Subsampling Cointegration Ranks in Large Systems

Pu Chen and Hsiao Chihying
Faculty of Economics
Bielefeld University
33501 Bielefeld, Germany
pchen@wiwi.uni-bielefeld.de
chsiao@wiwi.uni-bielefeld.de

Aug. 2005

Abstract

In this paper we investigate the possibility of the application of subsampling procedure for testing cointegration relations in large multivariate systems. The subsampling technique is applied to overcome the difficulty of nonstandard distribution and nuisance parameters in testing for cointegration rank without an explicitly formulated structural model. The contribution in this paper is twofold: theoretically this paper shows that the subsampling testing procedure is consistent and has asymptotically power 1; practically this paper demonstrates that the subsampling procedure can be applied to determine the cointegration rank in large scale models, where the standard procedures hits already its limit. For empirical relevant cases our simulation studies show that centered subsampling improves decisively the performance of subsampling test procedure and makes it applicable also for cases when the number of independent stochastic trends are very large.

KEY WORDS: Cointegration, Large Systems, Nonparametric Tests, Subsampling, PPP

1 Introduction

Over last decade considerable attention has been paid in empirical economics to testing for the existence of long-run relations by using cointegration analysis. There have been two main approaches: the two step residual-based procedure for testing either the null of no cointegration[1] or the null of cointegration[2] and the system-based reduced rank regression approach [3]. The residual-based approach is applicable to analyzing single cointegration relations between variables[4]. The system-based approach is applicable to small systems with a few variables due to its data intensive specifications of the error correction models[5].

Recently some non-parametric and semi parametric approaches are developed to test the cointegration relations. Bierens (1997) proposes a testing procedure that is free from the specification of the data-generating process. Breitung (2002) proposes a testing procedure that is invariant against log transformation which is often used in the empirical applications. Boswijk and Lucas (2002) aim at rising the power of testing procedure. More recently Bai and Ng (2004) propose a test of cointegration relations by analyzing the number of nonstationary factors in large systems. However, this approach works only for cases where the system is large but the numbers of independent stochastic trends are vary small. Following this line of research we propose a nonparametric subsampling approach to test of cointegration relations, targeting at handling cointegration analysis in large systems with moderate demand of data to identify either small or large numbers of independent stochastic trends.

Large systems have always been the interest of research. For instance, finding the number of independent stochastic trends in stock prices or bond prices that all behavior like stochastic trends will help to gain more insight into the driving forces of the financial markets, and knowing the number of independent stochastic trends among main economic indicators of an economy will help to understand the structure of the economy. But, the present standard procedure of multivariate cointegration analysis - the Johansen procedure can not applied to large systems. The basic reason is that the Johansen procedure requires a completely specified vector error model(ECM). In case of large scale systems this will certainly quickly use up degree of freedom for practically available economic data. Furthermore, the large number of

[1] See Engle and Granger (1987), Phillips and Quliaris (1990)

[2] See Kwiatkowsi, Phillips, Schmidt, and Shin (1992)

[3] Johansen (1991) and Johansen (1995)

[4] For problems of the application of this approach to many variables see Hamilton (1994)

[5] Empirical systems analyzed by this approach usually do not exceed the dimensionality of 8. Even the most used softwares provide only critical values of Johansen procedure up to the dimensionality of 12.

parameters in large VECM makes the statistical inference very unreliable.

The subsampling procedure for testing cointegration rank turns out in many cases to be simple, robust and has correct size. For systems up to 40 variables with a moderate requirement on data of 200 observations subsampling can provide reliable results for testing of cointegration ranks, while the standard procedure such as the Johansen procedure runs into problem already at the dimension of 20. The plan of this paper is as follows: Section 2 sets out the underlying model and motivates the basic idea of the testing procedure. Section 3 and Section 4 describe the testing procedure and consider the properties of the test statistics. Section 5 presents some simulation results. In Section 6 we give an empirical application. Section 7 provides some concluding remarks. Technical details are included in the Appendix.

2 The Assumption of the DGP and the Motivation of the Test

Let the $n \times 1$ time series y_t be specified as follows:

$$y_t = C\xi_t + u_t, \tag{2.1}$$

where u_t is $n \times 1$ stationary time series

$$u_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \tag{2.2}$$

with $\sum_{j=0}^{\infty} |\psi_j| < \infty$; and ξ_t is $g \times 1$ random walk [6]

$$\xi_t = \sum_{s=1}^{t} \epsilon_{2,t},$$

where $(\epsilon_t, \epsilon_{2,t})$ are jointedly gauss-distributed with $\epsilon_t \sim N(0, I_n)$, $\epsilon_{2,t} \sim N(0, I_g)$ and $\Omega = Cov(\epsilon_t, \epsilon'_{2,t})$. C is a $n \times g$ matrix with rank g. This model has g independent stochastic trends. Let β be $n \times h$ matrix ($h := n - g$) such that $\beta' C = 0$. β is called *(one) basis of cointegration vectors* because

$$\beta' y_t = \beta' u_t \tag{2.4}$$

[6] This model can accommodate the model:

$$\xi_t = \sum_{s=1}^{t} u_{2,t}$$

with $u_{2,t} = \sum_{j=0}^{\infty} \psi_{2,j} \epsilon_{2,t-j}$, $\sum_{j=0}^{\infty} j|\psi_{2,j}| < \infty$ by using the Beveridge-Nelson decomposition, see Hamilton (1994) P.504. In this context the formulation (2.1) encompasses the error correction model. To illustrate this we consider following error correction representation of a cointegrated system:

$$\Delta y_t = \beta \alpha' y_{t-1} + \pi \Delta y_{t-1} + \epsilon_t. \tag{2.3}$$

This model can be reformulated as follows:

$$\begin{aligned} y_t &= (I + \beta\alpha')y_{t-1} + \pi\Delta y_{t-1} + \epsilon_t \\ &= (\alpha^*, \alpha^*_\perp) \begin{pmatrix} \alpha' \\ \alpha'_\perp \end{pmatrix} y_{t-1} + \beta\alpha' y_{t-1} + \pi\Delta y_{t-1} + \epsilon_t \\ &= \alpha^*_\perp \alpha_\perp y_{t-1} + (\alpha^* + \beta)\alpha' y_{t-1} + \pi\Delta y_{t-1} + \epsilon_t \\ &= C\xi_t + u_t \end{aligned}$$

where $C = \alpha^*_\perp$ is a $n \times g$ matrix; $\xi_t = \alpha'_\perp y_{t-1}$ is $g \times 1$ stochastic trend; and $u_t = (\alpha^* + \beta)\alpha' y_{t-1} + \pi\Delta y_{t-1} + \epsilon_t$ is $n \times 1$ stationary series.

is stationary. Thus, this system has h cointegration relations[7]. The correspondence of the cointegration representation (2.1) and the standard representations, for example, the Phillips's triangular representation, the Stock-Watson common trends representation and the error correction representation can be seen in Hamiltion (1994) Chap.18. We choose the representation of (2.1) and (2.4) to pave the way for a test without specifying a structural model for the DGP.

(2.1) implies that cointegration is basically a phenomena of multicolinearity up to a I(0) residuals. Hence the rank of the regression coefficient matrix of the regression of Δy_t on y_{t-1} should reveal the number independent stochastic trends. Difficulty in finding a decision rule to determine the number of independent stochastic trends lies in the description of the distribution of the statistics on the rank of the coefficient matrix, because it may not have a standard distribution and may depend on nuisance parameters. This issue is filly discussed in Stock and Watson (1988), where they construct the statistical tests by looking at the corrected first order series correlation matrix of the observed data, which yields tabulable standard distributions. Johansen (1991) and Johansen (1995) take the maximum likelihood approach and construct tabulable statistical tests by controlling for correlation of the I(0) components through a completely specified structural model for the DGP.

Alternatively, we apply subsampling method to construct a subsampling confidence interval for the test, such that we avoid the problem of the description of the distribution of the test statistic by drawing from the "true" distribution- the DGP.

The first step of the procedure is to regress Δy_t on y_{t-1} by OLS like follows:

$$\Delta y_t = \Pi y_{t-1} + e_t \tag{2.5}$$

We get the OLS estimator $\hat{\Pi}$:

$$\hat{\Pi}' = \left(\sum_{t=1}^{T} y_{t-1} y_{t-1}'\right)^{-1} \left(\sum_{t=1}^{T} y_{t-1} \Delta y_t'\right) \tag{2.6}$$

This regression equation differs from the error correction model of Johansen procedure only in that it omits the specification of the short run dynamics. Owing to the super consistence of LS in regression with I(1) variables, we would nevertheless expect that the rank of $\hat{\Pi}$ would converge to the cointegration rank $n - g$. Thus the information about the rank of $\hat{\Pi}$, or equivalently,

[7]Following the general characterization of cointegrating vector in Hamilton (1994) P. 547, (2.1) and (2.4) consist a cointegrating system.

the number of zero eigenvalues of $\hat{\Pi}$ can be used to test the cointegration rank.

The advantage of Johansen procedure is that through controlling for the short run dynamics it achieves a test statistic based on the rank of the regression matrix of Δy_t on y_{t-1}[8] that is free from nuisance parameter and hence tables of critical values for the test statistics can be calculated. However, the full specification of the short run dynamics will quickly use up the degree of freedom for large systems, such that Johansen procedure is only practically applicable for small systems due to the availability of data for empirical research in the real world[9].

We sacrifice the specification of the short run dynamics for the ability to handle large cointegration systems at the same availability of data and hope that the super consistence of $\hat{\Pi}$ will nevertheless provides a good estimate for the calculation of the rank of Π. The difficulty is that the resulting test statistic of the rank of $\hat{\Pi}$ will be nonstandard and depends on nuisance parameters. We overcome this problem by using subsampling method that is immune against nonstandard distribution and the presence of nuisance parameters. This is the motivation behind the test procedure.

3 Property of the LS Regression

The last section suggests that we could infer the number of the cointegration relations from the rank of the estimated matrix $\hat{\Pi}$

$$\hat{\Pi}' = \big(\sum_{t=0}^{T-1} y_t y_t'\big)^{-1}\big(\sum_{t=0}^{T-1} y_t \Delta y_{t+1}'\big). \tag{3.7}$$

where $\Delta y_t = C\epsilon_{2,t} + u_t - u_{t-1} =: v_t$. In this section the distribution of the eigenvalues of $\hat{\Pi}$ will be derived. We will see that the eigenvalues can be separated into two groups which have different order of convergence as $T \to \infty$; one group corresponds to the cointegration space and the other group corresponds to the space of stochastic trends. Later we will take advantage of this fact to construct our test statistics.

Let $\beta_\perp$ be a $n \times g$ matrix including a basis of the complement subspace to β. Let $B = (\beta, \beta_\perp)$.

[8] The the matrix of canonical correlation coefficient has the same rank as the matrix $\hat{\Pi}$.

[9] Furthermore, for large systems small sample property of the Johansen test statistics dependent sensitively on the number of observations, especially when the number of independent stochastic trend is large. In these cases the Johansen test is not reliable (See Section 4).

First we look at the following terms

$$
\begin{aligned}
B'\sum_{t=0}^{T-1} y_t y_t' B &= \begin{pmatrix} \beta' \\ \beta_\perp' \end{pmatrix} \sum_{t=0}^{T-1} (C\xi_t\xi' C' + u_t\xi_t' C' + C\xi_t u_t' + u_t u_t') \begin{pmatrix} \beta & \beta_\perp \end{pmatrix} \qquad (3.8)\\
&= \begin{pmatrix} \beta'(\sum_{t=0}^{T-1} u_t u_t')\beta & \beta'(\sum_{t=0}^{T-1} u_t\xi_t' C' + \sum_{t=0}^{T-1} u_t u_t')\beta_\perp \\ \beta_\perp'(\sum_{t=0}^{T-1} C\xi_t u_t'\beta + \sum_{t=0}^{T-1} u_t u_t')\beta & \beta_\perp' \sum_{t=0}^{T-1}(C\xi_t + u_t)(\xi_t' C' + u_t')\beta_\perp \end{pmatrix} \\
&=: \begin{pmatrix} G_{1,T} & G_{2,T} \\ G_{3,T} & G_{4,T} \end{pmatrix}
\end{aligned}
$$

and

$$
\begin{aligned}
B'\sum_{t=0}^{T-1} y_t \Delta y_{t+1}' &= \begin{pmatrix} \beta' \\ \beta_\perp' \end{pmatrix} \sum_{t=0}^{T-1} (C\xi_t v_{t+1}' + u_t v_{t+1}') \qquad (3.9)\\
&= \begin{pmatrix} \beta \sum_{t=0}^{T-1} u_t v_{t+1}' \\ \beta_\perp' \sum_{t=0}^{T-1}(C\xi_t + u_t) v_{t+1}' \end{pmatrix} \\
&=: \begin{pmatrix} G_{5,T} \\ G_{6,T} \end{pmatrix}.
\end{aligned}
$$

Before we discuss the limit distributions of the matrices above we define $w_t = \sum_{s=0}^{\infty} \phi_{1,s}\epsilon_{t-s} + \sum_{s=0}^{\infty} \phi_{2,s}\epsilon_{2,t-s}$, where w_t is $n \times 1$ vector, $\phi_{1,s}$ are $n \times n$ matrices and $\phi_{2,s}$ are $n \times g$ matrices. Recall that the $n \times g$ matrix Ω is the covariance matrix of ϵ_t and $\epsilon_{2,t}$. Let $\xrightarrow{P}$ denote the convergence in probability and $\stackrel{D}{\Longrightarrow}$ denote the convergence in distribution-

Lemma 3.1 *(i)*

$$
\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} u_t w_t' \xrightarrow{P} \sum_{j=0}^{\infty} \psi_j \phi_{1,j}' + \sum_{j=0}^{\infty} \psi_j \Omega \phi_{2,j}'.
$$

(ii)

$$
\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \xi_t w_t' = \Omega'\phi_1(1)' + \phi_2(1)' + \int_0^1 W_{2,s} dW_s' \phi_1(1)' + \int_0^1 W_{2,s} dW_{2,s}' \phi_2(1)',
$$

where $(W_s, W_{2,s})$ *are jointedly Brownian motion with the same correlation as* $(\epsilon_s, \epsilon_{2,s})$ *and* $\phi_i(1) = \sum_{j=0}^{\infty} \phi_{i,j}$ *for* $i = 1, 2$.

(iii)

$$
\lim_{T\to\infty} \frac{1}{T^2} \sum_{t=0}^{T-1} \xi_t \xi_t' = \int_0^1 W_{2,s} W_{2,s}' ds.
$$

Proof
The statements (i), (ii) and (iii) follow the results of (f), (h) and (b) of Lemma 1 in Sims, Stock, and Watson (1990) P.121. □

Here we introduce two notations:

$$G_T = O_C(T^\alpha) \quad \text{iff} \quad \lim_{T\to\infty} \frac{G_T}{T^\alpha} \xrightarrow{P} g \ \ (g \text{ is a constant and } g \neq 0),$$
$$G_T = O_D(T^\alpha) \quad \text{iff} \quad \lim_{T\to\infty} \frac{G_T}{T^\alpha} \xRightarrow{D} G \ \ (G \text{ is a random variable and } G \neq 0) \ .$$

Remark $O_C(T^\alpha)$ can be seen as a special case of $O_D(T^\alpha)$ where $O_C(T^\alpha)/T^\alpha$ converges to a degenerated distribution (constant). Sometimes we do not distinguish these two cases if the difference is not relevant.

Using Lemma 3.1 we have

Property 3.2

$$\begin{array}{ll} G_{1,T} = O_C(T) & G_{2,T} = O_D(T) \\ G_{3,T} = O_D(T) & G_{4,T} = O_D(T^2) \\ G_{5,T} = O_C(T) & G_{6,T} = O_D(T) \end{array}$$

□

Let G_i be the limit (as a distribution) of $G_{i,T}$ for $i = 1, \cdots, 6$, that means

$$G_i := \lim_{T\to\infty} \frac{G_{i,T}}{T^{\alpha_i}} \quad \text{in distribution.}$$

Lemma 3.3 *$G_{i,T} = O_D(T^{\alpha_i})$ for $i = 1, \cdots, 6$ are defined in (3.8) and (3.9). G_i are the limit as described above. Then it holds*
(i) For $G_{i,T}$ and G_i invertible,

$$G_{i,T}^{-1} = O_D(T^{-\alpha_i})$$

(ii) For $G_{i,T}$, $G_{j,T}$ such that $G_i \neq 0$, $G_j \neq 0$ and that $G_{i,T} \cdot G_{j,T}$ and $G_i \cdot G_j$ are well-defined matrix multiplication and $G_i \cdot G_j \neq 0$, we have then

$$G_{i,T} \cdot G_{j,T} = O_D(T^{\alpha_i+\alpha_j}).$$

(iii) If $\alpha_i > \alpha_j$, then

$$G_{i,T} + G_{j,T} = O_D(T^{\alpha_i}).$$

Proof See Appendix.

Remark The results (i), (ii) and (iii) hold for $G_{i,T}$ as defined above, because these sequences $G_{i,T}$ are continuous functionals of the white noise $(\epsilon_t, \epsilon_{2,t})$

so that we can apply the continuous mapping theorem to them[10]. As the result, the product of two sequences converges to the product of the limits in distribution as (ii) in the Lemma above. Similarly it holds for inversion and summation as in (i) and (iii).

Let

$$\begin{pmatrix} G_{7,T} \\ G_{8,T} \end{pmatrix} := \left(B' \sum_{t=0}^{T-1} y_t y_t' B\right)^{-1} B' \sum_{t=0}^{T-1} y_t v_{t+1}' = \begin{pmatrix} G_{1,T} & G_{2,T} \\ G_{3,T} & G_{4,T} \end{pmatrix}^{-1} \begin{pmatrix} G_{5,T} \\ G_{6,T} \end{pmatrix} = B^{-1}\hat{\Pi}.$$

We denote particularly G_1 and G_5 as g_1 and g_5 in small capital to emphasize that the limits are constants instead of (non-degenerated) random variables. Then we have:

Lemma 3.4

$$\begin{aligned} \lim_{T\to\infty} G_{7,T} &\xrightarrow{P} g_1^{-1} g_5 =: g_7 \\ \lim_{T\to\infty} TG_{8,T} &\xRightarrow{D} -g_1 G_4^{-1} G_3 g_5 + G_6 G_4 =: G_8. \end{aligned}$$

Proof See Appendix.

Following Lemma 3.4 $T\hat{\Pi}$ behaves asymptotically like follows:

$$T\hat{\Pi} = B \begin{pmatrix} TG_{7,T} \\ TG_{8,T} \end{pmatrix} \sim B \begin{pmatrix} Tg_7 \\ G_8 \end{pmatrix}. \tag{3.10}$$

Now we are interested in the distribution of the *eigenvalues* of $T\hat{\Pi}$. We present the following theorem:

Theorem 3.5 *Let $\lambda_{1,T}, \cdots, \lambda_{n,T}$ be the eigenvalues of $T\hat{\Pi}$. Then there are h eigenvalues for $i = 1, \cdots, h$ such that*

$$\lim_{T\to\infty} \frac{\lambda_{i,T}}{T} \xrightarrow{P} \lambda_i \quad \textit{constant} \tag{3.11}$$

or equivalently

$$\lim_{T\to\infty} \lambda_{i,T} \xrightarrow{P} \infty \tag{3.12}$$

the rest g eigenvalues for $i = h+1, \cdots, n$

$$\lim_{T\to\infty} \lambda_{i,T} \xRightarrow{D} \lambda_i \quad \textit{distributions.} \tag{3.13}$$

Proof See Appendix.

[10] See Hamiltion (1994) P.482.

Corollary 3.6 *We have h eigenvalues, for* $i = 1, 2, ...h$,

$$\lim_{T\to\infty} \frac{|\lambda_{i,T}|}{T} \xrightarrow{P} |\lambda_i| \quad \textit{constants} \tag{3.14}$$

and the rest g eigenvalues, for $i = h + 1, \cdots, n$,

$$\lim_{T\to\infty} |\lambda_{i,T}| \xRightarrow{D} |\lambda_i| \quad \textit{distributions.} \tag{3.15}$$

4 The Test Procedure

The hypothesis of the existence of g independent stochastic trends or equivalently h cointegration relations implies that the g in absolute value smallest eigenvalues of $T\hat{\Pi}$ converge to some random variables λ_i, $(i = 1, 2, ...h)$ in distribution while other h eigenvalues will diverge. (3.12) and (3.13) tell us if we were able to construct a $1-\alpha$ percent confidence interval based on the distribution of $\lambda_i, i = 1, 2, ..., g$, then $1-\alpha$ percent of the g smallest eigenvalues of the matrix $T\hat{\Pi}$ will lie in the corresponding confidence interval, while the other h eigenvalues will asymptotically escape from any empirical confidence interval constructed by using subsamples. This difference between the divergence of the h largest eigenvalues of $T\hat{\Pi}$ and the convergence of the g smallest eigenvalues yields a sharp separation between these two groups of eigenvalues. In this sense this test of the cointegration rank has asymptotical power one.

Using Theorem 4.2.1 of Politis, Romano, and Wolf (1999), and the results of Theorem 3.5 we can get asymptotic valid test results.

Following is the result from Politis et al. (1999):

Let $\hat{\theta}_T = \hat{\theta}(Y_1, Y_2, ...Y_T)$ be an estimator of $\theta(P) \in \mathbb{R}$, the parameter of interest. Let $\hat{\theta}_{T,b,t} = \hat{\theta}_b(Y_t, Y_{t+1}, ...Y_{t+b-1})$ the estimator of $\theta(P)$ based on the subsample $\{Y_t, ...Y_{t+b-1}\}$. Define $J_{b,t}(P)$ be the sample distribution of $\tau_b(\hat{\theta}_{n,b,t} - \theta(P))$, where τ_b is an appropriate normalizing constant. Also, define the corresponding cumulative distribution function:

$$J_{b,t}(x, P) = Prob_P\{\tau_b(\hat{\theta}_{T,b,t} - \theta(P)) \leq x\}$$

For notation convenience, let $J_T(P) = J_{1,T}(P)$ and $J_T(., x) = J_{1,T}(., P)$. Denote the empirical distribution function as follows:

$$L_{T,b}(x) = \frac{1}{T-b+1} \sum_{t=1}^{T-b+1} 1\{\tau_b(\hat{\theta}_{n,b,t} - \hat{\theta}_T) \leq x\}.$$

Essentially the only assumption that will be needed to consistently estimate $J_T(P)$ is the following:

Assumption 4.1 *There exits a limiting law $J(P)$ such that*

- *i. $J_T(P)$ converges weakly to $J(P)$ as $T \to \infty$.*
- *ii. For every continuity point x of $J(P)$ and for any sequences T, b with $T, b \to \infty$ and $b/T \to 0$, we have $\frac{1}{T-b+1}\sum_{t=1}^{T-b+1} J_{b,t}(x,P) \to J(x,P)$*

Theorem 4.2 (subsampling) *Assume Assumption 4.1 and that $\tau_b/\tau_T \to 0$, $b/T \to 0$, and $b \to \infty$ as $T \to \infty$. $\alpha_{T,b}(h) \to 0$ as $T \to \infty$, where $\alpha_{T,b}(.)$ denotes the α-mixing coefficients corresponding to series of the test statistics $\{Z_{T,b,t}, t = 1, 2, ...T - b + 1\}$ with $Z_{T,b,t} = \tau_b(\hat{\theta}_{T,b,t} - \theta(P))$.*

- *i. If x is a continuity point of $J(., P)$, then*

$$L_{b,T}(x) \xrightarrow{P} J(x,P) \tag{4.16}$$

- *ii. If J(.,P) is continuous, then*

$$sup_x|L_{b,T}(x) - J(x,P)| \xrightarrow{P} 0 \tag{4.17}$$

- *iii. For $\alpha \in (0,1)$ let*

$$c_{T,b}(1-\alpha) = inf\{x : L_{T,b}(x) \geq 1-\alpha\}. \tag{4.18}$$

Define:

$$c(1-\alpha, P) = inf\{x : J(x,P) \geq 1-\alpha\}. \tag{4.19}$$

If J(.,P) is continuous at $c(1-\alpha, P)$, then

$$Prob_P\{\tau_T[\hat{\theta_T} - \theta(P)] \leq c_{T,b}(1-\alpha)\} \xrightarrow{P} 1-\alpha \quad as \quad T \to \infty. \tag{4.20}$$

Thus, the asymptotic coverage probability under P of the interval $I_1 = [\hat{\theta}_T - \tau_T^{-1}c_{T,b}(1-\alpha), \infty)$ is the nominal level $1-\alpha$.

Proof: see Politis et al. (1999) P. 273.

Comments:

For the application of the theorem above to testing for cointegration rank, the limiting law are the weakly convergence in (3.13). Because every summand in (*ii*) of Assumption 4.1 has the same limiting distribution for $b \to \infty$, the average has the same limiting distribution. Therefore, we can apply the subsampling procedure to get an asymptotically valid test for the cointegration rank.

Under H_0 of g independent stochastic trends, up to g absolutely smallest eigenvalue of $T\hat{\Pi}$ will lie in the corresponding empirical confidence interval constructed by subsampling with probability $1-\alpha$ while the other h eigenvalues of the test statistic of the whole sample will lie beyond the corresponding "subsampling confidence interval".

The testing procedure:

- Calculation of $T\Lambda(\hat{\Pi})$ as the whole sample test statistic for each eigenvalues, where $\Lambda(\hat{\Pi})$ is the diagonal matrix of the eigenvalues of $\hat{\Pi}$.
- Choice of a proper subsample size b_T [11]
- Calculation of $T-b+1$ subsample test statistics $|b_T\Lambda(\hat{\Pi}_{b_T,t})|$ $(t = b, b+1, ..., T)$ for each eigenvalues of $T\hat{\Pi}$
- Calculation of the empirical distribution function based on the subsample test statistics and calculation of the empirical confidence intervals at a given confidence level α for each norm of eigenvalues of $T\hat{\Pi}$.
- Check of the whole sample test statistic with the corresponding confidence interval and conclude the test result.

Remark 1

We can also calculate the sample canonical correlation coefficient as the test statistic. For this alternative way we have following results. Let Π^n be the matrix to calculate the sample canonical coefficient between ΔY_t and Y_{t-1}.

$$\Pi^n = \left(\sum_{t=0}^{T-1} \Delta y_{t+1}\Delta y_{t+1}'\right)^{-1} \left(\sum_{t=0}^{T-1} \Delta y_{t+1} y_t'\right) \left(\sum_{t=0}^{T-1} y_t y_t'\right)^{-1} \left(\sum_{t=0}^{T-1} y_t \Delta y_{t+1}'\right).$$

We have:

$$\begin{aligned}
& T\Pi^n \\
= \; & T\left(\sum_{t=0}^{T-1} \Delta y_{t+1}\Delta y_{t+1}'\right)^{-1} \left(\sum_{t=0}^{T-1} \Delta y_{t+1} y_t'\right) BB^{-1} \left(\sum_{t=0}^{T-1} y_t y_t'\right)^{-1} B^{-1}B \left(\sum_{t=0}^{T-1} y_t \Delta y_{t+1}'\right) \\
= \; & \begin{pmatrix} O_c(T) & O_c(T) \\ O_c(T) & O_c(T) \end{pmatrix}^{-1} \begin{pmatrix} O_C(T) & O_D(T) \end{pmatrix} \begin{pmatrix} G_{1,T} & G_{2,T} \\ G_{3,T} & G_{4,T} \end{pmatrix} \begin{pmatrix} G_{5,T} \\ G_{6,T} \end{pmatrix} \\
\xrightarrow{L} \; & (Tg_9, G_{10})
\end{aligned}$$

[11] Theoretically any subsample block sizes satisfying the condition $b \to \infty$ and $b/T \to 0$ as $T \to \infty$ will have the same asymptotical result. However the choice of an optimal is sensitive practical issue, because the test results may depend on the choice of the block size. See Politis et al. (1999) and Choi (2003) for detailed discussion.

The convergency in distribution in the last line above follows from the results of Lemma 4.3 and the Slucky theorem[12]. According to the results of Theorem 3.5 we can also apply subsampling procedure to the sample canonical coefficient to test the cointegration rank. One advantage of using canonical correlation coefficient instead of Π is that we do not need to care about the problem of complex eigenvalues, because the matrix used to calculate the canonical correlation coefficients is positive definite, while $\hat{\Pi}$ may contains complex eigenvalues.

Remark 2

In the presentation above we have assumed that there is no drift involved in the stochastic trends. However, the subsampling result can still apply for the cases with drifts, if $g > 1$. In the presence of drifts in the stochastic trends, only the smallest eigenvalue of $\hat{\Pi}$ will converge at the rate $T^{-\frac{3}{2}}$ to zero under null. The convergence rate of other $g-1$ eigenvalues that corresponds to the space of independent stochastic trends are T^{-1}, and the convergence rate of the remaining $n-g$ eigenvalues that corresponds to the cointegration space are $T^{-\frac{1}{2}}$. [13] For the null of $g = 1$ with drift we use the scaling factor $T^{\frac{3}{2}}$ to test for cointegration. Furthermore, we can apply different the scaling factor $T^{\frac{3}{2}}$ or T to the smallest eigenvalue to test the presence of drift in the stochastic trends.

5 Simulation Studies

5.1 Uncentered Subsampling

Subsampling procedure is an asymptotically valid test. For the application of this procedure to empirical research the performance of the this procedure in finite sample size is relevant. Therefore we conduct Monte Carlo simulation to assess the small sample property of the subsampling procedure. We would like to known how well the subsampling performs, when (i) the dimension of the DGP varies, and (ii) the coefficients of autocorrelation in u_t increase.

The DGP for the simulation studies is:

$$y_t = C\xi_t + u_t \tag{5.21}$$

[12] To apply Slucky theorem we need in fact stronger condition, namely the product of convergence in probability and convergence in distribution. We have here the product of convergence in distribution only. However, following Föllmer (1981), the convergence of unit root processes to functionals Brownian motion has a version of strong convergence and hence forth convergence in probability. Therefore we can apply Slusky theorem in our cases.

[13] See Sims et al. (1990) and Hamilton (1994) P.555 for more details.

with

$$\xi_t = \sum_{\tau=1}^{t} e_\tau \tag{5.22}$$

$$e_t = \sum_{\tau=0}^{4} \rho_{e\tau} \epsilon_{et-\tau}, \qquad \epsilon_{et} \sim i.i.d.N(0, I_g) \tag{5.23}$$

and

$$u_t = \sum_{\tau=0}^{4} \rho_\tau \epsilon_{t-\tau}, \qquad \epsilon_t \sim i.i.d.N(0, I_n) \tag{5.24}$$

Table 1 presents the results of simulation studies of the subsampling procedure. The numbers in the table are the relative frequency on non-rejection of the H_0 of zero eigenvalues calculated through 500 replications.

No.	n	g_0	T	b	$H_0 : g = g_0 - 1$	$H_0 : g = g_0$	$H_0 : g = g_0 + 1$
1	15	5	200	90	0.922	0.938	0.00
2	20	5	200	90	0.944	0.944	0.00
3	30	5	200	90	0.944	0.922	0.00
4	40	5	200	90	0.938	0.866	0.00
5	60	5	200	90	0.814	0.688	0.00

Table 1: *Estimated coverage probability of various nominal 90% subsampling confidence intervals. n is the dimension of the system. g_0 is the number of independent stochastic trends of the DGP. g is the hypothetical numbers of independent stochastic trends under test. T is the total sample size. b is the used subsample size.* $\rho_1 = \rho_2 = \rho_3 = \rho_4 = 0, \rho_{e1} = \rho_{e2} = \rho_{e3} = \rho_{e4} = 0$.

Table 1 shows that the subsampling procedure provides satisfactory results even at moderate number of observations. These simulation results show that the subsampling is a very promising method to handel the test of cointegration rank in large scale models[14]. In most cases the test will have correct size[15]. Obviously, the power of the test is very high, i.e. in the simulation the testing procedure did not falsely accept any stochastic trend. This is due to the sharp contrast between the asymptotical diverging property

[14]The standard Johansen procedure will fail to apply due to its data demanding formulation, i.a. in case of a system with 50 variables, an error correction model with 4 lags will use up all the degree of freedoms for 200 observations.

[15]The size downwards distortion becomes significant when the dimension of the system become large. According to Politis et al. (1999) Chapter 9, calibration method can be used to adjust the nominal coverage frequency to get a correct size. We show the subsampling coverage frequency to document this nominal downwards distortion in case of limited sample size.

of the h largest eigenvalues and the converging property of the g smallest eigenvalues. To illustrate this point we plot(See Figure 1) the whole sample test statistic and the confidence bounds based on the empirical quantiles of subsample test statistics together in one graph. For the converging eigenvalues the whole sample test statistic that is presented by the dot-and-dash constant line in the graph will lie with probability $1-\alpha$ within confidence bounds that are presented by a solid line for the upper bound and a dotted line for the lower bound respectively. Figure 1 shows the simulation case with following parameters: $T = 200, n = 16, g = 5, b \in (50, 100)$, $\rho_1 = \rho_2 = \rho_3 = \rho_4 = 0, \rho_{e1} = \rho_{e2} = \rho_{e3} = \rho_{e4} = 0$.

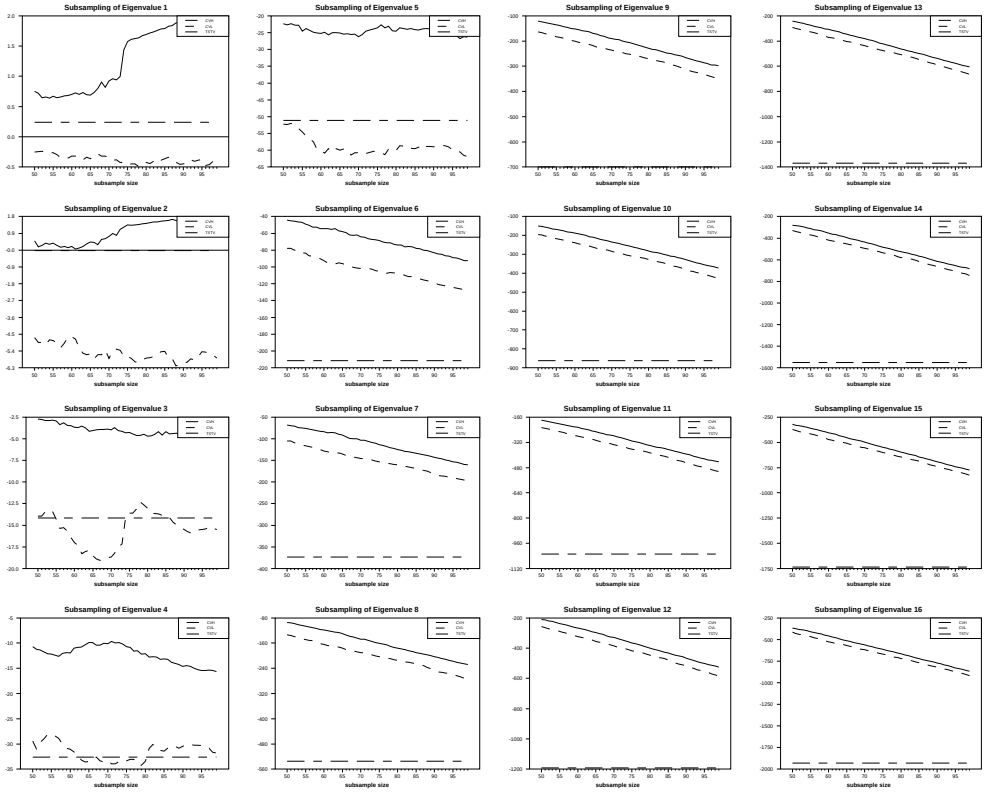

Figure 1: Subsampling confidence interval without autocorrelation
$N = 16, g = 5, T = 200, b \in (50, 100)$

The 4 graphs in the first column and first graph in the second column show the subsampling confidence intervals for the 5 smallest eigenvalues that correspond to 5 independent stochastic trends. The horizontal axis indicates the subsample sizes from 50 to 100, the vertical axis indicates the respective upper and lower bounds of the confidence intervals as well as the test statistics

over the whole sample. These graphs show that for the 5 smallest eigenvalues the whole sample test statistics and the confidence intervals are of the same order and in most cases the test statistics lie within the corresponding confidence intervals, while for the other 11 larger eigenvalues the whole sample test statistics and the subsample confidence interval bounds are not of the same order, and the whole test statistics are about T/b times larger (in absolute value) than the corresponding bounds (See Theorem 3.5). These graphs provide a clear picture about the consistence of the subsampling test.

In case of the presence of moderate autocorrelation[16] in the stationary components (u_t is autocorrelated.) this testing procedure works still quite well. The next panel of graphs are simulation results with following parameters: $T = 200, n = 16, g = 5, b \in (50, 100), \rho_1 = 0.4, \rho_2 = 0.3, \rho_3 = -0.2, \rho_4 = 0.1$

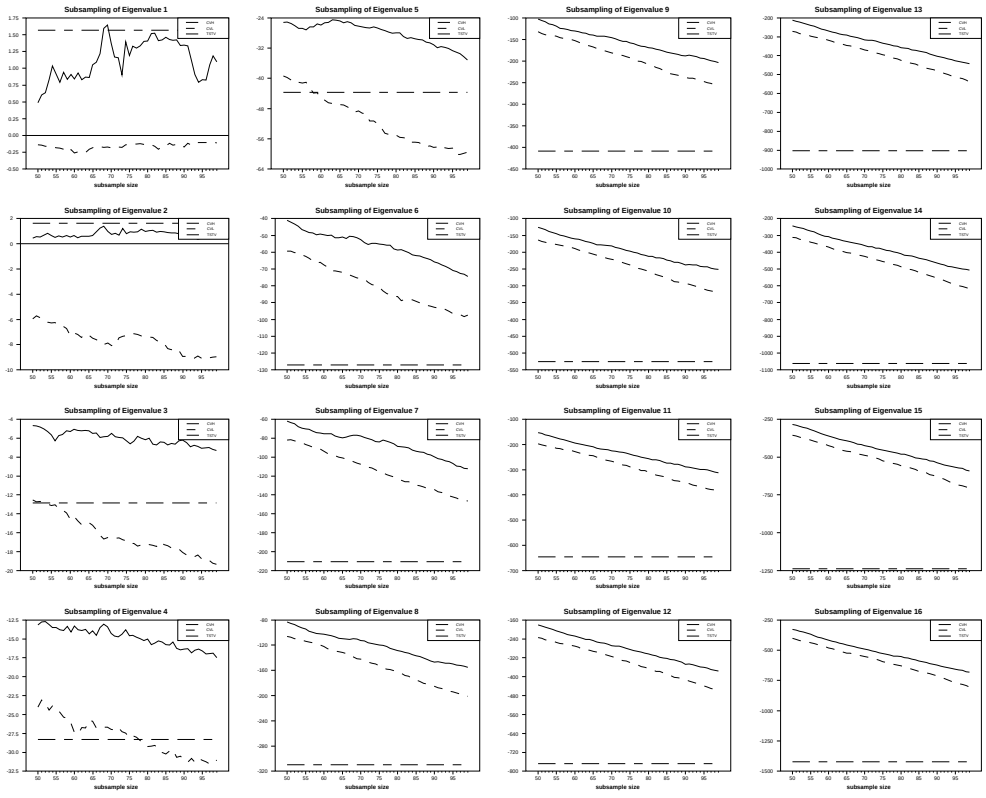

Figure 2: Subsampling Confidence Interval with Autocorrelation

[16]It is to note that if the autocorrelation in u_t are getting close to units, subsampling will not have good results with finite number of observations. This is due to the same reason why Dick-fuller test can not differ the unit root process from a AR process with a autocorrelationcoefficient close to unit if the number of observations are limited.

5.2 Centered Subsampling

The above subsampling test performs well, when the number of independent stochastic trends are small. However, it has generally a tendency to under estimate the number of stochastic trends when the number of stochastic trends is large. Following graph shows a typical result of the subsampling performance in the case when the number of independent stochastic trends are close to the dimension of the system.

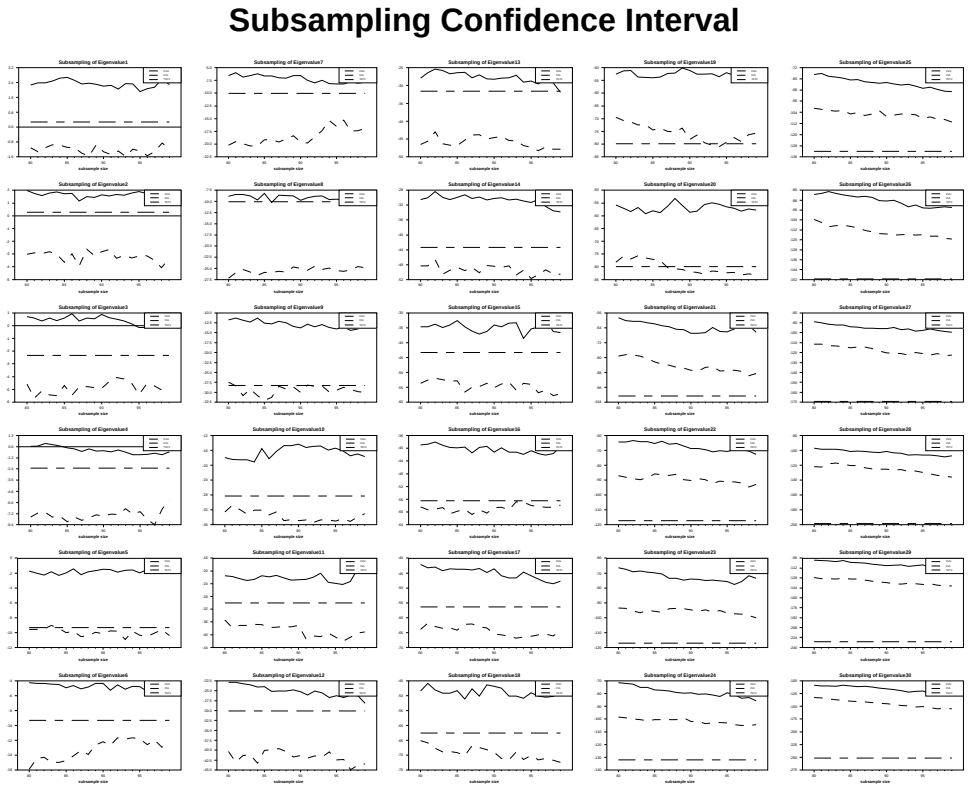

Figure 3: Uncentered Subsampling in the case of large number of stochastic trends. $N = 30$*,* $g = 28$*,* $T = 200$ *and* $b \in (80, 100)$

We have here 28 independent stochastic trends. But only 20 whole sample test statistics lie within the corresponding confidence bounds. This means that the subsampling would predict only 20 independent stochastic trends. The reason for the tendency of under estimation of the number of the independent stochastic trend is that the corresponding eigenvalues of $T\hat{\Pi}$ will converge to the asymptotical distribution one-sidedly. Consequently, the center of the approximated sample distribution of $\hat{\lambda}_{b_T,i}$ and $\hat{\lambda}_{T,i}$ are not the same as that of the asymptotical distribution of λ_i; and $\hat{\lambda}_{b_T,i}$ has larger

bias than $\hat{\lambda}_{T,i}$. To demonstrate the difference in bias we plot (see Figure 4) the (average) of T times the eigenvalues of $\hat{\Pi}_T$ that is the OLS estimate of $\Delta Y_t = \Pi Y_{t-1} + e_t$ and b_T times the eigenvalues of $\hat{\Pi}_{b_T}$ against the order of the eigenvalues. $T = 200, b_T = 70$ are numbers of observations that are used to calculate $\hat{\Pi}_T$ and $\hat{\Pi}_{b_T}$. Y_t is generated according to (2.1) with $n = 30$ and $g = 30$.

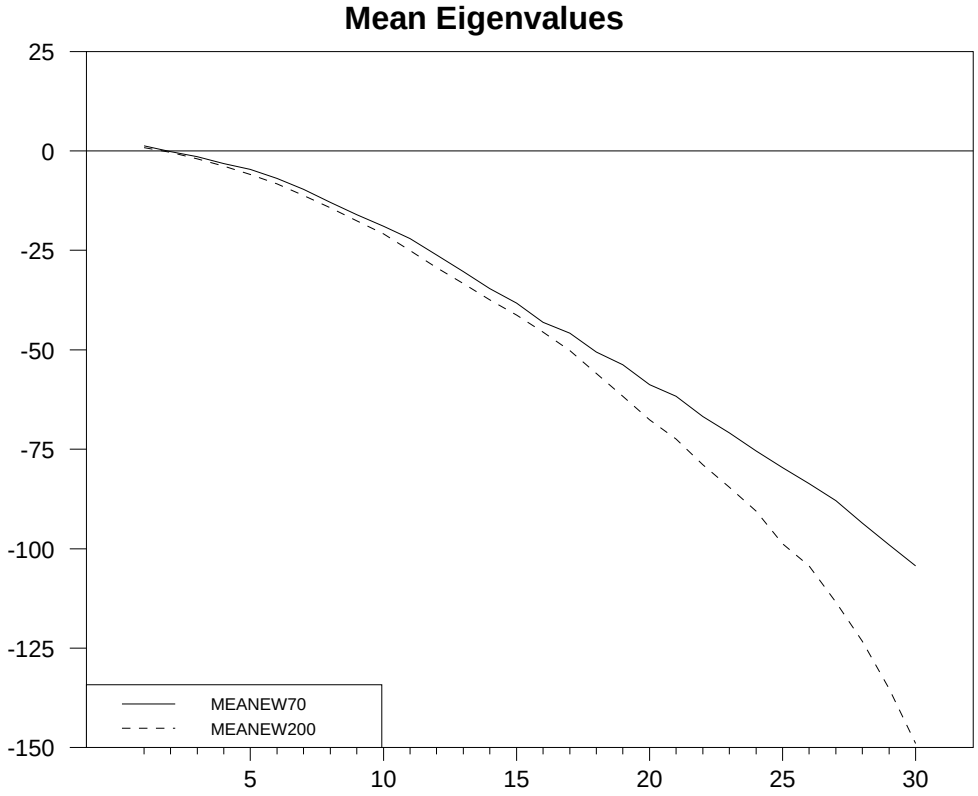

Figure 4: Mean eigenvalues of $\hat{\Pi}_T$ over 200 runs

Figure 4 shows these biases. Asymptotically these two curves should lie over each other on the curve $E(\lambda_i), i = 1, 2, ..n$ because Theorem 3.5 says $T\Lambda(\hat{\Pi}_T)$ and $b_T\Lambda(\hat{\Pi}_{b_T})$ should have same distribution and thus their expected values are the same. However the curve of $T\Lambda(\hat{\Pi}_T)$ lies below the curve of $b_T\Lambda(\hat{\Pi}_{b_T})$. The discrepancy between the two curves is directly related two the underestimation of the umber of stochastic trends. In subsampling tests we approximate the distribution of $T\hat{\Pi}_T$ by that of $b_T\Lambda(\hat{\Pi}_{b_T})$. Because $b_T\Lambda(\hat{\Pi}_{b_T})$ is biased one-sidedly upwards biased for $T\hat{\Pi}_T$, subsampling underestimates the number of independent stochastic trends. We observe also the the first few (absolutely smaller) eigenvalues calculated with difference number of observations are more close together than the latter (absolutely larger) eigenvalues.

In the context of subsampling test of cointegration rank in a system, if the number of stochastic trends in the system is small, only the first few (absolutely smaller) eigenvalues are relevant for test. Because in these cases the biases are small the subsampling can get good results. As the number of stochastic trends increases, more eigenvalues are relevant for test. The big biases have here significant influence on the performance of subsampling test. If we use $\hat{\lambda}_{b_T,i}$ as an approximation of $\hat{\lambda}_{T,i}$ as we did in the case of subsampling, $\hat{\lambda}_{T,i}$ tends to lie below the the confidence interval calculated base on $b_T\hat{\lambda}_{b_T,i}$. This is what we observe the Figure 3 and Table 2.

One way to solve this problem is to adjust the test statistics with respect to these biases. Instead of (3.12) and (3.13) we look at the centered distribution:

$$\lim_{T\to\infty}(\lambda_{i,T} - \overline{\lambda}_{i,T}), \tag{5.25}$$

where $\bar{\lambda}_{i,T}$ is the center of distribution of $\lambda_{i,T}$. It is calculated as the average of $\lambda_{i,T}$. For the centered subsampling version we have following result:

Theorem 5.1 *Let $\lambda_{i,T}$, λ_i be defined as in Theorem 3.5. If we replicate the estimation of $\hat{\Pi}_T$ by drawing from the same DGP as defined in (2.1) at the same starting point, then $\lambda_{i,T}$ can be seen as a function of realization of one draw. Let $\lambda_{i,T,j}$ be the j-th draw. $\lambda_{i,T,j}$ is independent and identical to $\lambda_{i,T}$. We define $\overline{\lambda}_{i,T} = \frac{1}{m}\sum_{j=1}^{m}\lambda_{i,T,j}$.*

For $i = 1, 2, ...h$

$$\lim_{T\to\infty}\frac{\lambda_{i,T} - \overline{\lambda}_{i,T}}{\sqrt{T}} \overset{D}{\Longrightarrow} \tilde{\lambda}_i(distribution), \tag{5.26}$$

where $\tilde{\lambda}_i$ is a random variable that has the limiting distribution of (5.26), or equivalently

$$\lim_{T\to\infty}\lambda_{i,T} - \overline{\lambda}_{i,T} \xrightarrow{P} \infty. \tag{5.27}$$

For $i = h+1, \cdots, n$

$$\lim_{T\to\infty,m\to\infty}(\lambda_{i,T} - \overline{\lambda}_{i,T}) \overset{D}{\Longrightarrow} \lambda_i - E(\lambda_i). \tag{5.28}$$

Proof: See Appendix

Theorem 5.1 implies that subsampling of $\lambda_{i,T} - \overline{\lambda}_{i,T}$ is a consistent test for identification of the cointegration rank, because for $i \leq h$ the whole sample test statistic $\lambda_{i,T} - \overline{\lambda}_{i,T}$ will asymptotically escape any empirical confidence interval based on the subsample test statistics $\{\lambda_{i,b_T} - \overline{\lambda}_{i,b_T}\}$.

The procedure of the centered subsampling test:

- For a chosen g $(g = 1, 2, ...n)$ simulate the (2.1) m times and get $\{Y_t^s\}, s = 1, 2, ..m$. The superindex s indicates Y_t^s are the simulated series.

- Calculate the eigenvalue of $\hat{\Pi}_T^s$ and $\hat{\Pi}_{b_T}^s$ based on the simulated series for each run, respectively. The superindex s indicates the estimation is based on the simulated data. T is the total number of observations and b_T is properly chosen subsample block size.

- Calculate the average of the $(n+1-g)$-th elements of the eigenvalues (ordered decreasingly according to the absolute values) over m runs, and take this as the $(n+1-g)$-th element of correction vector $\overline{\Lambda}_T$ and $\overline{\Lambda}_g$ respectively.

- Calculate $\Lambda(\hat{\Pi}_T)$ based on the observed data and take $T(\Lambda(\hat{\Pi}_T) - \overline{\Lambda}_T)$ as the centered whole sample test statistic for each eigenvalues, where $\Lambda(\hat{\Pi}_T)$ is the vector the eigenvalues of $\hat{\Pi}_T$.

- Calculate $T - b + 1$ subsample test statistics $b_T(\Lambda(\hat{\Pi}_{b_T,t} - \overline{\Lambda}_{b_T})$ $(t = b, b+1, ..., T)$ and calculate calculate the empirical confidence region at a given confidence level α based on the subsample statistics.

- If the centered whole the sample test statistic $T(\Lambda\hat{\Pi}_T - \overline{\Lambda}_T)$ lies within the corresponding confidence region based on subsample statistics, conclude that this eigenvalue belongs to the space of stochastic trends. Find the absolutely largest eigenvalues that lies within the corresponding confidence region. The index of this eigenvalue is the number of independent stochastic trend of the system.

Following graph is one typical result of centered subsampling test based on the Theorem above. A simulation assessment of centered subsampling is documented in Table 2.

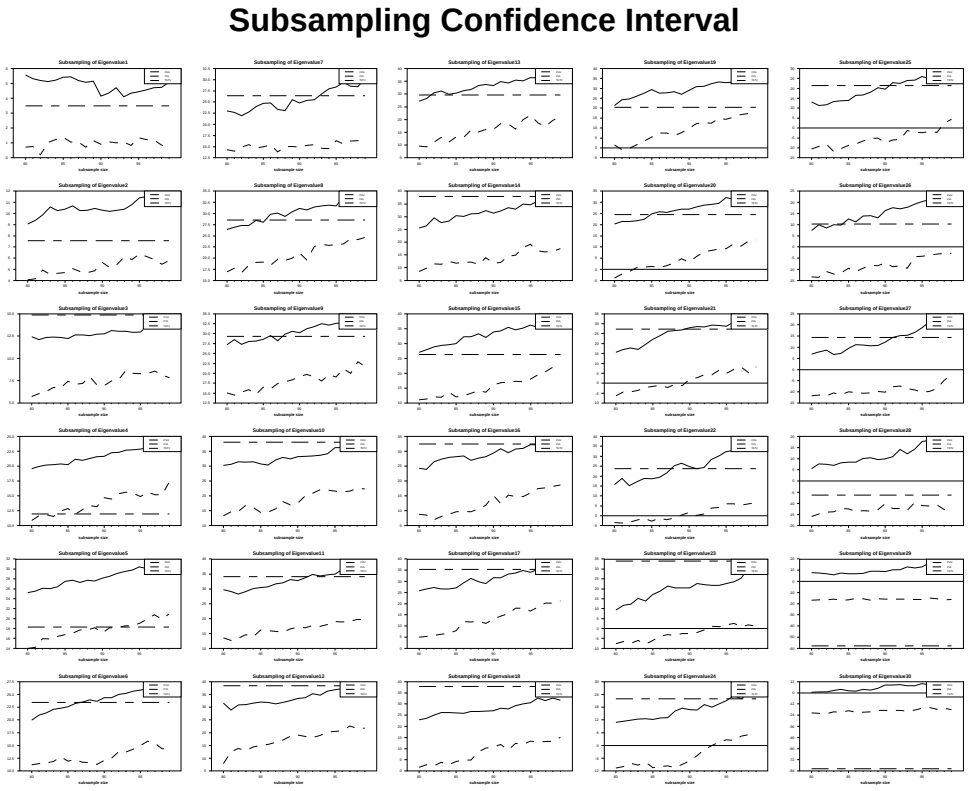

Figure 5: Centered Subsampling with large number of stochastic trends
$N = 30, g = 28, T = 200, b \in (80, 100)$

It can be clearly seen that after centering, the whole sample test statistic of first 28 eigenvalues lie within the corresponding confidence regions based on subsample statistics, while for the last two eigenvalues, the whole sample statistics lie far below the corresponding confidence regions.

In Table 2 we report some simulation results of tests based on the uncentered subsampling procedure, the centered subsampling and the Johansen procedure. The reason for taking the Johansen procedure as the benchmark for the comparison with the subsampling procedure is that the Johansen procedure seems to be the most popular one in applied cointegration research[17]. Further, the H_0 under test are the same - the number of independent stochastic trends.

[17]There are some other methods: Stock and Watson (1988), Phillips (1991), Bierens (1997), and Bierens (1997) which can be used to test for the number of cointegration relations. See Haug (1996) for a study of the comparison of some of these methods.

N	g_0	T	b	JH $g = g_0$	JH $g = g_0 + 1$	UCSS $g = g_0$	UCSS $g = g_0 + 1$	CSS $g = g_0$	CSS $g = g_0 + 1$
20	1	200	90	0.50	0.00	0.90	0.00	0.96	0.00
20	3	200	90	0.96	0.00	0.93	0.00	0.93	0.00
20	5	200	90	0.90	0.06	0.90	0.00	0.90	0.00
20	7	200	90	0.93	0.13	0.96	0.00	0.96	0.00
20	9	200	90	0.73	0.26	0.90	0.00	0.97	0.00
20	11	200	90	0.80	0.13	0.80	0.00	1.00	0.00
20	13	200	90	-	-	0.83	0.00	1.00	0.00
20	15	200	90	-	-	0.76	0.00	1.00	0.00
20	17	200	90	-	-	0.80	0.00	1.00	0.00
20	19	200	90	-	-	0.70	0.00	0.96	0.00
30	1	200	70	0.50	0.00	0.90	0.00	0.90	0.00
30	4	200	70	0.86	0.20	0.88	0.00	1.00	0.00
30	7	200	70	0.90	0.70	0.90	0.00	0.93	0.00
30	10	200	70	0.90	0.66	0.76	0.00	0.83	0.00
30	13	200	70	-	-	0.43	0.00	0.90	0.00
30	16	200	70	-	-	0.35	0.00	1.00	0.00
30	19	200	70	-	-	0.16	0.00	0.85	0.00
30	22	200	70	-	-	0.07	0.00	0.90	0.00
30	25	200	70	-	-	0.00	0.00	0.97	0.00
30	28	200	70	-	-	0.00	0.00	0.85	0.00
40	1	200	70	0.60	0.00	0.83	0.00	0.90	0.00
40	4	200	70	0.96	0.53	1.00	0.00	1.00	0.00
40	7	200	70	1.00	0.90	0.76	0.00	0.93	0.00
40	10	200	70	0.96	0.83	0.43	0.00	0.93	0.00
40	13	200	70	-	-	0.16	0.00	0.90	0.00
40	16	200	70	-	-	0.10	0.00	0.83	0.00
40	19	200	70	-	-	0.03	0.00	0.80	0.00
40	22	200	70	-	-	0.00	0.00	0.80	0.00
40	25	200	70	-	-	0.00	0.00	0.79	0.00
40	28	200	70	-	-	0.00	0.00	0.67	0.00
40	31	200	70	-	-	0.00	0.00	0.68	0.00
40	34	200	70	-	-	0.00	0.00	0.73	0.00
40	37	200	70	-	-	0.00	0.00	0.67	0.00

Table 2: *Relative frequency of acceptance of H_0 at the nominal level of 90%. N is the dimension of the system. g_0 is the number of the independent stochastic trends. T is the total sample size. b is the subsample size. H_0 under test are $g = g_0$ and $g = g_0 + 1$ respectively. $\rho_1 = \rho_2 = \rho_3 = \rho_4 = 0, \rho_{e1} = \rho_{e2} = \rho_{e3} = \rho_{e4} = 0$. u_t and u_{et} are i.i.d draws from standard normal distribution. The matrix C are filled with random numbers drawn from normal distribution with variance 4. The lag in the VAR of the ECM of Johansen procedure is 2. JH indicates the Johansen procedure. USS indicates the uncentered subsampling method. CSS indicates the centered subsampling.*

Obviously, the power of Johansen procedure in rejection of false stochastic tends is low. In the simulated cases it has positive probability to overestimate the number of stochastic trends. This probability becomes larger, when the number of the independent stochastic trends in the system gets larger. Because we have only critical values for Johansen trace statistic up to $g = 12$, the comparison with the Johansen procedure stops at $g = 13$. The uncentered subsampling performs well when g are small. When g is larger than 10 the uncentered subsampling is severely undercovered. The centered subsampling performs very well. Even for the cases when the number of the independent stochastic trend is close to the dimension of the system centered subsampling has still correct size and higher power.

In the last section we have discussed two asymptotically equivalent criteria, one is based in the eigenvalues of $T\hat{\Pi}_T$, the other is based on the T times the canonical correlation coefficient between Δy_t and y_{t-1}. In the following table we records the simulation results of these two tests using the same sets of data.

N	g_0	T	b	CSS RE	CSS CC
20	1	200	90	0.96	0.90
20	3	200	90	2.96	2.90
20	5	200	90	4.90	4.63
20	7	200	90	6.96	6.70.
20	9	200	90	8.94	8.83
20	11	200	90	11	10.57
20	13	200	90	13	12.40
20	15	200	90	15	14.12
20	17	200	90	17	16.57
20	19	200	90	18.96	18.60
30	1	200	90	0.95	0.89
30	4	200	90	4.00	3.76
30	7	200	90	7.00	6.74
30	10	200	90	9.98	9.66
30	13	200	90	12.96	12.6
30	16	200	90	15.95	15.56
30	19	200	90	18.90	18.68
30	22	200	90	21.85	21.64
30	25	200	90	24.84	24.41
30	28	200	90	27.93	27.74
40	1	200	70	0.9	0.70
40	4	200	70	4	3.30
40	7	200	70	6.97	6.92
40	10	200	70	9.95	9.39
40	13	200	70	12.90	12.60
40	16	200	70	15.83	15.65
40	19	200	70	18.80	18.87
40	22	200	70	21.73	21.53
40	25	200	70	24.57	24.45
40	28	200	70	27.74	27.54
40	31	200	70	30.75	30.25
40	34	200	70	33.56	33.48
40	37	200	70	36.74	36.45

Table 3: *The estimated number of independent stochastic trends averaged over 100 simulations based on the subsampling tests. N is the dimension of the system.* g_0 *is the number of the independent stochastic trends of the DGP. T is the total sample size. b is the subsample size.* $\rho_1 = \rho_2 = \rho_3 = \rho_4 = 0, \rho_{e1} = \rho_{e2} = \rho_{e3} = \rho_{e4} = 0$. u_t *and* u_{et} *are i.i.d draws from standard normal distribution. The matrix C are filled with random numbers drawn from normal distribution with variance 4. CSS RE indicates the centered subsampling based on* $T\hat{\Pi}$*, CSS CC indicates the centered subsampling based on the canonical correlation coefficients.*

Obviously, both two tests give good results for the simulated benchmark cases. Within finite sample size, the test based on the eigenvalue of $T\hat{\Pi}$ performs slightly better than that based on the canonical correlation coefficients.

For empirical application of the subsampling procedure, one critical issue is the autocorrelation in the disturbance u_t. Although the asymptotically theory on the subsampling tests will hold for autocorrelated disturbances, the finite sample property of the test procedure may change if the autocorrelation in the disturbance term is very high. It is well known that a unit root process and a highly positively autocorrelated stationary process are observationally indistinguishable with finite data. Therefore, subsampling will not be able to identify the number of independent stochastic trends if the autocorrelation in the disturbances are too high. To see how subsampling performs when the autocorrelation in the disturbance are moderate, we have run following simulations.

N	g_0	T	b	CSS RE	CSS CC
30	1	200	90	0.95	0.86
30	4	200	90	3.89	3.59
30	7	200	90	6.91	6.54
30	10	200	90	9.91	9.42
30	13	200	90	12.92	12.41
30	16	200	90	15.90	15.46
30	19	200	90	18.83	18.56
30	22	200	90	21.79	21.43
30	25	200	90	24.81	24.40
30	28	200	90	27.83	27.51

Table 4: *The estimated number of independent stochastic trends averaged over 100 simulations based on the subsampling tests. N is the dimension of the system. g_0 is the number of the independent stochastic trends of the DGP. T is the total sample size. b is the subsample size. $\rho_1 = \rho_2 = \rho_3 = \rho_4 = 0, \rho_{e1} = 0.4, \rho_{e2} = -0.2, \rho_{e3} = 0.2, \rho_{e4} = 0.1$. u_t and u_{et} are i.i.d draws from standard normal distribution. The matrix C are filled with random numbers drawn from normal distribution with variance 4. CSS Re indicates the centered subsampling based on the eigenvalues of $T\hat{\Pi}$, CSS CC indicates the centered subsampling based on the canonical correlation coefficient.*

The simulation results show that when the autocorrelation are moderate, subsampling test procedure will still perform satisfactory. Also here, the test based on the eigenvalues of $T\hat{\Pi}$ performs slightly better than the test based on the canonical correlation coefficients.

6 An Application Example

After we investigate the performance of subsampling testing procedure, we turn to an example of application of this testing procedure, namely the test of purchasing power parity(PPP).

Purchasing power parity is a benchmark model for nominal exchange rate determination. Over last decades a vast amount of empirical works have been conducted to assess the evidence of PPP in the real world. While most literature in 80s found evidence against PPP: the real exchange rate were found to follow a random walk, see for example Frankel (1981), and Enders (1988); the works in 90s deliver some results favor PPP in a long run cointegration relation, see for example Johansen and Juselius (1992), Kim (1990) and Abuaf and Jorion (1990). Resent panel studies do not provide a clear picture of PPP. For example Choi (2001), Oh (1996) found evidence favor PPP, while Fleissig and Strauss (2000) did not find evidence favor PPP.

The economical reasoning behind PPP is the goods arbitrage, i.e. the domestic price for products should equal to the foreign price for the same foreign products converted by the exchange rate:

$$P_d = S_{df} P_f, \tag{6.29}$$

where P_d is the domestic price, P_f is the foreign price, and S_{df} is the price of foreign currency - the exchange rate.

Taking the logarithm we get :

$$p_d = s_{df} + p_f, \tag{6.30}$$

where the small case letters denote the corresponding logs. Because p_d, p_f and s_{df} are I(1) variables the exact PPP relation (6.30) is then tested in the form that the real exchange rate s^r_{df} is a stationary I(0) variable:

$$s^r_{df} := s_{df} - p_d + p_f. \tag{6.31}$$

This implies at same time that the variable p_d, p_f and s_{df} are cointegrated with a particular cointegration vector $(1, -1, 1)'$. A large number of empirical studies were conducted in 90s to explore this specific cointegration relation among nominal exchange rates and price indices between different countries. Also in the panel studies, this cointegration relation is the object under test. Bilateral PPP implies many cointegration relations in a multi-country system.

In the following we are going to characterize PPP for a system of N countries. In the case of three countries, we may haver pair wise PPP:

$$-p_0 + s_{01} + p_1 \sim I(0), \tag{6.32}$$

$$-p_0 + s_{02} + p_2 \sim I(0), \tag{6.33}$$

and

$$-p_1 + s_{12} + p_2 \sim I(0). \tag{6.34}$$

In fact (6.34) is the difference between (6.33) and (6.32). Therefore PPP implies two linear independent particular cointegration relations among three countries. This simple example can be generalized to the case of N countries. In fact bilateral PPP relations with respect to one country implies bilateral PPP relations among other countries. PPP implies $N-1$ independent cointegration relations among a system of N countries with $2N-1$ variables. Choose arbitrarily a country as the base country, we may order the relevant variables of the system in the following way:

$$y_t = \begin{pmatrix} p_0 \\ s_{01} \\ p_1 \\ \vdots \\ s_{0(N-1)} \\ p_{N-1} \end{pmatrix}. \tag{6.35}$$

Then PPP implies that this $2N-1$ system is a cointegrated system with $N-1$ cointegration relations like follows:

$$\begin{pmatrix} -1 & 1 & 1 & 0 & 0 & & & \dots & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 & & & \vdots \\ \vdots & \vdots & 0 & 0 & 0 & \ddots & \ddots & & 0 \\ -1 & 0 & \dots & & & & 0 & 1 & 1 \end{pmatrix}. \tag{6.36}$$

Taking advantage of these particular cointegration relations, many empirical works were conducted in panel setting to increase the power of test.

In this paper we conduct a direct test of these $N-1$ cointegration relations in the $2N-1$ dimensional system. Here we investigate the relation among exchange rates and price levels among the most important 10 currencies in the world economy: US Dollars, Canadian Dollars, Euro, Japan Yen, Swiss Francs, British Pounds, Korea Wan, Australian Dollars, New Zealand Dollars, Mexican Pesos[18]. We look at a system of 9 exchange rates and 10 price indices. The data are daily from 2000:01:01 to 2002:10:02. The main summary statistics of the data are in the following table.

[18]There are still two important currencies: the Hongkong Dollars and the Chinese Yuan. Because they were institutionally under pegging system, we do not consider these two currencies.

Series	Obs	Mean	Std Error	Minimum	Maximum
CAD_USD	717	1.50130098	0.07950400	1.29240000	1.61320000
DEUR_USD	717	1.03754449	0.10297501	0.79386000	1.20871000
JPY_USD	717	117.61344491	7.68910553	102.65000000	134.79000000
GBP_USD	717	0.65868662	0.03765339	0.56031000	0.72563000
CHF_USD	717	1.57014533	0.15506100	1.23780000	1.82140000
AUD_USD	717	1.76061144	0.17266568	1.32810000	2.06980000
NZD_USD	717	2.11555411	0.27333018	1.52620000	2.54700000
KRW_USD	717	1214.26457462	68.86227937	1105.20000000	1368.30000000
MXN_USD	717	9.81171116	0.65366391	9.00230000	11.40970000
PCANADA	717	1.11234310	0.02287517	1.07000000	1.16000000
PEURO	717	1.12616457	0.02159016	1.09000000	1.16000000
PJAPAN	717	1.00903766	0.00834286	0.99000000	1.02000000
PUK	717	1.15896792	0.01697173	1.12000000	1.19000000
PSWISS	717	1.04523013	0.00857818	1.03000000	1.06000000
PAUS	717	1.13938633	0.03681317	1.07000000	1.19000000
PNZ	717	1.10020921	0.02380873	1.06000000	1.14000000
PKOREA	717	1.25447699	0.03443740	1.20000000	1.31000000
PMEXIKO	717	2.52722455	0.11249691	2.32000000	2.72000000
PUS	717	1.15408647	0.02211974	1.11000000	1.19000000

Source : Federal Reserve Bank of St. Louis
http://research.stlouisfed.org/fred2/categories/15\\
OECD: main economic indicator

All 19 series are graphed in the following graph panel. All series show strong stochastic trends. Except Japan all price indices have upward trend. Some exchange rates such as EURO/USD, BRP/USD, CHF/USD,AUD/USD and NZD/USD show well pronounced similar trend movement, which may suggest some cointegration relations among them.

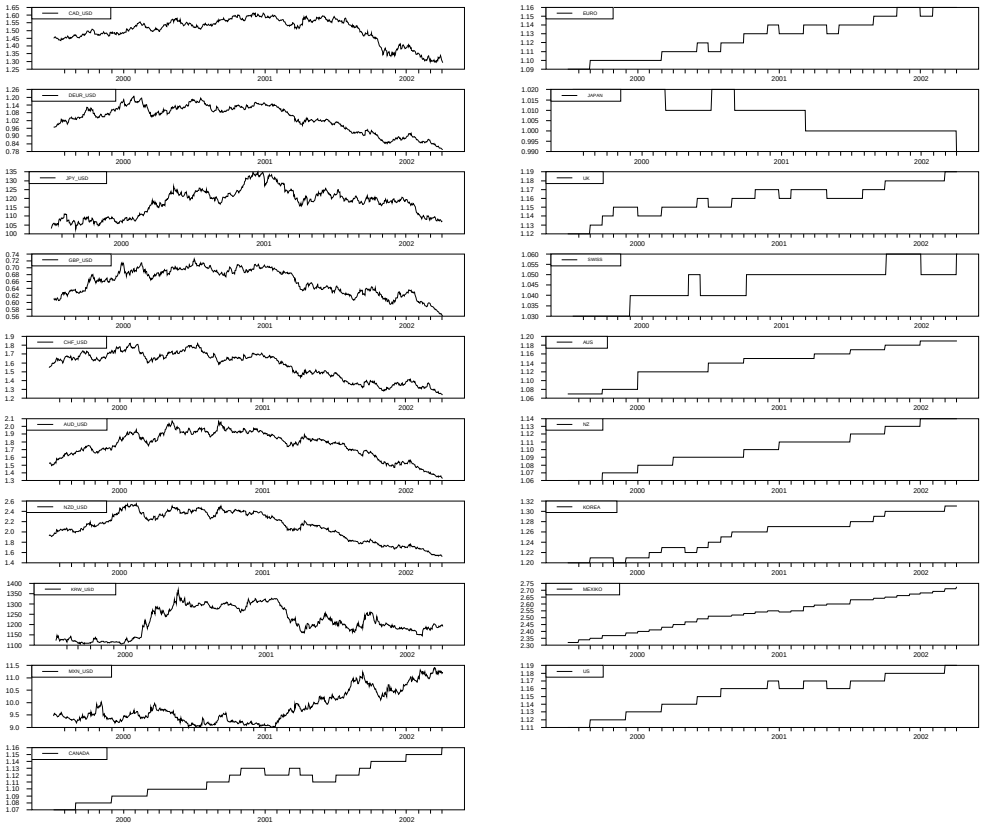

Figure 3: The fundamental data of the model.

The discussion in the last section shows that if the number of independent stochastic trends are small, we can use uncentered subsampling; if we expect large number of stochastic trends we should use centered subsampling. Hence we will start with uncentered subsampling test. If we then get any evidence that the stochastic trends are underestimated, we can run centered subsampling to get correct number of stochastic trends. If we can not find any cointegration relation based on the uncentered subsampling, this will be a more strong evidence against cointegration.

Our test starts therefore with uncentered subsampling. Contrary to the optical impression, the test shows that there is no cointegration relations among these 19 variables. In Figure 4 we can see clearly the even the smallest eigenvalue(in absolute value the largest) has the same order of the corresponding confidence bounds (see the last graph in the fourth column): it lies in fact within the subsampling confidence interval. This mean that there is no cointegration relations within 19 variables. This result implies that we can not find empirical evidence for PPP hypothesis with daily data within a short period of 3 years. This empirical result confirms the findings in the literature

that there is no evidence for PPP for short run period.

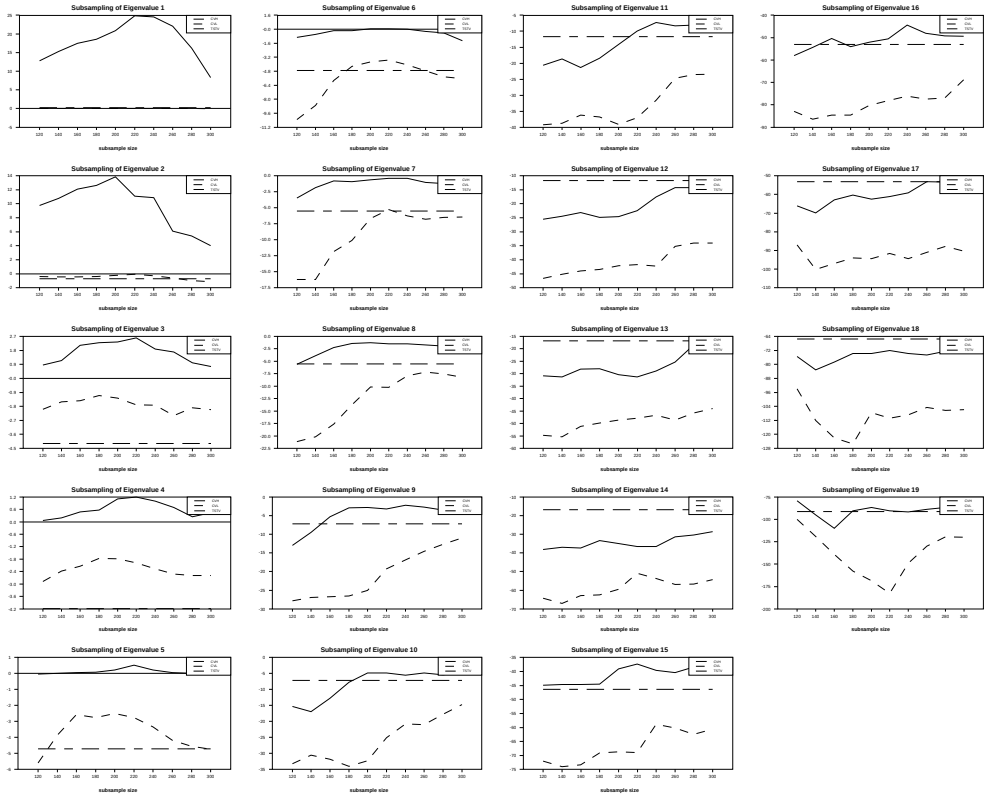

Figure 4: Subsampling confidence interval for block size: $b \in (100, 300)$.

Now we turn to the long run PPP relation. We investigate in fact the same system in the post-Bretton Woods period from 1973 to 2002. Due to data availability Korea and Mexico have to be excluded from the system. The data are monthly from 1973:1 to 2001:12. The main summary statistics of the data are in the following table.

Series	Obs	Mean	Std Error	Minimum	Maximum
CAD_USD	348	1.25128994	0.15740167	0.96230000	1.59220000
JPY_USD	348	181.40913793	68.00074320	83.69000000	305.67000000
GBP_USD	348	0.58802093	0.09610155	0.38816862	0.91482938
CHF_USD	348	1.83921638	0.52598882	1.13840000	3.72930000
DEM_USD	348	2.05346236	0.43630573	1.38120000	3.30250000
AUD_USD	348	1.21006869	0.31927103	0.67317402	1.99362041
NZD_USD	348	1.50079774	0.44567899	0.67276642	2.50626566
CAN445241K	348	0.74470489	0.26868023	0.25810000	1.12750000
GEM125241K	348	0.80161207	0.18652050	0.45510000	1.10210000
JAP465241K	348	0.84753764	0.17127051	0.37260000	1.03340000

SWZ685241K	348	0.78951437	0.18556799	0.44870000	1.05450000
USA425241K	348	0.74622989	0.26400929	0.28000000	1.17000000
GRB265241K	348	0.69524195	0.31238563	0.15190000	1.17140000
NZL595241K	348	0.65978621	0.33980043	0.12400000	1.11270000
AUS545241K	348	0.68831379	0.30557015	0.17560000	1.15970000

Source : Federal Reserve Bank of St. Louis
http://research.stlouisfed.org/fred2/categories/15\\
OECD: main economic indicator

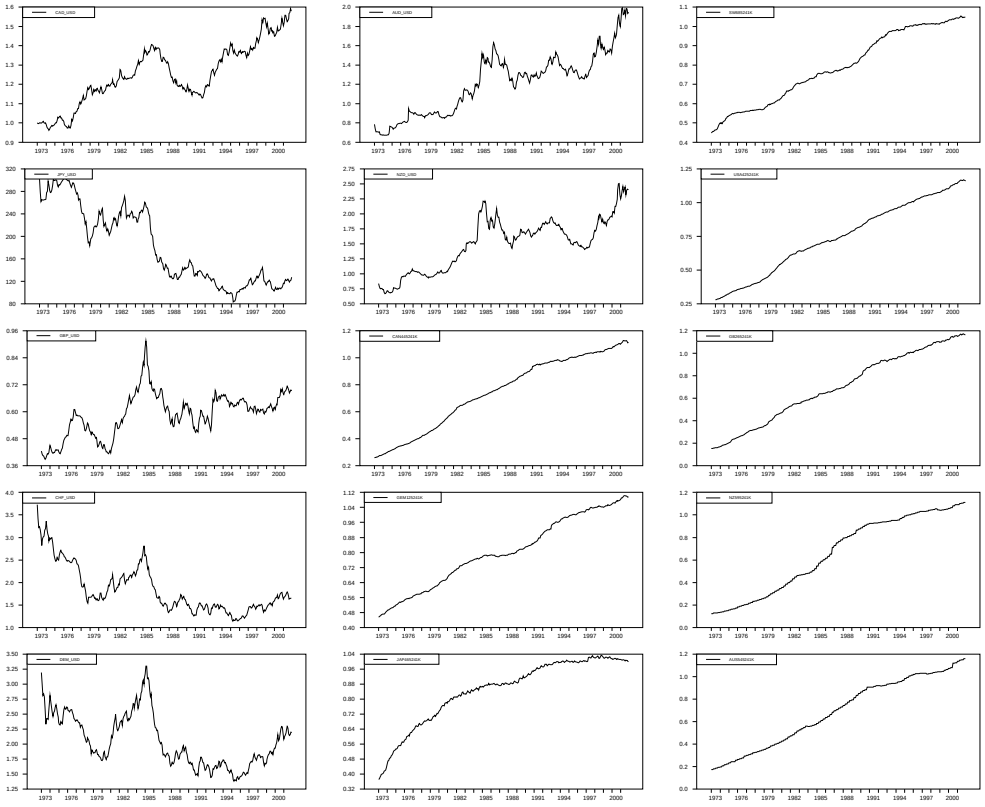

Figure 5: The fundamental data of the model.

On the graphs above there are well pronounced comovement among many series. We would expect some cointegration relation among them. Contrary to the optical impression, the test shows that there is no cointegration relations among these 15 variables. The subsampling test results are depicted in the following Figure 6.

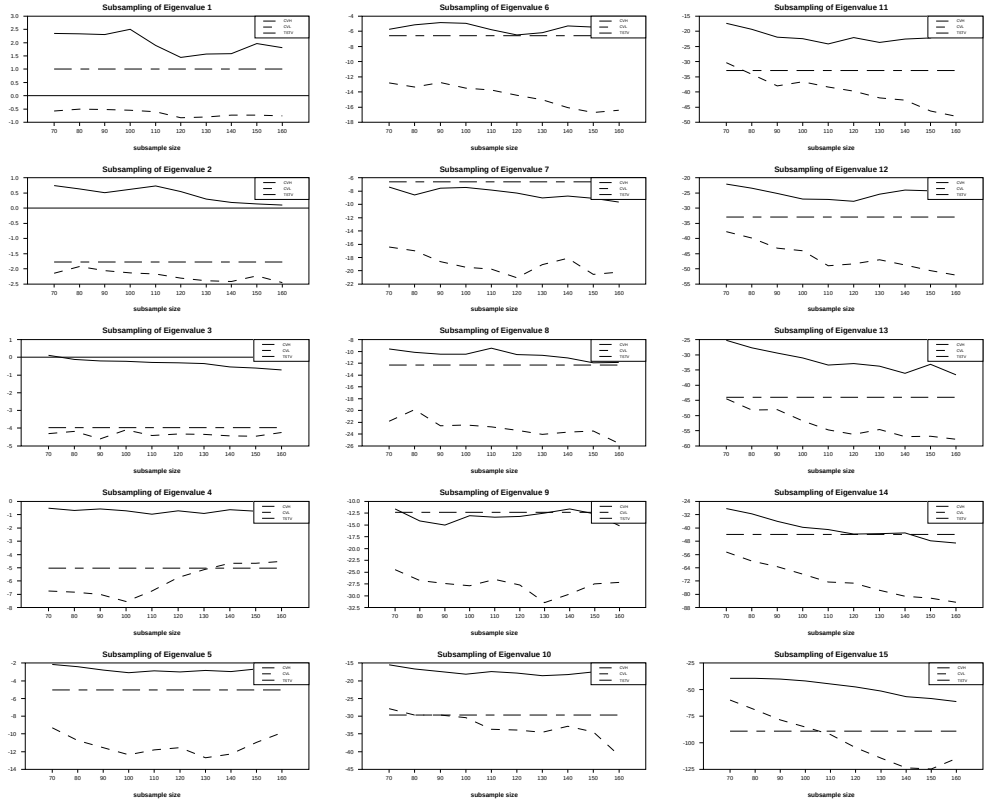

Figure 6: Subsampling confidence interval for block size: $b \in (70, 160)$.

In the graphs above we can not see any hint of diverging property of the eigenvalues. Even the smallest eigenvalue has the same order as the the subsample bounds. Therefore we conclude that the system does not contain any cointegration relations. This means that PPP hypothesis is rejected in this set of data.

7 Concluding Remarks

In this paper we develop a procedure to explore the cointegration relations among large numbers of variables and thus contribute to the econometric literature on multivariate cointegration analysis in large systems. The non-parametric subsampling test procedure is shown to be consistent and has asymptotically power 1. We show that uncenterd subsampling performs well with moderate sample size for small number of stochastic trends. We also show that the uncentered subsampling severely under estimates the number of stochastic trends, if the number is large. Centered subsampling procedure

is proposed to overcome this problem. The centered subsampling is able to identify the number independent stochastic trends even when the number of stochastic trends are close to the dimension of the system.

However, subsampling test is only a asymptotically valid test. The performance of the test in finite sample cases may depend on the parameter of the testing procedure such as the block size and the parameters of the DGP. Instead of searching for a optimal subsample size, we apply a range of reasonable subsample sizes to conduct our conclusion based on the qualitative difference between the divergence of the test statistics and the convergence of the test statistics.

Simulation studies show that the subsampling procedure perform very well for the cases of large systems with moderate requirement on data. The subsampling procedure is robust against moderate autocorrelation in the disturbance terms.

8 Appendix

Proof of Lemma 3.3.
Using the continuous mapping theorem we have

$$(\frac{G_{iT}}{T^{\alpha}})^{-1} = \frac{G_{iT}^{-1}}{T^{-\alpha}} \Rightarrow G_i^{-1}.$$

(i) is proved.

For (ii),

$$\frac{G_{iT}}{T^{\alpha_i}}\frac{G_{jT}}{T^{\alpha_j}} = \frac{G_{iT}\cdot G_{j,T}}{T^{\alpha_i+\alpha_j}} \overset{D}{\Longrightarrow} G_i \cdot G_j.$$

The proof for (iii) is similar.
□

Lemma 8.1

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A-BD^{-1}C)^{-1} & -(A-BD^{-1}C)^{-1}BD^{-1} \\ -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1} \end{pmatrix}.$$

Proof of Lemma 8.1 is done easily by matrix multiplication of the expression above and $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$.
□

Proof of Lemma 3.4
First, we calculate

$$\begin{pmatrix} G_{1,T} & G_{2,T} \\ G_{3,T} & G_{4,T} \end{pmatrix}^{-1} =: \begin{pmatrix} G_{1,T}^{*} & G_{2,T}^{*} \\ G_{3,T}^{*} & G_{4,T}^{*} \end{pmatrix}.$$

Using Lemma 8.1, the first block matrix is equal to

$$\begin{aligned}(G_{1,T}^*)^{-1} &= T\frac{G_{1,T}}{T} - T\frac{G_{2,T}}{T}T^{-2}(\frac{G_{4,T}}{T^2})^{-1}T\frac{G_{3,T}}{T} \\ &= T\underbrace{\frac{G_{1,T}}{T}}_{\xrightarrow{P} g_1} - \underbrace{\frac{G_{2,T}}{T}(\frac{G_{4,T}}{T^2})^{-1}\frac{G_{3,T}}{T}}_{\xRightarrow{D} G_2G_4^{-1}G_3}.\end{aligned}$$

The convergence of the second term in the last equation is guaranteed by the continuous mapping theorem, see Hamiltion (1994) P.482. Then,

$$\frac{G_{1,T}^*}{T^{-1}} \xrightarrow{P} \lim_{T\to\infty}\frac{(g_1T - G_2G_4^{-1}G_3)^{-1}}{T^{-1}} = g_1^{-1}. \tag{8.37}$$

Similarly, we have

$$\begin{aligned}\frac{G_{2,T}^*}{T^{-2}} &\xRightarrow{D} -g_1^{-1}G_2G_4^{-1} \\ \frac{G_{3,T}^*}{T^{-2}} &\xRightarrow{D} -G_4^{-1}G_3g_1^{-1} \\ \frac{G_{4,T}^*}{T^{-2}} &\xRightarrow{D} G_4^{-1}\,.\end{aligned}$$

Now we calculate the limit of $G_{7,T}, G_{8,T}$. Due to the continuous mapping theorem we are allowed to have the following "calculation"

$$\begin{aligned}\begin{pmatrix}G_{1,T} & G_{2,T}\\ G_{3,T} & G_{4,T}\end{pmatrix}^{-1}\begin{pmatrix}G_{5,T}\\ G_{6,T}\end{pmatrix} &\sim \begin{pmatrix}g_1^{-1}T^{-1} & -g_1^{-1}G_2G_4^{-1}T^{-2}\\ -G_4^{-1}G_3g_1^{-1}T^{-2} & G_4^{-1}T^{-2}\end{pmatrix}\begin{pmatrix}g_5T\\ G_6T\end{pmatrix} \\ &\sim \begin{pmatrix}g_1^{-1}g_5\\ (-g_5G_4^{-1}G_2g_1 + G_6G_4)T^{-1}\end{pmatrix}\end{aligned} \tag{8.38}$$

□

Lemma 8.2 *Let a, b, c, d be $h\times h, h\times g, g\times h, g\times g$ given matrices. Let $B = \begin{pmatrix}\beta_{11} & \beta_{12}\\ \beta_{21} & \beta_{22}\end{pmatrix}$ where $\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}$ are $h\times h, h\times g, g\times h, g\times g$ given matrices. Then, the eigenvalues of the matrix*

$$B\begin{pmatrix}Ta & Tb\\ c & d\end{pmatrix} \tag{8.39}$$

are separated into two groups: one has the order T and the other one has the order 1.

Note Lemma 8.2 will be applied later for $T\to\infty$.

Proof of the Lemma 8.2 The idea of the proof is that the eigenvalues of the

matrix X are invariant under the transformation $\Psi^{-1}X\Psi$.

Search Ψ_T of the form

$$\Psi_T = \begin{pmatrix} I & e_{12,T} \\ e_{21,T} & I \end{pmatrix}$$

where $e_{12,T}$ and $e_{21,T}$ are $h \times g$ and $g \times h$ matrices respectively.

From straightforward calculation, we obtain

$$\Psi_T^{-1} B \begin{pmatrix} Ta & Tb \\ c & d \end{pmatrix} \Psi_T = \begin{pmatrix} \tilde{A}_T & \tilde{B}_T \\ \tilde{C}_T & \tilde{D}_T \end{pmatrix}$$

where

$$\tilde{A}_T = (I - e_{12,T}e_{21,T})^{-1}\Big[\quad T(\beta_{11} - e_{12,T}\beta_{21})(a + be_{21,T}) \tag{8.40}$$
$$+ \ (\beta_{12} + e_{12,T}\beta_{22})(c + de_{21,T}) \quad \Big] ,$$

$$\tilde{B}_T = (I - e_{12,T}e_{21,T})^{-1}\Big[\quad T(\beta_{11} - e_{12,T}\beta_{21})(ae_{12,T} + b) \tag{8.41}$$
$$+ \ (\beta_{12} + e_{12,T}\beta_{22})(ce_{12,T} + d) \quad \Big] ,$$

$$\tilde{C}_T = (I - e_{21,T}e_{12,T})^{-1}\Big[\quad T(-e_{21,T}\beta_{11} + \beta_{21})(a + be_{21,T}) \tag{8.42}$$
$$+ \ (-e_{21,T}\beta_{12} + \beta_{22})(c + de_{21,T}) \quad \Big] ,$$

$$\tilde{D}_T = (I - e_{21,T}e_{12,T})^{-1}\Big[\quad T(-e_{21,T}\beta_{11} + \beta_{21})(ae_{12,T} + b) \tag{8.43}$$
$$+ \ (-e_{21,T}\beta_{12} + \beta_{22})(ce_{12,T} + d) \quad \Big].$$

We choose $e_{12,T}$ and $e_{21,T}$ such that $\tilde{B}_T \sim 0$ and $\tilde{D}_T \sim 0$.

From the constraint $\tilde{B}_T \sim 0$, we choose $e_{12,T}$ with the expression

$$e_{12,T} = -a^{-1}b + \frac{\Delta_{1,T}}{T} + O(T^{-2}) \tag{8.44}$$

where

$$\Delta_{1,T} = a^{-1}(\beta_{11} - \hat{e}_{12}\beta_{21})^{-1}(\beta_{12} + \hat{e}_{12}\beta_{22})(c\hat{e}_{12} + d)$$

with $\hat{e}_{12} = a^{-1}b$. The expression (8.44) provides an approximation

$$ae_{12,T} + b = O(T^{-1}) \ . \tag{8.45}$$

Similarly, under the setting $\tilde{C}_T \sim 0$, we choose $e_{21,T}$ with the expression

$$e_{21,T} = \beta_{21}\beta_{11}^{-1} + \frac{\Delta_{2,T}}{T} + O(T^{-2}) \tag{8.46}$$

where

$$\Delta_{2,T} = (-\hat{e}_{21}\beta_{12} + \beta_{22})(c + d\hat{e}_{21})(a + b\hat{e}_{21})^{-1}\beta_{11}^{-1}$$

with $\hat{e}_{21} = \beta_{21}\beta_{11}^{-1}$. This expression provides an approximation

$$-e_{21,T}\beta_{11} + \beta_{21} = O(T^{-1}) \ . \tag{8.47}$$

With the chosen $e_{12,T}$ and $e_{21,T}$, the leading term in (8.40) is the first term. Thus

$$\lim_{T\to\infty} \frac{\tilde{A}_T}{T} = (\beta_{11} - a^{-1}b\beta_{21})(a + b\beta_{21}\beta_{11}) =: A \ , \tag{8.48}$$

so $\tilde{A}_T$ has order T. Let $\lambda_{1,T}, \cdots, \lambda_{h,T}$ be the eigenvalues of $\tilde{A}_T$ and $\lambda_1, \cdots, \lambda_h$ be the eigenvalues of A in descending order in absolute value. Then due to the continuity of the eigenvalues

$$\lim_{T\to\infty} \frac{\lambda_{i,T}}{T} = \lambda_i, \quad \text{for } i = 1, \cdots, h.$$

Now, we calculate the eigenvalues of $\tilde{D}_T$. The most tricky part of this prove is that, by applying the approximations (8.45) and (8.47), the first term on the RHS in Eq. (8.43) hat order $\frac{1}{T}$ instead of T, therefore the leading term of $\tilde{D}_T$ in Eq. (8.43) is the second term. Then,

$$\lim_{T\to\infty} \tilde{D}_T = (-\beta_{21}\beta_{11}^{-1}\beta_{12} + \beta_{22})(ca^{-1}b + d) =: D. \tag{8.49}$$

The eigenvalues of $\tilde{D}_T$ converges to the eigenvalues of D.
□

Proof of Theorem 3.5
Theorem 3.5 is the stochastic version of Lemma 8.2. In order to use Lemma 8.2, we correspond $\begin{pmatrix} Ta & Tb \\ c & d \end{pmatrix}$ Eq. in (8.39) to $\begin{pmatrix} TG_{7,T} \\ TG_{8,T} \end{pmatrix}$ in Eq. (3.10). Since the correspondence of a, b, c, d are random variables and dependent on T, we add the subindex T to a, b, c, d and define

$$\begin{aligned} a_T &= G_{7,T,1} \quad b_T = G_{7,T,2} \\ c_T &= TG_{8,T,1} \quad d_T = TG_{8,T,2} \ . \end{aligned}$$

From Lemma 3.4 we have the convergence

$$\begin{aligned} \begin{pmatrix} a_T & b_T \end{pmatrix} &\xrightarrow{P} g_7 \ &=: \ \begin{pmatrix} g_{7,1} & g_{7,2} \end{pmatrix} \\ \begin{pmatrix} c_T & d_T \end{pmatrix} &\xRightarrow{D} G_8 \ &=: \ \begin{pmatrix} G_{8,1} & G_{8,2} \end{pmatrix} . \end{aligned}$$

These convergence has a strong version of convergence almost surely, see Föllmer (1981), that means the convergence

$$\begin{array}{rcl} \begin{pmatrix} a_T & b_T \end{pmatrix} \longrightarrow g_7 & =: & \begin{pmatrix} g_{7,1} & g_{7,2} \end{pmatrix} \\ \begin{pmatrix} c_T & d_T \end{pmatrix} \longrightarrow G_8 & =: & \begin{pmatrix} G_{8,1} & G_{8,2} \end{pmatrix} \end{array}$$

holds for almost all paths.

Applying Lemma 8.2 for all the path satisfying the convergence above, and define

$$\begin{aligned} A := \lim_{T\to\infty} \frac{\tilde{A}_T}{T} &= (\beta_{11} - g_{7,1}^{-1} g_{7,2}\beta_{21})(g_{7,1} + g_{7,2}\beta_{21}\beta_{11}) && (8.50) \\ D := \lim_{T\to\infty} \tilde{D}_T &= (-\beta_{21}\beta_{11}^{-1}\beta_{12} + \beta_{22})(G_{8,1} g_{7,1}^{-1} g_{7,2} + G_{8,2}). && (8.51) \end{aligned}$$

Then, A and D are well-defined for almost all paths (with probability one). Then, the eigenvalues of $\frac{\tilde{A}_T}{T}$ and $\tilde{D}_T$ converge to the eigenvalues of A and D with probability one. The statements (3.11) and (3.13) in Theorem 3.5 are proved.
□

Proof of Theorem 5.1

To proof the first part of the Theorem we look at Lemma 3.4 more closely. Instead of (8.37) we use the more detailed expression:

$$\begin{aligned} \frac{G^*_{1,T}}{T^{-1}} &\xrightarrow{P} \lim_{T\to\infty} \frac{(\beta'((\sqrt{T}\frac{1}{\sqrt{T}}\sum_{t=0}^{T-1}(u_t u_t' - Eu_t u_t') + TE(u_t u_t'))\beta - G_2 G_4^{-1} G_3)^{-1}}{T^{-1}} \\ &= g_1^{-1} + T^{-\frac{1}{2}}\tilde{G}_1. \end{aligned}$$

$\tilde{G}_1$ is the random variable defined by CLT for $\beta'\frac{1}{\sqrt{T}}\sum_{t=0}^{T-1}(u_t u_t' - Eu_t u_t')\beta$. similarly, we have

$$\frac{G_{5,T}}{T} \xrightarrow{P} = g_5 + T^{-\frac{1}{2}}\tilde{G}_5.$$

Inserting these two expressions into (8.38)

$$T\hat{\Pi} = \begin{pmatrix} G_{1,T} & G_{2,T} \\ G_{3,T} & G_{4,T} \end{pmatrix}^{-1} \begin{pmatrix} G_{5,T} \\ G_{6,T} \end{pmatrix} \sim \begin{pmatrix} g_1^{-1} g_5 T + g_1^{-1}\tilde{G}_5 T^{\frac{1}{2}} \\ -g_1 G_4^{-1} G_2 g_5 + G_6 G_4 \end{pmatrix}$$

Following Lemma 8.2, we know that the order of the eigenvalues is the same as the order of the elements of the corresponding subblock of the matrix. Hence we have for the ith eigenvalue of the matrix $T\hat{\Pi}$ from the jth draw:[19]

[19] Actually beside the leading term $g_{i,j}T$ we have at least one term with rate of convergency of $\sqrt{T}$ as given in the formula. This is enough for our proof. For Simplicity of presentation we assume the second order leading term is $\sqrt{T}\tilde{G}_{i;j}$.

$$\lambda_{i,T;j} \sim \tilde{g}_{i,j}T + \tilde{G}_{i;j}\sqrt{T} \quad \text{for j = 1,2,...,M; i = 1,2,...h,} \tag{8.52}$$

where $\sim$ reads "has the same order as". $\tilde{g}_{i,j}T$ is the leading term of order T. Because The constant $\tilde{g}_{i,j}$ is identical for all j, the index j can be omitted. We define $\overline{\lambda}_{i,j,T}$ as the average over m draws:

$$\overline{\lambda}_{i,T} = \frac{1}{m}\sum_{j=1}^{M} \lambda_{i,T;j}. \tag{8.53}$$

For m large enough we have

$$\overline{\lambda}_{i,T} \sim \tilde{g}_i T \tag{8.54}$$

because the second term in (8.52) has mean zero.

It follows then:

$$\lambda_{i,T;j} - \overline{\lambda}_{i,T} \sim \tilde{G}_{i,j}\sqrt{T} \quad \text{for i = 1,2,...h.} \tag{8.55}$$

This proofs the first part of Theorem 5.1. The proof of the second part is straight foreword.

$$\begin{aligned} & \lim_{T\to\infty, m\to\infty} (\lambda_{i,T} - \overline{\lambda}_{i,T}) \\ = & \lim_{T\to\infty} \lambda_{i,T} - \lim_{m\to\infty} \frac{1}{m}\sum_{j=1}^{m} \lim_{T\to\infty} \lambda_{i,T;j} \\ = & \lambda_i - \lim_{m\to\infty} \frac{1}{m}\sum_{j=1}^{m} \lambda_{i;j} \\ = & \lambda_i - E(\lambda_i). \end{aligned}$$

□

References

Abuaf, N. and Jorion, P. (1990). Purchasing power parity in the long run. *Jouranl of finance*, 45(1):157–74.

Bai, J. and Ng, S. (2004). A panic attack on unit roots and cointegration. *Econometrica*, No.72:1127–1177.

Bierens, H. (1997). Nonparametric cointegration analysis. *Journal of Econometrics*, No.77:379–404.

Boswijk, H. P. and Lucas, A. (2002). Semi-parametric cointegration testing. *Journal of econometrics*, No.108:253–280.

Breitung, J. (2002). Nonparametric tests for unit roots and cointegration. *Journal of econometrics*, No.108:343–363.

Choi, I. (2001). Unit root tests for panel data. *Jouranl of International Meney and Finance*, 20:249–272.

— (2003). Subsampling vector autoregressive tests for linear constraints. *Department of Economics, Hongkong University of Science and Technology*.

Enders, W. (1988). Arima and cointegration tests of ppp under ficed and flexible exchange rate regimes. *Review of Economics and Statistics*, 70:504–508.

Engle, R. F. and Granger, W. J. (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica*, 55:251–276.

Fleissig, A. R. and Strauss, J. (2000). Panel unit root tests of purchasing power parity for price indices. *Jouranl of International Meney and Finance*, 19:489–506.

Föllmer, H. (1981). Calcul des ito senc probabilites. *In: Sem. Probabilites Strasbourg XV, Lecture Note in Mathematics 850*, pages 143–150.

Frankel, J. (1981). The collaps of purchasing power parity during the 1970's. *European Economic Review*, 16(1):145–165.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, 1st edition.

Haug, A. (1996). Tests for cintegration a monte carlo comparison. *Journal of Econometrics*, 71:89–115.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gausian vector autoregressive models. *Econometrica*, 59:1151–1180.

— (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, 1st edition.

Johansen, S. and Juselius, K. (1992). Testig structural hypotheses in a multivariate cointegration anlysis of ppp and uip for uk. *Journal of Econometrics*, 53:211–244.

Kim, Y. (1990). Purchasing power parity: Another look in the long run data. *Economic Letters*, 32:339–344.

Kwiatkowsi, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54:91–115.

Oh, K. (1996). Purchasing power parity and unit root tests usng panel data. *Jouranl of International Meney and Finance*, 15:405–418.

Phillips, P. C. (1991). Optimal inference in cointegrated systems. *Econometrica*, 59:283–306.

Phillips, P. C. and Quliaris (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica*, 58:165–193.

Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling*. Springer, 1st edition.

Sims, C., Stock, J. H., and Watson, M. W. (1990). Inference in linear time series models with some unit roots. *Econometrica*, 58:113–144.

Stock, J. H. and Watson, M. W. (1988). Testing for common trends. *Jounal of American Statistical Association*, 83:1097–1107.