_{p-k-1},j_{p-k}}, \right. \\
&\left. \bigcap_{\substack{j_{p-k+1}=1 \\ j_{p-k+1} \neq i_1,...,i_{p-k}}}^{p} RH_{i_1,...,i_{p-k},j_{p-k+1}}, \bigcap_{h=1}^{p-k} \{c_h = i_h\}\right) \qquad (56) \\
\leq\ &\Pr_g\left(\bigcap_{\substack{j_{p-k+1}=1 \\ j_{p-k+1} \neq i_1,...,i_{p-k}}}^{p} RH_{i_1,...,i_{p-k},j_{p-k+1}}, \bigcap_{h=1}^{p-k} \{c_h = i_h\}\right) \qquad (57)
\end{aligned}$$

noting again that the residual-based test statistics $\widehat{\tau}_{i_1,...,i_{p-k},j_{p-k+1}}$ are well constructed because $z_{i_1,t}, ..., z_{i_{p-k},t}$, are true common trends. Then, by Theorem 1, if $r < k$, there exists (al least) a particular $l(i_1, ..., i_{p-k})$ such that $H_{i_1,...,i_{p-k},l\left(i_1,...,i_{p-k}\right)}$ is true, so the

right hand side of (57) is bounded by

$$\begin{aligned} & \Pr_g\left(\left\{\widehat{\tau}_{i_1,...,i_{p-k},l\left(i_1,...,i_{p-k}\right)} < \nu_{p-k+1}(\alpha)\right\}, \bigcap_{h=1}^{p-k}\{c_h = i_h\}\right) \\ \leq & \Pr_g\left(\left\{\widehat{\tau}_{i_1,...,i_{p-k},l\left(i_1,...,i_{p-k}\right)} < \nu_{p-k+1}(\alpha)\right\}\right), \end{aligned} \tag{58}$$

so, again, the result follows by (9) by taking limits on the right hand side of (58) and noting that the second term on the right hand side of (55) is clearly $o(1)$ by Theorem 2.

Next we show (44) for $r < p$. By the law of total probability

$$\begin{aligned} \Pr_g\left(\widehat{r}_{(g)} = r\right) &= \sum_{k_1=1}^{p}\sum_{\substack{k_2=1\\k_2\neq k_1}}^{p}...\sum_{\substack{k_{p-r}=1\\k_{p-r}\neq k_1,...,k_{p-r-1}}}^{p}\Pr_g\left(\widehat{r}_{(g)} = r, c_1 = k_1, c_2 = k_2, ..., c_{p-r} = k_{p-r}\right) \\ &= \Pr_g\left(\widehat{r}_{(g)} = r, \bigcap_{h=1}^{p-r}\{c_h = i_h\}\right) \\ &\quad + \sum_{(k_1,...,k_{p-r})\neq(i_1,...,i_{p-r})}....\sum\Pr_g\left(\widehat{r}_{(g)} = r, \bigcap_{h=1}^{p-r}\{c_h = k_h\}\right). \end{aligned} \tag{59}$$

First, let $r > 0$. Then, the first term on the right hand side of (59) equals

$$\begin{aligned} & \Pr_g\left(\bigcup_{j_1=1}^{p} AH_{j_1}, ..., \bigcup_{\substack{j_{p-r}=1\\j_{p-r}\neq i_1,...,i_{p-r-1}}}^{p} AH_{i_1,...,i_{p-r-1},j_{p-r}},\right. \\ & \left.\bigcap_{\substack{j_{p-r+1}=1\\j_{p-r+1}\neq i_1,...,i_{p-r}}}^{p} RH_{i_1,...,i_{p-r},j_{p-r+1}}, \bigcap_{h=1}^{p-r}\{c_h = i_h\}\right) \\ = & \Pr_g\left(\bigcup_{j_1=1}^{p}\left\{\widehat{\tau}_{j_1} \geq \nu_1(\alpha)\right\}, ..., \bigcup_{\substack{j_{p-r}=1\\j_{p-r}\neq i_1,...,i_{p-r-1}}}^{p}\left\{\widehat{\tau}_{i_1,...,i_{p-r-1},j_{p-r}} \geq \nu_{p-r}(\alpha)\right\}\right. \\ & \left., \bigcap_{\substack{j_{p-r+1}=1\\j_{p-r+1}\neq i_1,...,i_{p-r}}}^{p}\left\{\widehat{\tau}_{i_1,...,i_{p-r},j_{p-r+1}} < \nu_{p-r+1}(\alpha)\right\}, \bigcap_{h=1}^{p-r}\{c_h = i_h\}\right), \end{aligned} \tag{60}$$

noting again that the residual-based test statistics $\widehat{\tau}_{j_1}$, $\widehat{\tau}_{i_1,j_2}, ..., \widehat{\tau}_{i_1,...,i_{p-r},j_{p-r+1}}$ are well constructed because $z_{i_1,t}, ..., z_{i_{p-r},t}$, are true common trends. Then, given that the

cointegrating rank is r, by Theorem 1 there exist particular $l, l(i_1), l(i_1,i_2), ..., l(i_1,...,i_{p-r-1})$ such that the hypothesis H_l, $H_{i_1,l(i_1)}$, $H_{i_1,i_2,l(i_1,i_2)}, ..., H_{i_1,...,i_{p-r-1},l(i_1,...,i_{p-r-1})}$ are true. Thus, (60) is bounded below by

$$
\begin{aligned}
&= \Pr_g\left(\left\{\widehat{\tau}_l \geq \nu_1(\alpha)\right\}, \left\{\widehat{\tau}_{i_1,l(i_1)} \geq \nu_2(\alpha)\right\}, ..., \left\{\widehat{\tau}_{i_1,...,i_{p-r-1},l(i_1,...,i_{p-r-1})} \geq \nu_{p-r}(\alpha)\right\},\right. \\
&\qquad \left. \bigcap_{\substack{j_{p-r+1}=1 \\ j_{p-r+1}\neq i_1,...,i_{p-r}}}^{p} \left\{\widehat{\tau}_{i_1,...,i_{p-r},j_{p-r+1}} < \nu_{p-r+1}(\alpha)\right\}, \bigcap_{h=1}^{p-r}\{c_h = i_h\}\right) \\
&\geq \Pr_g\left(\widehat{\tau}_l \geq \nu_1(\alpha)\right) + \Pr_g\left(\widehat{\tau}_{i_1,l(i_1)} \geq \nu_2(\alpha)\right) + ... + \Pr_g\left(\widehat{\tau}_{i_1,...,i_{p-r-1},l(i_1,...,i_{p-r-1})} \geq \nu_{p-r}(\alpha)\right) \\
&\quad + \Pr_g\left(\bigcap_{\substack{j_{p-r+1}=1 \\ j_{p-r+1}\neq i_1,...,i_{p-r}}}^{p} \left\{\widehat{\tau}_{i_1,...,i_{p-r},j_{p-r+1}} < \nu_{p-r+1}(\alpha)\right\}\right) \\
&\quad + \Pr_g\left(\bigcap_{h=1}^{p-r}\{c_h = i_h\}\right) - (p-r+1). \qquad (61)
\end{aligned}
$$

Clearly, by (9) the limits of the first $p-r$ terms in the right hand side of (61) are bounded below by $1-\alpha$, so (44) follows on showing that as $n \to \infty$,

$$
\Pr_g\left(\bigcap_{\substack{j_{p-r+1}=1 \\ j_{p-r+1}\neq i_1,...,i_{p-r}}}^{p} \left\{\widehat{\tau}_{i_1,...,i_{p-r},j_{p-r+1}} < \nu_{p-r+1}(\alpha)\right\}\right) \to 1, \qquad (62)
$$

and also that $\Pr_g\left(\bigcap_{h=1}^{p-r}\{c_h = i_h\}\right) \to 1$. First, this latter result holds immediately by Theorem 2. Next, the left hand side of (62) is bounded below

$$
\sum_{\substack{j_{p-r+1}=1 \\ j_{p-r+1}\neq i_1,...,i_{p-r}}}^{p} \Pr_g\left(\widehat{\tau}_{i_1,...,i_{p-r},j_{p-r+1}} < \nu_{p-r+1}(\alpha)\right) - (r-1),
$$

noting that the intersection in (62) involves r events, so (62) holds by (10) because by Theorem 1 all null hypotheses $H_{i_1,...,i_{p-r},j_{p-r+1}}$ are false.

The proof of (44) for $r=0$ is much simpler. By (45), $\Pr_g\left(\widehat{r}_{(g)} = j\right) \leq \alpha$ for all $j = 1, ..., p$, which immediately implies that $\Pr_g\left(\widehat{r}_{(g)} = 0\right) \geq 1 - p\alpha$.

Finally, we show (46). First, let $k = r-1$. By the law of total probability,

$\Pr_g\left(\widehat{r}_{(g)}=r-1\right)$ equals

$$\sum_{k_1=1}^{p}\sum_{\substack{k_2=1\\k_2\neq k_1}}^{p}...\sum_{\substack{k_{p-r+1}=1\\k_{p-r+1}\neq k_1,...,k_{p-r}}}^{p}\Pr_g\left(\widehat{r}_{(g)}=r-1,c_1=k_1,c_2=k_2,...,c_{p-r+1}=k_{p-r+1}\right)$$

$$=\sum_{\substack{k_{p-r+1}=1\\k_{p-r+1}\neq i_1,...,i_{p-r}}}^{p}\Pr_g\left(\widehat{r}_{(g)}=r-1,\bigcap_{h=1}^{p-r}\left\{c_h=i_h\right\},\left\{c_{p-r+1}=k_{p-r+1}\right\}\right)$$

$$+\sum_{(k_1,...,k_{p-r})\neq(i_1,...,i_{p-r})}....\sum\sum_{\substack{k_{p-r+1}=1\\k_{p-r+1}\neq k_1,...,k_{p-r}}}^{p}\Pr_g\left(\widehat{r}_{(g)}=r-1,\bigcap_{h=1}^{p-r+1}\left\{c_h=k_h\right\}\right). \qquad (63)$$

First, by previous results, it is straightforward to show that the second term on the right hand side of (63) is $o(1)$. Next, the first term on the right hand side of (63) is

$$\sum_{\substack{k_{p-r+1}=1\\k_{p-r+1}\neq i_1,...,i_{p-r}}}^{p}\Pr_g\left(\bigcup_{j_1=1}^{p}AH_{j_1},...,\bigcup_{\substack{j_{p-r}=1\\j_{p-r}\neq i_1,...,i_{p-r-1}}}^{p}AH_{i_1,...,i_{p-r-1},j_{p-r}},\right.$$

$$\left.\bigcup_{\substack{j_{p-r+1}=1\\j_{p-r+1}\neq i_1,...,i_{p-r}}}^{p}AH_{i_1,...,i_{p-r},j_{p-r+1}},\right.$$

$$\left.\bigcap_{\substack{j_{p-r+2}=1\\j_{p-r+2}\neq i_1,...,i_{p-r},k_{p-r+1}}}^{p}RH_{i_1,...,i_{p-r},k_{p-r+1},j_{p-r+2}},\bigcap_{h=1}^{p-r}\left\{c_h=i_h\right\},\left\{c_{p-r+1}=k_{p-r+1}\right\}\right)$$

$$\leq\ p\Pr_g\left(\bigcup_{\substack{j_{p-r+1}=1\\j_{p-r+1}\neq i_1,...,i_{p-r}}}^{p}AH_{i_1,...,i_{p-r},j_{p-r+1}}\right)$$

$$=\ p\left(1-\Pr_g\left(\bigcap_{\substack{j_{p-r+1}=1\\j_{p-r+1}\neq i_1,...,i_{p-r}}}^{p}\left\{\widehat{\tau}_{i_1,...,i_{p-r},j_{p-r+1}}<\nu_{p-r+1}(\alpha)\right\}\right)\right)$$

so the required result follows by (62).

Next, we show (46) for a generic $k=0,...,r-2$. First, by the law of total

probability, $\Pr_g\left(\widehat{r}_{(g)}=k\right)$ equals

$$\begin{aligned}
&\sum_{k_1=1}^{p}\sum_{\substack{k_2=1\\k_2\neq k_1}}^{p}...\sum_{\substack{k_{p-k}=1\\k_{p-k}\neq k_1,...,k_{p-k-1}}}^{p}\Pr_g\left(\widehat{r}_{(g)}=k,c_1=k_1,c_2=k_2,...,c_{p-k}=k_{p-k}\right)\\
=&\sum_{\substack{k_{p-r+1}=1\\k_{p-r+1}\neq i_1,...,i_{p-r}}}^{p}...\sum_{\substack{k_{p-k}=1\\k_{p-k}\neq i_1,...,i_{p-r},k_{p-r+1},...,k_{p-k-1}}}^{p}\Pr_g\left(\widehat{r}_{(g)}=k,\bigcap_{h=1}^{p-r}\left\{c_h=i_h\right\},\bigcap_{h=p-r+1}^{p-k}\left\{c_h=k_h\right\}\right)\\
&+\sum_{(k_1,...,k_{p-r})\neq(i_1,...,i_{p-r})}....\sum\sum_{\substack{k_{p-r+1}=1\\k_{p-r+1}\neq k_1,...,k_{p-r}}}^{p}...\sum_{\substack{k_{p-k}=1\\k_{p-k}\neq k_1,...,k_{p-k-1}}}^{p}\Pr_g\left(\widehat{r}_{(g)}=k,\bigcap_{h=1}^{p-k}\left\{c_h=k_h\right\}\right).
\end{aligned}\tag{64}$$

Again, by previous results, it is straightforward to show that the second term on the right hand side of (64) is $o(1)$. Regarding the first term on the right hand side of (64), by an almost identical analysis to that for the first term on the right hand side of (63), it can be bounded by

$$p^{r-k}\left(1-\Pr_g\left(\bigcap_{\substack{j_{p-r+1}=1\\j_{p-r+1}\neq i_1,...,i_{p-r}}}^{p}\left\{\widehat{\tau}_{i_1,...,i_{p-r},j_{p-r+1}}<\nu_{p-r+1}(\alpha)\right\}\right)\right),$$

so, again, the required result follows by (62), to conclude the proof of (46).

Finally, (28)-(30) are straightforwardly derived from (44)-(46) just noting that for any $k=0,...,p$, by the law of total probability

$$\Pr(\widehat{r}=k)=\sum_{g=1}^{p!}\Pr_g(\widehat{r}=k)\Pr(o=g),$$

to conclude the proof of Theorem 3.

ACKNOWLEDGEMENTS

We are grateful to James Davidson, José García-Montalvo, Jesús Gonzalo, Adrian Pagan, Carlos Velasco and seminar participants at University of Sassari, Universidad Carlos III, Universitat Pompeu Fabra and Universidad Pública de Navarra for useful comments. Javier Gomez-Biscarri acknowledges support from the Ministerio de Educación y Ciencia through projects ECO2008-05155 and ECO2011-25607 and through the Severo Ochoa Program for Centers of Excellence in R&D (SEV-2011-

0075). Javier Hualde's research is supported by the Spanish Ministerio de Ciencia e Innovación through project ref. ECO2011-24304.

REFERENCES

Ahn, B.C., and I. Choi, 1995, Testing for cointegration in a system of equations. Econometric Theory 11, 952-983.

Ahn, S.K., and G.C. Reinsel, 1990, Estimation for partially nonstationary multivariate autoregressive models. Journal of the American Statistical Association 85, 813-823.

Bierens, H.J., 1997, Nonparametric cointegration analysis. Journal of Econometrics 77, 379-404.

Boswijk, H.P., 1996, Testing identifiability of cointegrating vectors. Journal of Business and Economic Statistics 14, 153-160.

Breitung, J., 2002, Nonparametric tests for unit roots and cointegration. Journal of Econometrics 108, 343-363.

Diebold, F.X., and C. Li, 2006, Forecasting the term structure of government bond yields. Journal of Econometrics 130, 337-364.

Engle, R.F., and C.W.J. Granger, 1987, Cointegration and error correction model. Representation, estimation and testing. Econometrica 55, 251-276.

Giese, J.V., 2008, Level, slope, curvature: characterising the yield curve in a cointegrated VAR model. Economics: The Open-Access, Open-Assessment E-Journal 2, 2008-2028.

Gonzalo, J., and T.H. Lee, 1998, Pitfalls in testing for long run relationships. Journal of Econometrics 86, 129-154.

Hall, A.D., Anderson, H.M., and C.W.J. Granger, 1992, A cointegration analysis of treasury bill yields. Review of Economics and Statistics 74, 116–126.

Harris, D., 1997, Principal components analysis of cointegrated time series. Econometric Theory 13, 529-557.

Haug, A.A., 1996, Tests for cointegration. A Monte Carlo comparison. Journal of Econometrics 71, 89-115.

Hoover, K., Johansen, S., and K. Juselius, 2008, Allowing the data to speak freely: the macroeconometrics of the cointegrated vector autoregression. American Economic Review 98, 251-255.

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 cointegrating vectors in Gaussian vector autoregressive models. Econometrica, 59 1551-1580.

Johansen, S., 1995, Likelihood based inference in cointegrated vector autoregressive models. Oxford University Press, Oxford.

Johansen, S., 2005, Interpretation of cointegrating coefficients in the cointegrated vector autoregressive model. Oxford Bulletin of Economics and Statistics 67, 93-104.

Johansen, S., 2010, Some identification problems in the cointegrated vector autoregressive model. Journal of Econometrics 158, 262-273.

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.

Kurozumi, E., 2005, The rank of a submatrix of cointegration. Econometric Theory 21, 299-325.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992, Testing the null hypothesis of stationary 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.

Luukkonen, R., Ripatti, A., and P. Saikkonen, 1999, Testing for a valid normalization of cointegrating vectors in vector autoregressive processes. Journal of Business and Economic Statistics 17, 195-204.

Nyblom, J., and A. Harvey, 2000, Tests of common stochastic trends. Econometric Theory 16, 176-199.

Paruolo, P., 2006, The likelihood ratio test for the rank of a cointegration submatrix. Oxford Bulletin of Economics and Statistics 68, 921-948.

Pesaran, M.H., and Y. Shin, 2002, Long-run structural modelling. Econometric Reviews 21, 49-87.

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. Tauchen (Eds.), Nonparametric and Semiparametric Methods in Econometrics and Statistics. Cambridge University Press, Cambridge, pp. 413-436.

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 S. Ouliaris, 1990, Asymptotic properties of residual based tests for cointegration. Econometrica 58, 165-193.

Saikkonen, P., 1993, Estimation of cointegration vectors with linear restrictions. Econometric Theory 9, 19-35.

Snell, A., 1999, Testing for r versus $r-1$ cointegrating vectors. Journal of Econometrics 88, 151-191.

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.

Tanaka, K., 1990, Testing for a moving average root. Econometric Theory 6, 433-444.

Toda, H.Y., 1995, Finite sample performance of likelihood ratio tests for cointegrating ranks in vector autoregressions. Econometric Theory 11, 1015-1032.

Table 1. Performance of the GBH procedure - White noise error

	r	4			3			2			1			0		
$\widehat{r}$	$_n\backslash^\alpha$	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
	50	.851	.800	.692	.097	.053	.017	.013	.004	.001	.003	.001	.000	.001	.000	.000
	100	.997	.993	.981	.091	.047	.011	.009	.002	.000	.002	.000	.000	.000	.000	.000
4	200	1.00	1.00	1.00	.092	.042	.009	.008	.002	.000	.001	.000	.000	.000	.000	.000
	500	1.00	1.00	1.00	.091	.044	.008	.006	.002	.000	.001	.000	.000	.000	.000	.000
	1000	1.00	1.00	1.00	.094	.044	.008	.008	.002	.000	.001	.000	.000	.000	.000	.000
	50	.131	.171	.242	.752	.754	.706	.148	.080	.026	.027	.011	.001	.008	.002	.000
	100	.003	.007	.019	.900	.938	.959	.125	.073	.018	.022	.007	.001	.005	.001	.000
3	200	.000	.000	.000	.908	.958	.990	.112	.061	.013	.015	.003	.000	.002	.000	.000
	500	.000	.000	.000	.909	.956	.992	.101	.054	.010	.013	.003	.000	.002	.000	.000
	1000	.000	.000	.000	.906	.956	.992	.091	.052	.011	.011	.003	.000	.002	.000	.000
	50	.017	.027	.055	.132	.163	.221	.702	.741	.723	.239	.152	.054	.078	.029	.004
	100	.000	.000	.000	.009	.015	.029	.855	.906	.943	.201	.126	.041	.045	.015	.001
2	200	.000	.000	.000	.000	.000	.001	.880	.936	.985	.161	.094	.027	.024	.009	.001
	500	.000	.000	.000	.000	.000	.000	.893	.944	.990	.120	.067	.016	.016	.005	.000
	1000	.000	.000	.000	.000	.000	.000	.901	.946	.989	.106	.056	.011	.016	.004	.000
	50	.001	.002	.010	.017	.027	.049	.125	.160	.221	.644	.722	.787	.300	.226	.087
	100	.000	.000	.000	.000	.000	.001	.011	.019	.039	.764	.848	.926	.255	.163	.054
1	200	.000	.000	.000	.000	.000	.000	.000	.001	.002	.823	.901	.970	.193	.116	.032
	500	.000	.000	.000	.000	.000	.000	.000	.000	.000	.866	.930	.984	.137	.076	.017
	1000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.882	.941	.989	.114	.057	.012
	50	.000	.000	.001	.002	.003	.007	.012	.015	.029	.087	.114	.158	.613	.743	.909
	100	.000	.000	.000	.000	.000	.000	.000	.000	.000	.011	.019	.032	.695	.821	.945
0	200	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.002	.003	.781	.875	.967
	500	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.845	.919	.983
	1000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.868	.939	.988

The cells show the proportion of 10,000 replications where the estimated rank is $\widehat{r}$, given the correct rank r, obtained following the GBH procedure. Significance levels $\alpha = \{.10, .05, .01\}$ were used in the tests. The number of lags in the ADF tests is chosen using the BIC. The innovation vector u_t is generated from $u_{j,t} = \varepsilon_{j,t},\ \ j = 1, ..., 4$, with Gaussian ε_t such that $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = I_4$

Table 2. Performance of Johansen's trace test - White noise error

	r	4			3			2			1			0		
$\widehat{r}$	$_n\backslash^\alpha$	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
	50	.999	1.00	.999	.066	.031	.006	.008	.003	.000	.004	.002	.002	.005	.007	.002
	100	1.00	1.00	1.00	.064	.031	.006	.007	.001	.000	.001	.000	.000	.000	.000	.000
4	200	1.00	1.00	1.00	.063	.033	.005	.006	.002	.000	.001	.000	.000	.000	.000	.000
	500	1.00	1.00	1.00	.062	.033	.007	.005	.001	.000	.001	.000	.000	.000	.000	.000
	1000	1.00	1.00	1.00	.066	.033	.006	.005	.002	.000	.000	.000	.000	.000	.000	.000
	50	.000	.000	.001	.933	.963	.925	.048	.024	.006	.006	.003	.003	.003	.003	.004
	100	.000	.000	.000	.936	.969	.994	.048	.024	.005	.002	.001	.000	.000	.000	.000
3	200	.000	.000	.000	.937	.967	.995	.050	.021	.005	.002	.001	.000	.000	.000	.000
	500	.000	.000	.000	.938	.967	.993	.047	.025	.005	.003	.001	.000	.000	.000	.000
	1000	.000	.000	.000	.934	.967	.994	.051	.024	.004	.002	.001	.000	.000	.000	.000
	50	.001	.000	.000	.001	.006	.069	.920	.907	.729	.050	.021	.005	.004	.003	.002
	100	.000	.000	.000	.000	.000	.000	.945	.975	.995	.042	.021	.005	.004	.001	.000
2	200	.000	.000	.000	.000	.000	.000	.944	.977	.995	.044	.021	.003	.004	.001	.000
	500	.000	.000	.000	.000	.000	.000	.948	.974	.995	.038	.023	.004	.003	.001	.000
	1000	.000	.000	.000	.000	.000	.000	.944	.974	.996	.041	.017	.004	.002	.000	.000
	50	.000	.000	.000	.000	.000	.000	.024	.066	.265	.886	.860	.660	.059	.030	.008
	100	.000	.000	.000	.000	.000	.000	.000	.000	.000	.955	.978	.994	.050	.023	.005
1	200	.000	.000	.000	.000	.000	.000	.000	.000	.000	.953	.978	.997	.040	.021	.003
	500	.000	.000	.000	.000	.000	.000	.000	.000	.000	.958	.976	.996	.038	.019	.003
	1000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.957	.982	.996	.040	.019	.002
	50	.000	.000	.000	.000	.000	.000	.000	.000	.000	.054	.114	.330	.929	.957	.984
	100	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.946	.976	.995
0	200	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.956	.978	.997
	500	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.959	.980	.997
	1000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.958	.981	.998

The cells show the proportion of 10,000 replications where the estimated rank is $\widehat{r}$, given the correct rank r, obtained following Johansen's trace test. Significance levels $\alpha = \{.10, .05, .01\}$ were used in the tests. The number of lags in the estimated VARs is chosen using the BIC. The innovation vector u_t is generated from $u_{j,t} = \varepsilon_{j,t},\ \ j = 1, ..., 4,$ with Gaussian ε_t such that $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = I_4$

Table 3. Performance of the GBH procedure - Autoregressive error

	r	4			3			2			1			0		
$\widehat{r}$	$_n\backslash^\alpha$	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
	50	.109	.031	.001	.082	.047	.011	.015	.004	.001	.003	.001	.000	.001	.000	.000
	100	.821	.592	.106	.061	.030	.006	.008	.002	.000	.001	.000	.000	.000	.000	.000
4	200	.999	.995	.970	.061	.026	.006	.006	.003	.000	.001	.000	.000	.000	.000	.000
	500	1.00	1.00	1.00	.061	.029	.006	.008	.002	.000	.001	.000	.000	.000	.000	.000
	1000	1.00	1.00	1.00	.069	.034	.006	.006	.002	.000	.001	.000	.000	.000	.000	.000
	50	.091	.036	.003	.161	.084	.015	.054	.029	.005	.017	.005	.001	.009	.003	.000
	100	.073	.110	.049	.683	.489	.150	.061	.031	.005	.011	.003	.000	.004	.001	.000
3	200	.001	.003	.014	.937	.967	.925	.065	.028	.005	.009	.003	.000	.002	.001	.000
	500	.000	.000	.000	.939	.971	.994	.072	.032	.005	.007	.002	.000	.002	.001	.000
	1000	.000	.000	.000	.931	.966	.994	.066	.030	.007	.009	.003	.000	.002	.000	.000
	50	.196	.126	.025	.191	.117	.031	.211	.114	.027	.120	.065	.018	.072	.032	.004
	100	.061	.117	.121	.169	.227	.141	.590	.413	.125	.095	.049	.011	.040	.014	.001
2	200	.000	.002	.012	.002	.007	.058	.923	.951	.847	.082	.039	.007	.026	.008	.000
	500	.000	.000	.000	.000	.000	.000	.920	.966	.995	.075	.031	.006	.018	.004	.000
	1000	.000	.000	.000	.000	.000	.000	.928	.968	.993	.071	.032	.007	.011	.005	.000
	50	.306	.308	.178	.276	.263	.148	.338	.292	.148	.346	.265	.122	.288	.205	.074
	100	.034	.110	.266	.065	.159	.251	.253	.334	.289	.606	.499	.230	.243	.158	.050
1	200	.000	.000	.003	.000	.000	.009	.006	.018	.127	.899	.927	.829	.175	.106	.030
	500	.000	.000	.000	.000	.000	.000	.000	.000	.000	.917	.967	.994	.137	.075	.016
	1000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.919	.965	.993	.109	.060	.014
	50	.298	.499	.793	.290	.489	.795	.382	.561	.819	.514	.664	.859	.630	.760	.922
	100	.011	.071	.458	.022	.095	.452	.088	.220	.581	.287	.449	.759	.713	.827	.949
0	200	.000	.000	.001	.000	.000	.002	.000	.000	.021	.009	.031	.164	.797	.885	.970
	500	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.843	.920	.984
	1000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.878	.935	.986

The cells show the proportion of 10,000 replications where the estimated rank is $\widehat{r}$, given the correct rank r, obtained following the GBH procedure. Significance levels $\alpha = \{.10, .05, .01\}$ were used in the tests. The number of lags in the ADF tests is chosen using the BIC. The innovation vector u_t is generated from $u_{j,t} = 0.8u_{j,t-1} + \varepsilon_{j,t}$, $j = 1, ..., 4$, with Gaussian ε_t such that $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = I_4$

Table 4. Performance of Johansen's trace test - Autoregressive error

	r	4			3			2			1			0		
$\widehat{r}$	$_n\backslash^\alpha$	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
	50	.017	.005	.001	.032	.022	.012	.106	.084	.044	.220	.177	.125	.379	.301	.236
	100	.517	.245	.015	.038	.012	.001	.020	.004	.000	.018	.006	.000	.025	.006	.000
4	200	1.00	.997	.900	.096	.045	.004	.020	.005	.000	.010	.002	.000	.006	.002	.000
	500	1.00	1.00	1.00	.102	.049	.011	.018	.006	.000	.006	.001	.000	.002	.001	.000
	1000	1.00	1.00	1.00	.099	.052	.011	.017	.006	.000	.005	.000	.000	.001	.000	.000
	50	.026	.010	.003	.037	.025	.023	.066	.049	.054	.093	.065	.066	.130	.092	.067
	100	.189	.230	.097	.109	.049	.006	.050	.020	.003	.039	.017	.001	.055	.020	.002
3	200	.000	.003	.098	.728	.581	.214	.092	.045	.007	.026	.011	.000	.015	.004	.001
	500	.000	.000	.000	.898	.951	.989	.094	.053	.011	.020	.007	.001	.006	.002	.000
	1000	.000	.000	.000	.901	.948	.989	.095	.049	.011	.015	.006	.000	.004	.001	.000
	50	.065	.032	.007	.115	.064	.020	.175	.122	.056	.199	.154	.066	.200	.182	.095
	100	.135	.214	.196	.276	.209	.065	.189	.109	.026	.152	.091	.019	.143	.089	.023
2	200	.000	.000	.002	.169	.344	.597	.576	.449	.179	.141	.080	.020	.063	.026	.004
	500	.000	.000	.000	.000	.000	.000	.888	.941	.987	.112	.067	.015	.026	.009	.001
	1000	.000	.000	.000	.000	.000	.000	.888	.945	.989	.111	.053	.011	.015	.006	.001
	50	.225	.160	.057	.433	.394	.256	.364	.337	.241	.302	.316	.242	.210	.271	.272
	100	.108	.194	.337	.376	.409	.339	.421	.396	.221	.407	.351	.185	.351	.318	.173
1	200	.000	.000	.000	.007	.029	.172	.293	.457	.614	.597	.551	.317	.242	.176	.062
	500	.000	.000	.000	.000	.000	.000	.000	.000	.002	.862	.925	.970	.144	.093	.026
	1000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.869	.941	.989	.116	.071	.015
	50	.667	.793	.932	.383	.495	.689	.289	.408	.605	.186	.288	.501	.081	.154	.330
	100	.051	.117	.355	.201	.321	.589	.320	.471	.750	.384	.535	.795	.426	.567	.802
0	200	.000	.000	.000	.000	.001	.013	.019	.044	.200	.226	.356	.663	.674	.792	.933
	500	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.014	.822	.895	.973
	1000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.864	.922	.984

The cells show the proportion of 10,000 replications where the estimated rank is $\widehat{r}$, given the correct rank r, obtained following Johansen's trace test. Significance levels $\alpha = \{.10, .05, .01\}$ were used in the tests. The number of lags in the estimated VARs is chosen using the BIC. The innovation vector u_t is generated from $u_{j,t} = 0.8u_{j,t-1} + \varepsilon_{j,t}$, $j = 1, ..., 4$, with Gaussian ε_t such that $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = I_4$

Table 5

Performance of GBH procedure with correlated errors - Correct rank estimation

r	4			3			2			1			0		
$n \backslash \alpha$	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
	Panel A. $Cov(\varepsilon_{i,t}, \varepsilon_{j,t}) = 0.4$, $i \neq j$														
50	.853	.812	.688	718	.717	.668	.674	.699	.691	.632	.704	.764	.621	.760	.917
WN 100	.996	.993	.982	.890	.922	.943	.840	.880	.916	.755	.836	.896	.706	.825	.946
200	1.00	1.00	1.00	.912	.953	.988	.879	.933	.978	.822	.899	.962	.785	.885	.967
500	1.00	1.00	1.00	.904	.955	.992	.879	.945	.985	.859	.930	.984	.846	.925	.985
50	.138	.043	.002	.160	.079	.016	.179	.106	.022	.338	.243	.104	.634	.771	.923
AR 100	.832	.630	.150	.660	.477	.144	.570	.394	.115	.583	.476	.217	.712	.832	.947
200	.999	.997	.970	.931	.961	.918	.930	.950	.832	.910	.932	.818	.791	.883	.974
500	1.00	1.00	1.00	.929	.967	.995	.931	.969	.995	.923	.970	.996	.845	.922	.983
	Panel B. $Cov(\varepsilon_{i,t}, \varepsilon_{j,t}) = 0.8$, $i \neq j$														
50	.880	.838	.758	.672	.662	.618	.580	.620	.609	.589	.650	.669	.632	.776	.920
WN 100	.996	.992	.984	.870	.905	.897	.743	.782	.797	.691	.760	.786	.714	.824	.947
200	1.00	1.00	1.00	.910	.951	.985	.827	.888	.920	.765	.842	.875	.781	.887	.970
500	1.00	1.00	1.00	.904	.958	.991	.853	.926	.983	.838	.911	.972	.845	.915	.985
50	.275	.130	.020	.159	.084	.020	.155	.085	.018	.285	.204	.077	.634	.770	.921
AR 100	.877	.724	.291	.643	.452	.135	.529	.340	.091	.522	.398	.176	.724	.831	.950
200	.999	.998	.976	.935	.962	.914	.923	.943	.811	.911	.925	.769	.785	.885	.972
500	1.00	1.00	1.00	.932	.970	.994	.925	.967	.996	.927	.972	.997	.850	.922	.983
	Panel C. $Cov(\varepsilon_{i,t}, \varepsilon_{j,t}) = -0.3$, $i \neq j$														
50	.853	.800	.693	.778	.764	.717	.738	.748	.729	.670	.736	.791	.656	.799	.934
WN 100	.997	.991	.984	.920	.947	.965	.865	.916	.948	.767	.854	.933	.728	.845	.959
200	1.00	1.00	1.00	.924	.968	.992	.889	.942	.984	.838	.904	.973	.797	.892	.978
500	1.00	1.00	1.00	.907	.958	.994	.893	.948	.989	.866	.929	.984	.848	.928	.985
50	.319	.042	.002	.173	.085	.015	.211	.110	.028	.330	.247	.108	.653	.783	.933
AR 100	.825	.610	.141	.671	.473	.150	.582	.391	.124	.584	.470	.228	.728	.848	.957
200	.999	.998	.972	.924	.961	.923	.918	.943	.831	.896	.920	.812	.801	.897	.974
500	1.00	1.00	1.00	.931	.965	.993	.921	.965	.994	.915	.957	.994	.847	.922	.985

The cells show the proportion of 10,000 replications with estimated rank $\widehat{r} = r$ obtained using GBH's procedure for significance levels $\alpha = \{.10, .05, .01\}$. Lag length of ADF tests is chosen using the BIC. The innovation vector u_t is generated as $u_{j,t} = \varepsilon_{j,t}$ (WN) and $u_{j,t} = 0.8u_{j,t-1} + \varepsilon_{j,t}$ (AR), $j = 1, ..., 4$, with Gaussian ε_t such that $E(\varepsilon_t) = 0$, $Var(\varepsilon_{i,t}) = 1$ and $Cov(\varepsilon_{i,t}, \varepsilon_{j,t}) = 0.4$ (panel A); $= 0.8$ (panel B); $= -0.3$ (panel C), $i \neq j$

Table 6

Performance of Johansen's procedure with correlated errors - Correct rank estimation

r		4			3			2			1			0		
$n \backslash \alpha$		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
		Panel A. $Cov(\varepsilon_{i,t}, \varepsilon_{j,t})$= 0.4, $i \neq j$														
	50	1.00	1.00	.998	.940	.965	.949	.934	.943	.849	.931	.946	.890	.925	.960	.983
WN	100	1.00	1.00	1.00	.932	.968	.993	.945	.974	.995	.945	.977	.996	.951	.974	.994
	200	1.00	1.00	1.00	.927	.964	.992	.937	.965	.993	.933	.961	.992	.921	.958	.991
	500	1.00	1.00	1.00	.901	.949	.990	.899	.946	.990	.893	.940	.989	.876	.937	.986
	50	.015	.004	.001	.042	.028	.024	.203	.151	.078	.297	.318	.241	.085	.152	.336
AR	100	.516	.237	.015	.118	.053	.008	.201	.117	.032	.408	.365	.203	.414	.561	.803
	200	.998	.994	.865	.729	.597	.219	.593	.464	.193	.616	.561	.344	.676	.798	.933
	500	1.00	1.00	1.00	.897	.950	.991	.889	.940	.968	.853	.917	.936	.802	.888	.968
		Panel B. $Cov(\varepsilon_{i,t}, \varepsilon_{j,t})$= 0.8, $i \neq j$														
	50	.652	.471	.181	.401	.291	.095	.403	.320	.159	.463	.463	.343	.335	.434	.576
WN	100	.849	.693	.271	.195	.359	.107	.465	.353	.134	.567	.532	.369	.503	.622	.820
	200	.999	.986	.826	.810	.749	.401	.690	.616	.337	.697	.694	.516	.672	.793	.927
	500	1.00	1.00	1.00	.898	.951	.990	.880	.942	.987	.854	.927	.986	.832	.905	.976
	50	.029	.012	.003	.118	.105	.077	.231	.216	.173	.158	.207	.283	.017	.031	.075
AR	100	.053	.013	.000	.061	.026	.003	.241	.168	.064	.351	.409	.399	.072	.123	.288
	200	.491	.248	.022	.203	.102	.017	.286	.200	.069	.336	.364	.413	.159	.198	.284
	500	1.00	1.00	.997	.899	.943	.865	.665	.720	.810	.291	.340	.441	.037	.053	.084
		Panel C. $Cov(\varepsilon_{i,t}, \varepsilon_{j,t})$= -0.3, $i \neq j$														
	50	.998	.998	.992	.928	.956	.952	.926	.945	.871	.919	.954	.975	.896	.934	.971
WN	100	1.00	1.00	.999	.911	.958	.986	.907	.953	.972	.881	.941	.981	.854	.913	.977
	200	1.00	1.00	1.00	.900	.948	.989	.886	.945	.983	.865	.929	.975	.826	.909	.976
	500	1.00	1.00	1.00	.897	.946	.989	.889	.942	.986	.863	.928	.985	.829	.906	.979
	50	.019	.004	.001	.048	.028	.022	.225	.170	.100	.308	.323	.248	.089	.162	.348
AR	100	.430	.186	.009	.129	.060	.009	.207	.132	.033	.416	.379	.206	.400	.552	.781
	200	.965	.880	.424	.666	.505	.190	.499	.367	.132	.563	.524	.313	.584	.720	.894
	500	1.00	1.00	.998	.908	.943	.889	.857	.866	.696	.803	.833	.727	.740	.848	.954

The cells show the proportion of 10,000 replications with estimated rank $\widehat{r} = r$ obtained using Johansen's trace test for significance levels $\alpha = \{.10, .05, .01\}$. Lag length of VAR is chosen using the BIC. The innovation vector u_t is generated as $u_{j,t}$= $\varepsilon_{j,t}$ (WN) and $u_{j,t}$= $0.8u_{j,t-1}+\varepsilon_{j,t}$ (AR), $j = 1, ..., 4$, with Gaussian ε_t such that $E(\varepsilon_t) = 0$, $Var(\varepsilon_{i,t})$ = 1 and $Cov(\varepsilon_{i,t}, \varepsilon_{j,t})$= 0.4 (panel A); = 0.8 (panel B); = -0.3 (panel C), $i \neq j$

Table 7 Additional analyses

Panel A. Large systems with a mixture of $I(1)$ and $I(0)$ variables											
		$r=9$ $p_1=6$		$r=8$ $p_1=6$		$r=7$ $p_1=7$		$r=6$ $p_1=7$		$r=5$ $p_1=8$	
$\alpha=.05$	n	GBH	Joh	GBH	Joh	GBH	Joh	GBH	Joh	GBH	Joh
$\widehat{r}=r$	500	.953	.955	.943	.943	.948	.867	.947	.747	.946	.651
	1000	.960	.950	.962	.948	.958	.942	.958	.935	.964	.921
		$r=4$ $p_1=8$		$r=3$ $p_1=9$		$r=2$ $p_1=9$		$r=1$ $p_1=10$		$r=0$ $p_1=10$	
$\alpha=.05$	n	GBH	Joh	GBH	Joh	GBH	Joh	GBH	Joh	GBH	Joh
$\widehat{r}=r$	500	.946	.635	.939	.606	.925	.574	.914	.486	.794	.335
	1000	.959	.902	.956	.875	.959	.823	.948	.780	.885	.709

Panel B. "Pitfalls" of Johansen's procedure													
		B.1: Roots larger than 1						B.2: Fractional roots					
		GBH			Joh			GBH			Joh		
$\widehat{r}$	$n\backslash^{\alpha}$	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
	50	.309	.225	.095	.007	.06	.006	.218	.141	.046	.053	.051	.041
	100	.269	.174	.057	.000	.000	.000	.129	.073	.025	.009	.006	.003
1	200	.203	.119	.031	.000	.000	.000	.055	.027	.007	.001	.002	.001
	500	.156	.087	.024	1.00	1.00	.999	.022	.008	.001	.191	.152	.089
	1000	.146	.096	.028	1.00	1.00	1.00	.009	.004	.000	.281	.246	.189
	50	.609	.740	.901	.993	.994	.994	.743	.846	.953	.942	.947	.958
	100	.682	.808	.942	1.00	1.00	1.00	.861	.924	.975	.991	.994	.997
0	200	.759	.868	.968	1.00	1.00	1.00	.942	.972	.993	.999	.998	.999
	500	.769	.885	.973	.000	.000	.001	.977	.992	.999	.807	.847	.911
	1000	.655	.816	.958	.000	.000	.000	.990	.996	1.00	.701	.740	.804

Panel A. System with $p=10$ variables ($p_1$of them are $I(1)$). The cells show the proportion of replications (out of 10,000) where $\widehat{r}=r$ obtained by GBH and Johansen's trace (Joh) test procedures. Number of lags in GBH chosen by BIC and set to 1 in Joh. $u_{j,t}=0.8u_{j,t-1}+\varepsilon_{j,t}$, $j=1,...,10$, with Gaussian ε_t, $E(\varepsilon_t)=0$, $Var(\varepsilon_t)=I_{10}$. Panel B: The cells show the proportion of replications (out of 10,000) where the estimated rank is $\widehat{r}=0$ (true rank) or $\widehat{r}=1$ (spurious cointegration) applying the GBH and Johansen's (Joh) test procedures in two "pitfall" settings. Number of lags in GBH and Joh chosen by BIC. Panel B.1: $z_{j,t}=1.01z_{j,t-1}+\varepsilon_{j,t}$, $j=1,2,3,4$, with Gaussian ε_t, $E(\varepsilon_t)=0$, $Var(\varepsilon_t)=I_4$. Panel B.2: $z_{j,t}=(1-L)^{-1.4}\{\varepsilon_{j,t}1(t>0)\}$, $j=1,2,3,4$, with Gaussian ε_t such that $E(\varepsilon_t)=0$, $Var(\varepsilon_t)=I_4$.

Table 8. Cointegration in interest rates - GBH procedure test steps

Example 1: short-term maturities						
		i_1	i_3	i_9	i_{12}	i_{15}
$H(1)$	$\widehat{\tau}_m$	-2.15	-2.24	-1.84	-2.64*	-2.75
	$\widehat{\kappa}_m$	.187	.166	.147	.143	.133
	$\widehat{\theta}_m$	4.61	5.04	3.41	6.95	7.57
	$\widehat{\theta}^*_m$	.862	.837	.501	.994	1.01
$H(2)$	$\widehat{\tau}_{c_1,m}$	-3.97***	-3.50**	-4.16***	-5.84***	–
Example 2: short and long-term maturities						
		i_1	i_3	i_{18}	i_{48}	i_{120}
$H(1)$	$\widehat{\tau}_m$	-2.15	-2.24	-2.85*	-3.11**	-2.89**
	$\widehat{\kappa}_m$	.187	.166	.126	.089	.072
	$\widehat{\theta}_m$	4.61	5.04	8.13	9.68	8.34
	$\widehat{\theta}^*_m$	.862	.837	1.02	.862	.600
$H(2)$	$\widehat{\tau}_{c_1,m}$	-1.98	-1.04	-1.84	-2.08	–
	$\widehat{\kappa}_{c_1,m}$	.617	.552	.418	.358	–
	$\widehat{\theta}_{c_1,m}$	3.92	1.09	3.39	4.34	–
	$\widehat{\theta}^*_{c_1,m}$	2.42	.597	1.42	1.55	–
$H(3)$	$\widehat{\tau}_{c_1,c_2,m}$	-9.57***	–	-4.40***	-4.44***	–

* Significant at 10%; ** Significant at 5%; *** Significant at 1%. $H(1) : \widehat{\tau}_m$ are ADF tests for the null $i_{m,t} \sim I(1)$. Automatic lag selection (BIC) in an ADF regression with intercept. Critical values: –2.57 (10%) -2.86 (5%) -3.44 (1%). $\widehat{\kappa}_m$ is the inverse of the variance ratio statistic for $i_{m,t}$, $\widehat{\kappa}_m = n\Sigma_{t=1}^{n}(i_{m,t}-\bar{i}_m)^2 / \Sigma_{t=1}^{n}\left(\Sigma_{s=1}^{t}(i_{m,s}-\bar{i}_m)\right)^2$, where $\bar{i}_m = n^{-1}\Sigma_{t=1}^{n} i_{m,t}$. $\widehat{\theta}_m = \widehat{\tau}^2_m$, $\widehat{\theta}^*_m = \widehat{\theta}_m \widehat{\kappa}_m$. $H(2) : \widehat{\tau}_{c_1,m}$ are ADF tests on residuals of the regression $i_{m,t} = a_m + b_m c_{1,t} + u_{m,t}$,where $c_{1,t}$ is the first variable chosen as common trend; the ADF tests are computed with no intercept or trend and automatic lag selection (BIC). Critical values from Phillips and Ouliaris (1990): –3.07 (10%) -3.37 (5%) -3.96 (1%). $\widehat{\kappa}_{c_1,m}$ is the inverse of the variance ratio statistic applied to the residuals of the regression; $\widehat{\theta}_{c_1,m} = \widehat{\tau}^2_{c_1,m}$, $\widehat{\theta}^*_{c_1,m} = \widehat{\theta}_{c_1,m}\widehat{\kappa}_{c_1,m}$. $H(3) : \widehat{\tau}_{c_1,c_2,m}$ are ADF tests on residuals of the regression $i_{m,t} = a_m + b_{m,1} c_{1,t} + b_{m,2} c_{2,t} + u_{m,t}$, where $c_{2,t}$ is the second variable chosen as common trend; the ADF tests are computed with no intercept or trend and automatic lag selection (BIC). Critical values from Phillips and Ouliaris (1990): –3.45 (10%) -3.77 (5%) -4.31 (1%).

Table 9. Cointegration in interest rates - Estimates of the cointegrating vectors

Example 1: only short maturities				
		$\widehat{a}_m$	$\widehat{b}_m$	
i_1	OLS	−0.22 (0.17)	0.89 (0.03)	
	DOLS	−0.49 (0.26)	0.93 (0.05)	
	Johansen	-0.49 (0.48)	0.94 (0.02)	
i_3	OLS	−0.31 (0.12)	0.95 (0.02)	
	DOLS	−0.54 (0.22)	0.99 (0.04)	
	Johansen	−0.56 (0.40)	1.00 (0.06)	
i_9	OLS	−0.23 (0.05)	0.99 (0.01)	
	DOLS	−0.32 (0.10)	1.00 (0.02)	
	Johansen	−0.30 (0.15)	1.00 (0.02)	
i_{12}	OLS	−0.13 (0.03)	1.00 (0.01)	
	DOLS	−0.18 (0.06)	1.00 (0.01)	
	Johansen	−0.16 (0.09)	1.00 (0.01)	
Example 2: short and long maturities				
		$\widehat{a}_m$	$\widehat{b}_{m,3}$	$\widehat{b}_{m,120}$
i_1	OLS	0.16 (0.09)	1.00 (0.02)	−0.06 (0.02)
	DOLS	0.16 (0.13)	1.00 (0.02)	−0.06 (0.02)
	Johansen	0.25 (0.10)	1.00 (0.02)	−0.07 (0.02)
i_{18}	OLS	−0.33 (0.10)	0.75 (0.02)	0.33 (0.02)
	DOLS	−0.14 (0.15)	0.78 (0.03)	0.29 (0.03)
	Johansen	−0.12 (0.22)	0.79 (0.04)	0.28 (0.04)
i_{48}	OLS	−0.33 (0.08)	0.36 (0.01)	0.71 (0.01)
	DOLS	−0.13 (0.15)	0.38 (0.03)	0.66 (0.03)
	Johansen	−0.05 (0.20)	0.40 (0.04)	0.64 (0.05)

OLS: OLS estimates from regressions of the form $i_{m,t}= a_m+b_m i_{15,t}+u_{m,t}$, $m = 1, 3, 9, 12$ (Ex1) and $i_{m,t}= a_m+b_{m,3}i_{3,t}+ + b_{m,120}i_{120,t}+u_{m,t}$, $m = 1, 18, 48$ (Ex2); DOLS: dynamic OLS estimates where four lags and leads of $\Delta i_{15,t}$ (Ex1) and of $\Delta i_{3,t}$ and $\Delta i_{120,t}$ (Ex2) are included as additional regressors (Stock and Watson, 1993). Regular (OLS) and Newey-West (DOLS) standard errors in parentheses. Johansen: Likelihood-based estimates of the $(p \times r)$ matrix of cointegrating vectors β in the reduced-rank system $\Delta z_t= \alpha a + \Pi z_{t-1}+ \sum_{j=1}^{4} \Gamma_j \Delta z_{t-j}+u_t$, where $z_t= (i_{1,t}, i_{3,t}, i_{9,t}, i_{12,t}, i_{15,t})'$ (Ex1) and $z_t= (i_{1,t}, i_{3,t}, i_{18,t}, i_{48,t}, i_{120,t})'$ (Ex2). Standard errors in parentheses. Cointegrating vectors are normalized to have coefficient equal to one in the interest rate in the row and the signs of the estimates are multiplied by (-1) to facilitate comparison with OLS and DOLS estimates.