Searching for cointegration in a dynamic system

Zhongjun Qu*

Boston University

This version: September 1, 2004

Abstract

Outlying events or regime changes may mask the cointegrating relationships, making cointegration tests fail to detect them. In this paper we propose a procedure to detect such relationships. First, we propose a test against the existence of cointegration, or a higher cointegration rank, in some sub-sample with unknown timing. If the test rejects, then we use an estimator to consistently estimate the boundary of such regimes. We illustrate the procedure with two applications. The first application is pertaining to the test of Expectation Hypothesis and the second is a study of the linkage between European interest rates. Useful features of the procedure are: it is simple to implement; it is valid in a multivariate setting and it applies when multiple regime shifts exist.

JEL Classification Number: C12; C32

Keywords: Cointegration, Expectation hypothesis, Hypothesis testing, Model selection, Structural change,

*Address: Economics Department, Boston University, Boston, MA, 02215. Email: qu@bu.edu. Tel.: 617-3534455. I am grateful to Professor Pierre Perron for his guidance. I also want to thank Eiji Kurozumi for his comments. The errors are mine. The gauss codes used to implement the tests and to simulate the critical values are available upon request.

1 Introduction:

In practice, cointegration tests often fail to find the cointegration relationship implied by economic theory. One reason could be structural instability, which may arise in various situations, for example, war, economic crisis, or regime change. In some cases the cointegration may be temporarily switched off, in others the cointegrating vector may have changed permanently. Accordingly, simply casting a cointegration test using the sample at hand often fails to be informative. Statistical procedures are needed to account for such instability when analyzing cointegration. Recently, there has been some progress in this direction. For example, Siklos and Granger (1997) propose the notion of *"regime-sensitive cointegration"* and apply it to the study of interest-rate parity. They find that the existence of cointegration is regime dependent. Hansen (2003) finds stronger support for the rational expectation hypothesis, using US data, by separating the interest rate target regime from the regime with *non-borrowed reserves operating procedure*. Zhou (2004) identifies the close interest rates linkage within the European Monetary System by separating the sample into three different regimes. Other works include Kleimeier and Sander (2000), among others.

However, in most studies, the timming of the regime changes or outlying events is assumed to be known. However, in most cases it is determined endogenously. This raises the issue of pre-testing and the induced distorted size of the resulting test procedure. Hence, we need statistical tools which can detect the regime change endogenously and a distribution theory that allows for this feature.

In this paper, we propose a procedure to search for cointegration and to determine the location of the changes endogenously. Our formulation of the problem is based on the fact that when there is a regime shift, the standard cointegration tests based on the full sample will find no cointegration or will under-specify the cointegration rank. However, such a result is inconsistent with at least one subset of the sample. We can build tests exploring such inconsistencies and answer whether cointegration exists in some sub-sample. More specifically, we test the null hypothesis of r_0, which can be zero, stable cointegrating vectors against the alternative of more than r_0 cointegrating vectors existing in some sub-sample. If a rejection of the null hypothesis occurs, we then use a method to consistently estimate the locations of the change points and base subsequent inference on the partitioned sample.

Our test has several useful features. It can detect cointegration even when multiple regime shifts happen. It is eigenvalue-based and hence does not involve arbitrary normalization of the cointegrating vector. It can be used against a broad class of alternatives. A simple extension of the test allows us to test for the structural change near the boundary of the sample, which is useful for out of sample prediction.

We illustrate our procedure with two empirical applications. The first is pertaining to the test of the Expectation Hypothesis. We show that regime change helps to explain the rejection of the hypothesis. The second application is a study of the interest rates linkage between European countries. Again we show that it is important to account for the regime change when trying to identify the linkage between the series. Our result reveals the unattractiveness of modelling a relationship either by imposing cointegration over the sample of interest or by ignoring it altogether, which was also stressed by Saiko and Granger (1997). We conclude that our procedure helps to provide a statistically adequate treatment for such a situation

The organization of the paper is as follows. Section 2 proposes the test for regime sensitive cointegration. Section 3 investigates its finite sample properties and propose a simple correction to improve the size of the test. Section 4 introduces an estimator which consistently estimates the location of the structural changes. Section 5 contains simulations studying the finite sample size and power. Section 6 illustrates our procedure with two applications. All proofs are collected in the appendix.

2 Testing the homogeneity of the cointegration rank

2.1 The null and the alternative hypothesis

Suppose we have an n vector of series, denoted as $\{Y_t\}_{t=1}^{T}$ where T is the number of observations. The series are assumed to be non-trending. Let r denote the cointegrating rank, which could be zero if the series are all $I(1)$ and are not cointegrated.

Suppose each series admits the following representation, under the null hypothesis that no structural instability exists in the system:

$$\begin{aligned} y_{it} &= \alpha_i + u_{it} \\ u_{it} &= u_{it-1} + e_{it} \end{aligned} \tag{1}$$

where α_i is the deterministic component for series i and u_{it} is the mean zero $I(1)$ component. Define $\alpha = [\alpha_1, ...\alpha_n]$ and $U_t = [u_{1t}, ...u_{nt}]'$ and $E_t = [e_{1t}, ...e_{nt}]'$. We make the following assumptions regarding the Data Generating Process (DGP) under the null hypothesis. As a matter of notation, we use [.] for the integer part of the value in the bracket and use $\Rightarrow$ for weak convergence under skorohod topology.

- **Assumption 1**: $T^{-1/2}U_{[Ts]} \Rightarrow \Sigma^{1/2}W(s)$, where $W(s)$ is an n-dimensional Wiener process and Σ is a positive semi-definite matrix not necessarily of full rank.

- **Assumption 2**: There exists an invertible matrix $H = [\beta, \gamma]$, where β and γ are linearly independent $n \times r_0$ and $n \times (n - r_0)$ matrices, respectively, with $0 \leq r_0 < n$, such that

    1. $H'(Y_t - \alpha) = \begin{bmatrix} \beta'(Y_t - \alpha) \\ \gamma'(Y_t - \alpha) \end{bmatrix} \equiv \begin{bmatrix} v_t \\ \xi_t \end{bmatrix}$,
    2. $T^{-1/2}\xi_{[Ts]} \Rightarrow \Omega^{1/2}W_{n-r_0}(s)$, with $W_{n-r_0}(s)$ being an $n - r_0$ dimensional Wiener process and Ω a positive definite matrix.
    3. $T^{-2}\sum_{t=1}^{[Ts]} v_t v_t' \rightarrow 0$ uniformly in s.

Assumption 1 and 2 determines the cointegration property of the process, which are similar to the ones used by Breitung (2002). The assumptions imply that the series can be decomposed into an $n - r_0$ dimensional vector of stochastic trend components and an r_0 vector of transitory components. The partial sum of the error process E_t is required to satisfy a functional central limit theorem, which allows substantial correlation and heterogeneity.

Here the matrix H applies a rotation to the system. A constant H plus the convergence of $T^{-1/2}\xi_{[Ts]}$ to a Brownian motion process implies that the system is stable. Hence, there is no change in the cointegrating space. The parameter α is constant hence there is no structural change in the mean. Also, $T^{-2}\sum_{t=1}^{[Ts]} v_t v_t' \rightarrow 0$ means that the transitory components are dominated by the stochastic trends. Note v_t needs not be a linear process. Instead, it can be generated by a nonlinear process with short memory properties, as pointed by Breitung (2002).

We test for r_0 stable cointegrating relationships against the existence of more than r_0 cointegrating relationships in some sub-sample. If $r_0 = 0$, this becomes a test for no cointegration against segmented cointegration, which was treated in Kim (2003). If $n = 1$, this is a test against a change in persistence as considered in Leybourne, Kim, Smith and Newbold (2003). Our framework is therefore quite general.

More specifically, our null (H_0) and alternative hypothesis (H_1) are as follows:

- H_0 : There exists r_0 stable cointegrating relationships in the system, i.e. the cointegrating rank is r_0.

- H_1 : There exists $r > r_0$ cointegrating relationships in some sub-sample $[T_1, T_2]$, i.e. there exists at least one regime with cointegrating rank higher than r_0.

Remark 1 *We assume T_1 and T_2 to be unknown. In particular we do not assume $T_1 = 1$ or $T_2 = T$. Hence the segment of interest can be located anywhere in the sample. Also, we do not assume that there is only one shift. Instead, multiple changes are allowed and our test will be consistent against such alternatives. Finally, as will be seen later, the existence of a segment with cointegrating rank less than r_0 does not affect the consistency of the test. To have consistency, all we need is an asymptotically non-vanishing segment with cointegrating rank higher than r_0.*

Remark 2 *We do assume that $lim_{T\to\infty} (T_2 - T_1)/T > 0$, i.e. the regime with higher cointegrating rank is non-vanishing asymptotically. This is because our procedure is asymptotic in nature.*

2.2 The test statistics

2.2.1 A test statistic allowing for m regime changes under the alternative

We will construct a test which allows m regime changes, hence $m + 1$ regimes, under the alternative hypothesis. For the test to be consistent, we need that at least one regime has a cointegrating rank higher than r_0. We also need to impose a minimal length for a single regime, which is called the trimming parameter and denoted as ε. Different specifications of ε will lead to different asymptotic distributions. A larger ε reflects less uncertainty about the system, which may consequently deliver higher power. In the following, we set $\varepsilon = 0.2$, leaving optimal choice of ε for future research.[1]

For a given partition of the sample $(T_0, T_1, ..., T_{m+1})$, with the convention that $T_0 = 1, T_{m+1} = T$, we construct the demeaned process $\hat{U}_t$. It is obtained by regressing Y_t on a constant segment by segment. More specifically, for $t \in [T_k + 1, T_{k+1}]$, $\hat{U}_t = Y_t - \sum_{T_k+1}^{T_{k+1}} Y_t = Y_t - \bar{Y}_k$. More concisely,

$$\hat{U}_t = \sum_{k=0}^{m} 1_{(T_k+1 \leq t \leq T_{k+1})} \left(Y_t - \bar{Y}_k\right)$$

Note that we demean the series segment by segment because we want the test to be consistent against mean shifts. We could also demean the series under H_0 if the possibility of mean

[1] For other values of ε, the relevent quantiles of the limit distribution can be simulated using our gauss code, which is available upon request.

shifts is excluded. We think, however, that this would not be appropriate for most economic applications.

Consider the moment matrices:

$$A_k = (T_{k+1} - T_k)^{-2} \sum_{t=T_k+1}^{T_{k+1}} \hat{U}_t \hat{U}_t'$$

$$B_k = (T_{k+1} - T_k)^{-4} \sum_{t=T_k+1}^{T_{k+1}} (\sum_{j=T_k+1}^{t} \hat{U}_j)(\sum_{j=T_k+1}^{t} \hat{U}_j)'$$

for $k = 0, ..., m$. Note that A_k corresponds to the second moment of $\hat{U}_t$ and B_k to the second moment of the partial sum, within regime k $(k = 0, ..., m)$. Now, define

$$Q_k = (B_k)^{-1/2} A_k (B_k)^{-1/2}$$

the ratio of the moment matrices. Let $\rho_i(Q_k)$, $i = 1, ..., n$, denote the ith smallest eigenvalues of Q_k. Hence, $(\rho_1(Q_k), ..., \rho_n(Q_k))$ are the ordered solutions to the eigenvalue problem

$$|\rho B_k - A_k| = 0$$

The construction of Q_k follows Breitung (2002) and is motivated by the fact that the degree of cointegration affects the rank of the moment matrices, which is captured by the eigenvalues of Q_k. When there exists r_0 cointegrating vectors, the $n - r_0$ smallest eigenvalues of Q_k converge to a non-degenerate distribution, the r_0 largest eigenvalues diverges to infinity. Hence, when the cointegrating rank is higher than r_0, the sum of the smallest $n - r_0$ eigenvalues diverges to infinity. This fact is used by Breitung (2002) to construct non-parametric tests for unit root and no cointegration. More generally, this is also the idea behind variance-ratio type statistics, which have been the subject of extensive studies, see Nyblom (1989), Nyblom and Harvey (2000). Johansen (1991) also formulates his cointegration test as an eigenvalue problem. The main difference is that our construction is based on the partial sum of $I(1)$ processes while Johansen's approach is based on first differences. Hence, both approaches are complementary.

Since our alternative only specifies a cointegrating rank higher than r_0 in some subsample, we need to explore the behavior of the smallest $n - r_0$ eigenvalues over all admissible partitions of the sample. More specifically, for an admissible partition $(T_1, ..., T_m)$, construct

$$Q^m (T_1, ..., T_m) = \sum_{k=0}^{m} \sum_{i=1}^{n-r_0} \rho_i(Q_k)$$

It is the sum of the $n - r_0$ smallest eigenvalues over the $m + 1$ segments specified by the partition $(T_1, ..., T_m)$. We shall show that under the null hypothesis of r_0 stable cointegrating vectors, $Q^m(T_1, ..., T_m)$ converges to a non-degenerate distribution. Under the alternative, it diverges to infinity for at least one partition. This partition is, however, unknown, hence a search is needed. To this end, define $\Pi = (T_1, ..., T_m)$ and

$$\Pi_\varepsilon = \{(T_1, ..., T_m); |T_{k+1} - T_k| \geq \varepsilon T, T_1 \geq \varepsilon T, T_m \leq (1 - \varepsilon) T\}$$

The test statistic is then defined as

$$SupQ^m = \sup_{\Pi \in \Pi_\varepsilon} \sum_{k=0}^{m} \sum_{i=1}^{n-r_0} \frac{\rho_i(Q_k)}{100\,(n - r_0)}$$

We have sclaled the test by $100\,(n - r_0)$ to control the critical values. Note that when $m = 0$, no break is allowed and the test reduces to Breitung's (2002) non-parametric test for unit root and cointegration. His simulation shows that the test has size and power comparable to the existing parametric tests, for example, Augmented Dickey-Fuller test. The test may enjoy better size and power when certain forms of non-linearity are present.

One nice feature of the above test statistic is that the null limiting distribution does not depend on the parameters determining the short run dynamics. It is also rotation free which means that it does not involve arbitrary normalization of the cointegrating vector. Indeed, the null limit distribution only depends on $(n - r_0, \varepsilon)$. More specifically, we have the following result:

Proposition 1 *Define* $\Lambda = (\lambda_1, ...\lambda_m)$ *with* $\lambda_i = T_i/T$ *and for a given* $\varepsilon > 0$, $\Lambda_\varepsilon = \{(\lambda_1, ..., \lambda_m); |\lambda_{k+1} - \lambda_k| \geq \varepsilon, \lambda_1 \geq \varepsilon, \lambda_m \leq (1 - \varepsilon)\}$. *Under Assumption 2, and under the null hypothesis that the n-vectors* Y_t *has cointegrating rank* r_0, *we have, as* $T \to \infty$,

$$SupQ^m \Rightarrow \frac{1}{100\,(n - r_0)} \sup_{\Lambda \in \Lambda_\varepsilon} tr \left\{ \sum_{k=0}^{m} (\lambda_{k+1} - \lambda_k)^2 \int_{\lambda_k}^{\lambda_{k+1}} W_\lambda(s) W_\lambda(s)' ds \left[\int_{\lambda_k}^{\lambda_{k+1}} V_\lambda(s) V_\lambda(s)' ds \right]^{-1} \right\}$$

where $W_\lambda(s)$ *is the residual function from the regression of the* $n - r_0$ *dimensional Wiener process on a constant segment by segment, defined as*

$$W_\lambda(s) = \sum_{k=0}^{m} 1_{s \in (\lambda_k, \lambda_{k+1}]} \left(W(s) - \int_{\lambda_k}^{\lambda_{k+1}} W(r)\, dr \right)$$

with $W(s)$ *an* $n - r_0$ *dimensional Wiener process. Also,* $V_\lambda(s)$ *is defined as*

$$V_\lambda(s) = \sum_{k=0}^{m} 1_{s \in (\lambda_k, \lambda_{k+1}]} \int_{\lambda_k}^{s} W_\lambda(a)\, da$$

Consider the behavior of $SupQ^m$ under the alternative. In this case, there will be at least one regime with cointegrating rank $r > r_0$. Then, when we search over Π_ε, there will be at least one partition that contains a segment with cointegrating rank r. For this segment, say the kth segment, Q_k has $n-r$ eigenvalues of order $O_p(1)$ and r eigenvalues diverging to infinity, which means that $\sum_{i=1}^{n-r_0} \rho_i(Q_k)$ diverges. Hence, the test is consistent under the alternative.

Some remarks are in order. First, our approach is non-parametric. Alternatively, one can use a vector error correction model (ECM) and do everything parametrically. The advantage of our approach is that we do not need to choose the order of the autoregression. In our case, an appropriate specification of the lag length is particularly difficult because we would need to run the regression using different sample sizes at least $O(T^2)$ times. The appropriate choice of lag order in a fixed sample is already a delicate problem, especially in a multivariate context. Another advantage that our test is robust to non-linear transitory dynamics. Hence it is expected to show robust size when some nonlinearity is present.

However, the nonparametric test does have its draw back. Although the serial correlation does not enter the distribution asymptotically, they do affect the finite sample distribution. In other words, the nonparametric test automatically corrects for the serial correlation only to some degree. The correction might not be sufficient in small samples. In fact, as we shall see, the test suffers from over rejection when the mean reversion is strong and it is conservative when the first differenced series are positively correlated. We will come back to this in the next section and propose a simple procedure to improve the size of the test. It will be seen that the improvement is important.

2.2.2 The WQ and the SQ test

So far the number of regimes was taken as given, we now propose a modified test that allows an unknown number of regimes.

Suppose that we do not know the number of regimes under the alternative. Can we simply construct the test assuming $m = 1$? The answer is no. To see this, consider the following situation. Suppose we have a system of integrated processes. Three regimes exist in the sample, which are $\left[1, \frac{1}{3}T\right]$, $\left[\frac{1}{3}T+1, \frac{2}{3}T\right]$,and $\left[\frac{2}{3}T+1, T\right]$. Suppose cointegration only exists in the second regime. Then the test assuming $m = 1$ is not consistent. This is because with only two segments, no matter how we do the partition, there can not be a single segment which only contains observations from the second regime. On the other hand, using a large m may yield a test with low power even though it would be consistent.

It turns out that $m = 2$ is sufficient to generate consistency for any unknown number of regimes. This is because using $m = 2$ can yield a segment containing only the regime with a higher cointegrating rank under at least one partition of the sample. This is sufficient to yield a diverging statistic as $T \to \infty$.

Given these considerations, we construct two tests, called WQ and SQ, which follows the construction of Bai and Perron (1998).

- The test WQ is constructed to make the critical value at given significance level to be equalized across $m = 1$ and $m = 2$. More specifically

$$WQ = \max\left(SupQ^1, SupQ^2 * \frac{c(\alpha, 1, n - r_0, \varepsilon)}{c(\alpha, 2, n - r_0, \varepsilon)} \right)$$

  where $c(\alpha, i, n-r, \varepsilon)$ is the critical value of the test against i breaks at significance level α. $n - r_0$ stands for the number of stochastic trends and ε stands for the trimming.

- The test SQ is constructed in a similar way but computes the sum instead of the maximum.

$$SQ = SupQ^1 + SupQ^2 * \frac{c(\alpha, 1, n - r, \varepsilon)}{c(\alpha, 2, n - r, \varepsilon)}$$

We study the finite sample size and power of these tests in section 5. What will transpire from the results is that there is not much power loss from not knowing the exact number of regimes under the alternative.

2.2.3 Asymptotic critical values of the tests

Now we present the critical values of the tests we proposed above. The values are obtained via simulations. The Wiener process $W_{n-r_0}(r)$ is approximated by the partial sums $T^{-1/2}\sum_{t=1}^{[Tr]} e_t$ with $e_t \sim i.i.d\ N(0, I_{n-r_0})$ and $T = 1,000$. The number of replications is $10,000$. For each replication, the supremum of $Q^m(\lambda_1, ..., \lambda_m)$ with respect to $(\lambda_1, ..., \lambda_m)$ is obtained via a dynamic programming algorithm as described in Bai and Perron (2003). We present, in Table 1, critical values covering cases with up to 3 breaks and up to 5 stochastic trends.

Table 1. Asymptotic critical values of the $SupQ^m$

The entries are quantiles x such that $P\left(SupQ^m \leq x\right) = \alpha$

		number of breaks, m				
$q = n - r_0$	α	1	2	3	WQ	SQ
1	.90	2.688	4.460	4.763	3.125	5.110
	.95	3.380	5.316	5.601	3.906	6.279
	.99	4.826	7.387	7.539	5.275	8.644
2	.90	3.938	6.357	7.211	4.295	7.553
	.95	4.449	7.098	8.042	4.857	8.466
	.99	5.810	8.633	9.822	6.350	10.922
3	.90	5.461	8.439	10.271	5.851	10.637
	.95	6.008	9.177	11.019	6.381	11.583
	.99	7.283	10.669	12.840	7.673	14.098
4	.90	7.366	11.257	13.896	7.764	14.475
	.95	7.921	12.015	14.639	8.410	15.466
	.99	9.039	13.435	16.563	9.532	17.427
5	.90	9.533	14.584	18.218	10.023	18.849
	.95	10.180	15.403	18.988	10.630	19.855
	.99	11.143	16.624	20.881	11.532	21.513

2.3 Other useful tests with more prior information about the break points

When we have some prior information about the break dates, it makes sense to incorporate it into the test. For example, we may know that a regime shift happens close to the end of the sample. Or we could know that the cointegration is not stable at the beginning. This leads to some variants of the tests, presented below.

2.3.1 The forward recursive test

Suppose we conjecture that the relationship is perturbed only at the end of the sample. This could be incorporated into the test to improve power. We can then use a forward recursive procedure which generates the test:

$$SupQ^F = \frac{1}{100\,(n-r_0)} \sup_{t\in[T_1+1,T]} \sum_{i=1}^{n-r_0} \rho_i((B_{1,t})^{-1/2} * A_{1,t} * (B_{1,t})^{-1/2})$$

where $A_{1,t}$ is the matrix A constructed using observations from 1 to t and similarly for $B_{1,t}$. Additional power gains can be achieved by specifying a large trimming $\varepsilon = T_1/T$ and only searching forward. The limiting distribution of the test is stated in the following corollary.

Corollary 1 *For a given $\varepsilon > 0$, define $\Lambda_\varepsilon^F = \{\lambda_1 : \varepsilon \leq \lambda_1 \leq 1\}$. Under Assumption 2, and under the null hypothesis that the n-vectors Y_t has cointegrating rank r_0, we have, as $T \to \infty$,*

$$SupQ^F \Rightarrow \frac{1}{100\,(n-r_0)} \sup_{\lambda_1\in\Lambda_\varepsilon^F} tr\left\{\lambda_1^2 \int_0^{\lambda_1} W_{\lambda_1}^F(s)\, W_{\lambda_1}^F(s)'\, ds \left[\int_0^{\lambda_1} V_{\lambda_1}^F(s)\, V_{\lambda_1}^F(s)'\, ds\right]^{-1}\right\}$$

where $W_{\lambda_1}^F(s)$ is defined as

$$W_{\lambda_1}^F(s) = 1_{s\in(0,\lambda_1]}\left(W(s) - \int_0^{\lambda_1} W(r)\, dr\right)$$

with $W(s)$ being $n-r_0$ dimensional Wiener process. Also, $V_{\lambda_1}^F(s)$ is defined as

$$V_{\lambda_1}^F(s) = 1_{s\in(0,\lambda_1]} \int_0^s W_{\lambda_1}^F(a)\, da$$

The distribution is easily tabulated. The critical values are presented at Table 2 of the next page, which are generated by the same procedure used for Table 1.

2.3.2 The reverse recursive test

Suppose now it is believed that the system is perturbed only at the beginning of the sample, then we can use a reverse recursive test defined by

$$SupQ^R = \frac{1}{100\,(n-r_0)} \sup_{t\in[1,T_1]} \sum_{i=1}^{n-r_0} \rho_i((B_{t,T})^{-1/2} * A_{t,T} * (B_{t,T})^{-1/2})$$

Table 2. Asymptotic critical values of the forward recursive test $SupQ^F$

The entries are quantiles x such that $P\left(SupQ^F \leq x\right) = \alpha$

		$q = n - r_0$				
ε	α	1	2	3	4	5
	.90	1.895	2.708	3.626	4.680	5.876
0.2	.95	2.504	3.256	4.111	5.243	6.414
	.99	4.010	4.593	5.395	6.376	7.752
	.90	1.556	2.291	3.244	4.316	5.515
0.4	.95	2.070	2.855	3.805	4.796	6.056
	.99	3.725	3.984	5.059	5.992	7.204
	.90	1.295	2.049	2.924	3.958	5.065
0.6	.95	1.746	2.493	3.429	4.428	5.552
	.99	3.196	3.643	4.554	5.607	6.717
	.90	0.980	1.752	2.624	3.599	4.729
0.8	.95	1.376	2.197	3.095	4.068	5.272
	.99	2.392	3.298	4.229	5.236	6.435

Again, additional power gain can be achieved by specifying a large trimming $\varepsilon = (T - T_1)/T$ and only searching backward. The distribution of the test is given in the following corollary and the critical values are presented in table 3.

Corollary 2 *For a given $\varepsilon > 0$, define $\Lambda_\varepsilon^R = \{\lambda_1 : 0 \leq \lambda_1 \leq 1 - \varepsilon\}$. Under assumption 2, and under the null hypothesis that the n-vectors Y_t has cointegrating rank r_0, we have, as $T \to \infty$,*

$$SupQ^R \Rightarrow \frac{1}{100\,(n - r_0)} \sup_{\lambda_1 \in \Lambda_\varepsilon^R} tr \left\{ (1 - \lambda_1)^2 \int_{\lambda_1}^1 W_{\lambda_1}^R(s)\, W_{\lambda_1}^R(s)'\, ds \left[\int_0^{\lambda_1} V_{\lambda_1}^R(s)\, V_{\lambda_1}^R(s)'\, ds \right]^{-1} \right\}$$

where $W_{\lambda_1}^R(s)$ is defined as

$$W_{\lambda_1}^R(s) = 1_{s \in (\lambda_1, 1]} \left(W(s) - \int_{\lambda_1}^1 W(r)\, dr \right)$$

with $W(s)$ being $n - r_0$ dimensional Wiener process. Also, $V_{\lambda_1}^R(s)$ is defined as

$$V_{\lambda_1}^R(s) = 1_{s \in (\lambda_1, 1]} \int_{\lambda_1}^s W_{\lambda_1}^R(a)\, da$$

Table 3. Asymptotic critical values of the reverse recursive test $SupQ^R$

The entries are quantiles x such that $P\left(SupQ^R \leq x\right) = \alpha$

		$q = n - r_0$				
ε	α	1	2	3	4	5
	.90	1.884	2.674	3.592	4.681	5.773
0.2	.95	2.365	3.157	4.108	5.186	6.498
	.99	3.861	4.545	5.218	6.384	7.688
	.90	1.542	2.343	3.232	4.352	5.383
0.4	.95	2.024	2.831	3.723	4.798	6.139
	.99	3.434	4.256	4.828	6.041	7.454
	.90	1.282	2.055	2.918	3.956	5.030
0.6	.95	1.793	2.527	3.365	4.541	5.748
	.99	2.883	3.706	4.556	5.612	6.965
	.90	1.013	1.787	2.556	3.602	4.564
0.8	.95	1.471	2.244	3.030	4.106	5.328
	.99	2.435	3.155	4.173	5.210	6.316

2.3.3 The rolling window test

Suppose now we conjecture that the cointegration relationship is stable at the middle of the sample with a cointegrating rank $r > r_0$, which is perturbed at the beginning and/or the end. Then we could use the rolling window test defined by

$$SupQ^W = \frac{1}{100\,(n-r_0)} \sup_{(t_1,t_2)\in\Pi_\varepsilon^W} \sum_{i=1}^{n-r_0} \rho_i((B_{t_1,t_2})^{-1/2} * A_{t_1,t_2} * (B_{t_1,t_2})^{-1/2})$$

with

$$\Pi_\varepsilon^W = \{(T_1, T_2)\,; T_2 - T_1 \geq \varepsilon T, T_1 \geq 1, T_2 \leq T\}$$

Additional power gain can be achieved by allowing a large trimming at the middle of the sample. The limiting distribution follows directly and critical values are presented in table 4.

Corollary 3 *Define* $\Lambda_{\varepsilon}^{W}=\{(\lambda_1,\lambda_2);\lambda_2-\lambda_1\geq\varepsilon,\lambda_1\geq 0,\lambda_2\leq 1\}$. *Under Assumption 2, and under the null hypothesis that the n-vectors* Y_t *has cointegrating rank* r_0,*we have, as* $T\to\infty$,

$$SupQ^{W}\Rightarrow\frac{1}{100\left(n-r_0\right)}\sup_{\Lambda\in\Lambda_{\varepsilon}^{W}}tr\left\{\left(\lambda_2-\lambda_1\right)^2\int_{\lambda_1}^{\lambda_2}W_{\lambda}^{W}\left(s\right)W_{\lambda}^{W}\left(s\right)'ds\left[\int_{\lambda_1}^{\lambda_2}V_{\lambda}^{W}\left(s\right)V_{\lambda}^{W}\left(s\right)'ds\right]^{-1}\right\}$$

where $W_{\lambda}^{W}\left(s\right)$ *is defined as*

$$W_{\lambda}^{W}\left(s\right)=1_{s\in(\lambda_1,\lambda_2]}\left(W\left(s\right)-\int_{\lambda_1}^{\lambda_2}W\left(r\right)dr\right)$$

with $W\left(s\right)$ *being* $n-r_0$ *dimensional Wiener process. Also,* $V_{\lambda}^{W}\left(s\right)$ *is defined as*

$$V_{\lambda}^{W}\left(s\right)=1_{s\in(\lambda_1,\lambda_2]}\int_{\lambda_1}^{s}W_{\lambda}^{W}\left(s\right)\left(a\right)da$$

Table 4. Asymptotic critical values of the rolling window test $SupQ^{W}$

The entries are quantiles x such that $\mathrm{P}\left(SupQ^{W}\leq x\right)=\alpha$

		$q=n-r_0$				
ε	α	1	2	3	4	5
	.90	4.450	5.033	5.826	7.041	8.174
0.2	.95	5.324	5.617	6.429	7.659	8.829
	.99	7.458	7.376	8.017	9.069	10.065
	.90	3.119	3.798	4.856	5.969	7.097
0.4	.95	3.841	4.522	5.415	6.630	7.768
	.99	5.472	5.816	6.804	7.923	8.946
	.90	2.153	3.010	3.951	5.109	6.213
0.6	.95	2.778	3.602	4.560	5.681	6.781
	.99	4.473	5.138	5.689	7.310	8.019
	.90	1.391	2.157	3.155	4.202	5.322
0.8	.95	1.935	2.651	3.681	4.816	5.844
	.99	3.584	3.625	4.989	6.319	6.817

3 Finite sample property of the tests and a simple adjustment to improve the size

First, we look at the finite sample size of the tests under a standard DGP of n independent random walks, with $n = 1, 2$ and 3. The null hypothesis is no cointegration at any subsample, which means $r_0 = 0$. We examine the size of the test at the 5% nominal level with $\varepsilon = 0.2$ and $T = 200$. In Table 5, we report the percentage rejection over $1,000$ replications.

Table 5. The exact size of the tests at 5% nominal size under standard DGP

n	SupQ^1	SupQ^2	WQ	SQ
1	4.1	4.4	3.8	4.0
2	5.1	3.0	4.1	5.0
3	3.9	2.1	3.3	3.4

The above result suggests that the asymptotic distribution provides a reasonable approximation to the finite sample distribution when the DGP specifies series whose first differences are uncorrelated. Among the four tests, the size of $SupQ^2$ is affected most by the dimension of the system, while the other three appear more robust.

Next, we introduce transitory dynamics into the system. Specifically, we consider the following two DGPs.

- **DGP 1:** $\Delta Y_t = \rho * \Delta Y_t + e_t$, with $e_t \sim i.i.d.\ N(0, I_n)$, $\rho = -0.5, 0.5$
- **DGP 2:** $\Delta Y_t = e_t + \theta * e_{t-1}$, with $e_t \sim i.i.d.\ N(0, I_n)$, $\theta = -0.5, 0.5$

Again, we look at the exact size with a 5% nominal level, with $\varepsilon = 0.2$ and $T = 200$. The results are summarized in Table 6. It is seen that the size is very unstable. It suggests that although the transitory dynamics does not enter the limit, it does have an important impact on the finite sample distribution. When the series are close to I(2), i.e, the first differenced series show strong positive correlation, the test is then very conservative. When the series are weak I(1) processes, i.e, the first differenced series show strong mean reversion, the test over rejects the null hypothesis. The latter phenomenon is called near-observational equivalence and is widely documented in the unit root literature. The reason is that the transitory component dominates the random walk component in finite sample. This problem also occurs when certain combination of the two series yields an integrated process showing near-observational equivalence. It appears that we need some finite sample correction to improve the finite sample size of the tests.

Table 6. The exact size of the tests at 5% nominal level under different transitory dynamics

AR(1) errors								
	$\rho = -0.5$				$\rho = 0.5$			
n	$SupQ^1$	$SupQ^2$	WQ	SQ	$SupQ^1$	$SupQ^2$	WQ	SQ
1	9.6	13.8	12.0	12.2	2.0	0.2	1.0	1.6
2	15.6	19.5	19.0	21.6	0.7	0.3	0.3	0.3
3	27.9	41.9	39.7	40.7	0.3	0.0	0.3	0.2
MA(1) errors								
	$\theta = -0.5$				$\theta = 0.5$			
n	$SupQ^1$	$SupQ^2$	WQ	SQ	$SupQ^1$	$SupQ^2$	WQ	SQ
1	18.0	26.6	23.4	23.4	2.8	2.1	1.7	2.8
2	37.5	54.5	50.7	53.8	1.9	0.7	1.2	1.5
3	60.2	82.3	78.8	80.7	1.3	0.2	0.6	0.8

3.1 A simple procedure to improve the finite sample size

An ideal solution to the above problem would be to separate the random walk component from the transitory component. Then we could use only the random walk component to construct the test. However, that is not possible in most cases. Fortunately, all we need is to extract the strong serial dependence of the transitory component from the system. The weak dependence would then not affect the distribution very much, hence making the asymptotic distribution a good approximation. The goal can be achieved by the following procedure based on BN decomposition.

Let $\{Y_t\}$ be an n vector of series and not cointegrated, satisfying Assumption 1. The first differenced series can then be well approximated by an infinite order autoregressive process, i.e, $A(L)\Delta Y_t = M_t$, with $A(L) = I + \sum_{i=1}^{\infty} a_i L^i$ and M_t being an n vector of martingale difference sequences. Then, using the BN decomposition,

$$M_t = A(L)\Delta Y_t = A(1)\Delta Y_t + A^*(L)(1-L)\Delta Y_t \qquad (2)$$

with $A^*(L) = -\sum_{i=0}^{\infty} \widetilde{a}_i L^i$ and $\widetilde{a}_i = \sum_{j=k+1}^{\infty} a_j$. Taking the sums, (2) becomes

$$A(1)Y_t + A^*(L)\Delta Y_t = \sum_{j=1}^{t} M_j$$

and

$$Y_t + A(1)^{-1} A^* (L) \Delta Y_t = A(1)^{-1} \sum_{j=1}^{t} M_j \equiv \sum_{j=1}^{t} V_j$$

with $Y_t + A(1)^{-1} A^* (L) \Delta Y_t$ being a multivariate random walk process. In practice, $A(1)$ and $A^* (L)$ are unknown. However, they can be estimated by regressing ΔY_t on the lagged values, using a finite order polynomial of order, say, k. Denoting the resulting estimates with a hat, we have

$$Y_t + \hat{A}(1)^{-1} \hat{A}^* (L) \Delta Y_t = \sum_{j=1}^{t} \hat{V}_j$$

Then $Y_t + \hat{A}(1)^{-1} \hat{A}^* (L) \Delta Y_t$ is approximately a random walk process and we can construct the tests using this quantity instead of Y_t. Notice that $Y_t + \hat{A}(1)^{-1} \hat{A}^* (L) \Delta Y_t$ has the same stochastic trends as Y_t, because the transformation is simply adding a stationary process to an integrated process, which does not affect the stochastic trends. More generally, we have the following proposition:

Proposition 2 *Assume the DGP of Y_t satisfies Assumption 1 and the rank condition satisfies assumption 2. Define $\widetilde{Y}_t = Y_t + \hat{A}(1)^{-1} \hat{A}^* (L) \Delta Y_t$, where $\hat{A}^* (L)$ is the fitted polynomial of the lag operator with the order fixed as $T \to \infty$, then $\widetilde{Y}_t$ satisfies,*

1. $H' \left(\widetilde{Y}_t - \alpha \right) = \begin{bmatrix} \beta' \left(\widetilde{Y}_t - \alpha \right) \\ \gamma' \left(\widetilde{Y}_t - \alpha \right) \end{bmatrix} \equiv \begin{bmatrix} \widetilde{v}_t \\ \widetilde{\xi}_t \end{bmatrix}$,
2. $T^{-1/2} \widetilde{\xi}_{[Ts]} \Rightarrow \Omega^{1/2} W_{n-r_0} (s)$,
3. $T^{-2} \sum_{t=1}^{[Ts]} \widetilde{v}_t \widetilde{v}_t' \to 0$ *uniformly in s.*

The correction has no impact on the cointegrating rank, which means the correction does not affect either the limit distribution under the null or the consistency of the test under the alternative. It is not essential that $\hat{A}(1)^{-1} \hat{A}^* (L)$ be precisely estimated, nor is it important to select the *"true order"* of the autoregression. All we need is to extract the strong correlation. In the rest of the paper, we use $k = 2$, which is motivated by the hump-shaped response typically encountered in economics. We will show that such a correction brings a significant improvement to the size.[2]

[2] Whether a data dependent order chosen by some information criteria could further improve the size is a problem remain to be studied.

When applying to our non-parametric test, theoretically, we can use either the full-sample based or the sub-sample based parameter estimates for the adjustment. However, our simulation shows that, if the sample size is relatively small, as typically encountered in economics, the full-sample based adjustment yields more dependable size, while the latter often appears to be too conservative. Hence full-sample based adjustment is suggested.

When the system has a reduced rank structure, i.e, the series are cointegrated under the null, it is necessary to include lagged dependent variable in the regression when estimating $\hat{A}(1)$ and $\hat{A}^*(L)$. This consideration is also important for the power of the test. Hence we suggest to always include lagged dependent variable and a constant term in the regression. More specifically, our proposal is as follows,

1. Estimate the following regression using the full sample: $\Delta Y_t = u + CY_{t-1} + A_1 \Delta Y_{t-1} + A_2 \Delta Y_{t-2} + e_t$

2. Construct $\hat{A}(1)$ and $\hat{A}^*(L)$ as following: $\hat{A}(1) = I_n - \hat{A}_1 - \hat{A}_2$ and $\hat{A}^*(L) = \hat{A}_1^* + \hat{A}_2^* L$, with $\hat{A}_1^* = \hat{A}_1 + \hat{A}_2$ and $\hat{A}_2^* = \hat{A}_2$.

3. Construct the adjusted process:

$$\widetilde{Y}_t = Y_t + \hat{A}(1)^{-1} \hat{A}^*(L) \Delta Y_t$$

4. Use $\widetilde{Y}_t$ to construct the test statistic.

A final note is that our proposed procedure may have wider use in non-parametric tests. Interestingly, our procedure has some flavor of *"pre-whitening"*, although the application is in a non-stationary framework.

4 Estimating the location of the changes

Once the null hypothesis is rejected, the next step is to estimate the locations of the changes. Unlike the stationary case, the partition of the sample that maximizes $SupQ$ does not necessarily consistently estimate the location of the change. This is because under the alternative hypothesis the appropriately normalized test converges to a random variable instead of a constant and the maxima need not correspond to the change point.

Hence, another approach is needed to find the change points. To this end, we adopt the method proposed by Bai, Lumsdaine and Stock (1998), BLS henceforth. Under the assumption that only one regime change happens, they show that by maximizing the likelihood function, we obtain a consistent estimate of the break fraction under some mild conditions, e.g. a mean shift of order higher than $T^{-1/2}\log T$ and/or a cointegrating vector change of order higher than $T^{-1}\log T$. The convergence rate of the break fraction is then sufficient to guarantee the standard root-T asymptotic normality of the estimated coefficients, which allows the subsequent inference to be conducted as if the break were known.

We conjecture that their results continue to hold when multiple regime changes exist. To show this, we could follow the strategy of Bai and Perron (1998), i.e. the different breaks are asymptotically distinct hence the problem essentially reduces to the one-break case. This strategy is used by Qu and Perron (2004) to prove the consistency of multiple change points estimate, in a linear multivariate regression model with stationary regressors. Here the main complication is that a subset of the regressors is non-stationary. Hence some specific regularity conditions are needed. Although this might be technically demanding, the main results are unlikely to change. In the following, we will use the conjecture without proof, leaving a rigorous treatment for future research.

5 Finite sample performance of the test

5.1 The finite sample size

In this section, we investigate the finite sample size of the tests with the proposed adjustment. We consider the following two DGPs:

- **DGP 3:** $\Delta Y_t = \rho * \Delta Y_t + e_t$, with $e_t \sim i.i.d\ N(0, I_n)$, $\rho = -0.5, -0.2, 0.2, 0.5$.
- **DGP 4:** $\Delta Y_t = e_t + \theta * e_{t-1}$, with $e_t \sim i.i.d\ N(0, I_n)$. $\theta = -0.5, -0.2, 0.2, 0.5$.

We study the exact size at 5% nominal level with $\varepsilon = 0.2$ and $T = 200$. The rejection frequencies are reported in table 8.

Compared with the case with no adjustment, we see a dramatic improvement in the size. In the case of AR(1) errors, the over-rejection problem disappears and the test shows dependable size. In the case of MA errors, the over-rejection problem still persists for some parameter values, although the size is significantly improved. For example, for n=3, the rejection frequency with no data adjustment often goes above 60%, and now it is below 13%. The overall picture shows that the adjustment is quite effective.

Table 8. The exact size of the tests at 5% nominal level using proposed data adjustment

AR(1) errors																
	$\rho=-0.5$				$\rho=-0.20$				$\rho=0.2$				$\rho=0.5$			
n	SupQ[1]	SupQ[2]	WQ	SQ	SupQ[1]	SupQ[2]	WQ	SQ	SupQ[1]	SupQ[2]	WQ	SQ	SupQ[1]	SupQ[2]	WQ	SQ
1	4.4	4.4	4.2	3.6	5.4	5.8	4.4	5.4	5.6	4.6	3.6	5.2	4.4	5.8	5.2	4.8
2	5.2	3.0	4.1	4.4	4.6	3.4	3.7	4.6	4.5	2.2	3.1	3.2	3.8	2.7	2.8	3.0
3	4.3	2.5	3.3	3.8	4.4	2.4	3.6	4.1	3.8	3.1	2.9	3.6	3.8	1.2	2.1	2.1
MA(1) errors																
	$\theta=-0.5$				$\theta=-0.20$				$\theta=0.20$				$\theta=0.50$			
n	SupQ[1]	SupQ[2]	WQ	SQ	SupQ[1]	SupQ[2]	WQ	SQ	SupQ[1]	SupQ[2]	WQ	SQ	SupQ[1]	SupQ[2]	WQ	SQ
1	7.2	7.9	7.1	7.5	4.9	5.3	4.2	4.4	4.7	5.6	4.9	5.8	2.9	2.0	2.5	2.3
2	9.7	8.8	9.1	9.9	5.6	3.1	4.7	4.9	4.4	3.0	3.8	3.6	3.8	1.4	2.9	2.8
3	12.8	10.7	12.3	12.8	6.0	3.0	3.9	4.7	3.0	1.9	2.2	2.3	1.8	0.7	1.3	1.0

In the next, we look at the performance of the adjustment when the data is cointegrated under the null. For that purpose, we adopt the model studied by Yap and Reinsel (1995). More specifically, it is a tri-variate system with an $ARMA(1,1)$ structure, as follows,

$$\Delta Y_t = P\left(\begin{bmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_2 & 0\\ 0 & 0 & \lambda_3\end{bmatrix} - I_3\right) QY_{t-1} + \varepsilon_t - P_\theta\left(\begin{bmatrix}\lambda_\theta & 0 & 0\\ 0 & 0.297 & 0\\ 0 & 0 & -0.202\end{bmatrix}\right) Q_\theta \varepsilon_{t-1}$$

with

$$Q = \begin{bmatrix}-0.29 & -0.47 & -0.57\\ -0.01 & -0.85 & 1.00\\ -0.75 & 1.39 & -0.55\end{bmatrix}, \quad P_\theta = \begin{bmatrix}-0.816 & -0.657 & -0.822\\ -0.624 & -0.785 & 0.566\\ 0.488 & 0.475 & 0.174\end{bmatrix}$$

$$\Sigma_\varepsilon = \begin{bmatrix}0.47 & 0.20 & 0.18\\ 0.20 & 0.32 & 0.27\\ 0.18 & 0.27 & 0.30\end{bmatrix}$$

The value of $(\lambda_1, \lambda_2, \lambda_3)$ determines the cointegration property of the process. More precisely, the number of λ_i with absolute value less than one is the cointegrating rank of the system. A λ_i with value less than but close to one indicates persistent deviation from the equilibrium. The parameter λ_θ controls the strength of mean reversion of the process, which has absolute value less than one to guarantee the invertibility of the MA process. We has chosen $\lambda_\theta = 0$ and $\lambda_\theta = \pm 0.5$ respectively. Again, we use $\varepsilon = 0.2$, and $T = 200$.

Table 9 presents the empirical rejection frequencies under different parameterization of the model. Two facts emerge. First, the effect of λ_θ is effectively reduced. To see this, we may compare the size to the one without adjustment. Take the case with $\lambda_1 = \lambda_2 = \lambda_3 = 1$ as example, the rejection rates without adjustment are above 30% in some cases, but now are uniformly below 9%. Hence there is a dramatic improvement. Second, the tests are conservative when the deviation from the equilibrium is persistent. In this case, the size is not remedied by the proposed adjustment.

Table 9. The exact size of the test in the VARMA(1,1) model using the proposed adjustment

The null	$\lambda_\theta = -0.5$				$\lambda_\theta = 0.0$				$\lambda_\theta = 0.5$			
$(\lambda_1, \lambda_2, \lambda_3)$	$SupQ^1$	$SupQ^2$	WQ	SQ	$SupQ^1$	$SupQ^2$	WQ	SQ	$SupQ^1$	$SupQ^2$	WQ	SQ
$(1, 1, 1)$	5.4	1.8	3.4	3.4	4.7	2.8	3.2	3.5	8.3	6.5	7.3	8.0
$(1, 1, 0.7)$	0.8	0.1	0.2	0.1	1.0	0.1	0.2	0.2	0.6	0.2	0.4	0.6
$(1, 1, 0)$	2.4	0.4	1.3	1.5	2.3	0.7	1.4	1.2	5.3	2.2	3.5	4.1
$(1, 0.7, 0.7)$	0.2	0.0	0.1	0.1	0.2	0.1	0.1	0.1	0.4	0.1	0.1	0.3
$(1, 0, 0)$	2.4	1.0	1.5	1.8	3.1	1.0	1.4	2.0	4.3	3.1	3.6	3.9

5.2 The finite sample power of the test

As mentioned before, our non-parametric approach allows the test to have power against a broad class of alternatives. In particular, we expect it potentially to be able to detect the regime-sensitive cointegration in the presence of changes in the cointegrating rank, in the cointegrating vector or in the deterministic component. In the following, we use some simple simulations to examine the performance of the tests in the above situations.

5.2.1 Changes in the cointegrating rank

We first look at the power of the tests when there are changes in the cointegrating rank. For that purpose, we adopt a DGP similar to Kim (2003). More specifically,

$$\begin{aligned} Y_{1t} &= Y_{1t-1} + v_{1t} \\ Y_{2t} &= \alpha + \beta Y_{1t} + \varepsilon_t \ with \ \varepsilon_t = \rho\varepsilon_{t-1} + v_{2t}, \ if \ t \in \mathcal{C}_t \\ Y_{2t} &= \alpha + \beta Y_{1t} + \varepsilon_t \ with \ \varepsilon_t = \varepsilon_{t-1} + v_{2t}, \ if \ t \in \mathcal{N}_t \\ v_t &= (v_{1t}, v_{2t})' \sim i.i.d \ N(0, 0.01 I_n) \end{aligned}$$

It is a system with n series, with Y_{1t} being an $n-1$ vector of random walks and Y_{2t} a univariate series that subjects to regime changes. We use $t \in \mathcal{C}_t$ to denote that the observation belongs to a cointegrated regime and otherwise $t \in \mathcal{N}_t$. We consider two cases, $\mathcal{C}_t = [0.2T, T]$ and $\mathcal{C}_t = [0.2T, 0.8T]$, which correspond to break at one and both ends of the sample. We let the autoregressive parameter vary between 0.95 and 0.0 to represent decreasing persistence. The other specifications are: $\alpha = 1, \beta = 1_{n-1}$ (n-1 row vector of ones), $T = 200$ and $\varepsilon = 0.2$. Figure 1 and 2 plots the size adjusted power of the tests based on 5% significance level, as a function of the persistence parameter $1-\rho$. We also include the full sample based Johansen's trace test in the figure for the reason of comparison.

Figure 1 pertains to the case $\mathcal{C}_t = [0.2T, T]$, in which the cointegration was temporally switched off at the beginning of the sample. The full-sample based test constantly reports no-cointegration. This reflects the fact that cointegration tests often fail to find cointegration when the data is contaminated by short-lived extreme event. Our tests show a different picture. Their ability to detect cointegration is not much affected by the extreme events. And their power increases consistently when the error correction becomes more effective. Interestingly, the power of WQ and SQ is comparable to that of $SupQ^1$, which suggests that there is not much power loss by not specifying the number of breaks.

Figure 2 is concerned with the case $\mathcal{C}_t = [0.2T, 0.8T]$, in which the cointegration only exists in the middle segment of the sample. As expected, the power of $SupQ^1$ is very low. This shows the unattractiveness of always testing against a one-break alternative. Our WQ and SQ still show desired power property, with WQ dominating SQ in this case. Overall, the power of the tests is considerably lower than that in figure 1, which is partly due to the fact that now the cointegrated regime is only 60% of the sample.

To see the performance of the tests when the series are cointegrated under the null, we repeat the above simulation using the VARMA(1,1) model as in Yap (1995). The results are shown in Figure 3, where we have chosen $\lambda_\theta = 0$ and $(\lambda_1, \lambda_2, \lambda_3) = (1, \rho, 0)$, with ρ varies between 0.95 and 0. Again, the rejection frequencies based on 5% level are plotted, as a function of $1 - \rho$. Similar results are observed except that, when ρ is close to 1, the power of the tests is lower than that of full sample based trace test. Actually, this comes with no surprise, which reflects the fact that when the structural change is small, the property of the system is better captured by tests ignoring the break. The relatively low power against persistent alternatives appears to be the price we pay for the one more dimension of uncertainty incorporated in our tests.

To conclude, commonly applied cointegration tests fail to be informative when the sample of interest is contaminated by outlying events or subjects to regime shifts, while our tests are able to capture the inhomogenity in the data and report a rejection of the null. In practice, WQ or SQ should be used when not knowing the number of changes.

5.2.2 Change in the cointegrating vector

In the next we examine the situation where the cointegrating rank is maintained while the cointegrating vector changes. Our DGP is the same as above except that, there is a one time shift in the cointegrating vector happening at the middle of sample. More precisely, the DGP is,[3]

$$\begin{aligned} Y_{1t} &= Y_{1t-1} + v_{1t} \\ Y_{2t} &= \alpha + \beta_1 Y_{1t} + \varepsilon_t \; with \; \varepsilon_t = \rho\varepsilon_{t-1} + v_{2t}, \; if \; t \in [1, 0.5T] \\ Y_{2t} &= \alpha + \beta_2 Y_{1t} + \varepsilon_t \; with \; \varepsilon_t = \rho\varepsilon_{t-1} + v_{2t}, \; if \; t \in [0.5T + 1, T] \end{aligned}$$

In the above, $\beta_2 \neq \beta_1$, hence there is a change in the cointegrating vector. We consider two parameterizations: $(\alpha, \beta_1, \beta_2) = (1, 1_{n-1}, 1.5)$ and $(\alpha, \beta_1, \beta_2) = (1, 1_{n-1}, 2)$, which stand for different magnitudes of change (1_{n-1} denotes the n-1 demensional row vector of ones). To have some feeling about the magnitude, we plot a typical realization of the process in Figure 4. It is seen that the structural change is hardly visually detectable even for the large shift case.

The power of the tests under the first parametrization is plotted in Figure 5. Our tests shows desirable power, which is able to reject 80% when ρ is about 0.5. Interestingly, the power of the trace test is bell shaped in the case with n=2. This is due to the fact that when the persistence of the series decreases, the structural change becomes statistically more

[3]The series Y_{2t} is adjusted to avoid the spurious mean shift at $t = [0.5T]$

important, hence having a larger effect on the power of the test. Figure 6, which considers a larger shift, shows a similar picture, with the difference between our tests and trace test being more pronounced.

In a previous draft, we also examined the tests in models with mean shifts. The overall result is the same, hence omitted.

6 Empirical applications

In this section, we provide two simple applications. The first is pertaining to the test of the Expectation Hypothesis (EH) of the term structure of interest rates and the second is a study of the linkage between some European interest rates.

6.1 The term structure of US interests rates

The Expectation Hypothesis (EH) states that the long-term interest rate equals the market's expectation for the short-term rate plus a time-invariant term premium, which implies cointegration between long and short term interest rates if they are individually I(1). The EH is often examined by testing cointegration between interest rates of different maturities, with the empirical evidence far from being conclusive. One conjecture is that the term premium may have shifted due to regime changes. For example, Hansen (2003) use a ECM model allowing two regime changes and finds stronger support for EH. However, he takes the structural changes as exogenous. It is of interest to see whether his finding still holds if we treat the structural changes endogenously.

To this end, we examine the term structure using $US\ LIBOR$ data. We choose $LIBOR$ because we expect it to have a simple structure, which allows us to focus on the effect of regime change. The data is monthly taken from IFS data base and and ranges from $1978:1$ to $1997:6$. The rates including overnight (ON), 1-month($1M$), 3-month($3M$) and 1-year($1Y$) rates, which are plotted in figure 7.

Applying Johansen's (1991) tract test using the full sample shows mixed evidence toward EH. The test is constructed with a restricted constant, the lag selected by AIC[4], with the maximal lag length set to $int(12 * (T/100)\hat{}(1/4))$, which equals to 15 for this particular sample. The test is applied to all possible pairs and the results are reported in table 10,

[4] Note that the test statistic is $(T-(p+1)*2-1)\sum_{i=1}^{2}\ln(1-\lambda_i)$, where T is the effective sample size, p is the number of lags, and the $\lambda_i' s$ are the corresponding eigenvalues. The test is adjusted by the number of regressors to avoid a spurious rejection due to the small sample and large lag length.

where Y denotes a rejection at the the specified significance level and N denotes a non rejection. The number in the bracket denotes the lag length selected.

Table 10. Full sample based test for the null of no cointegration

	Pairs					
Significance level	ON-1M	ON-3M	ON-1Y	1M-3M	1M-1Y	3M-1Y
0.01	N (10)	N (10)	N (9)	N (9)	N (9)	N (9)
0.05	N	N	N	Y	N	N
0.10	Y	N	Y	Y	N	N

The null is never rejected at the 1% significance level. Even at the 10%, the null is only rejected in 3 relationships among 6. So the evidence in favour of EH is, at most, weak. To see whether regime change could explain the non-rejections, the SQ test is applied to all pairs. The result is presented at table 11, with the number in the bracket standing for the significance level at which the null is rejected:

Table 11. Test of no-cointegration against segmented cointegration

Pairs	ON-1M	ON-3M	ON-1Y	1M-3M	1M-1Y	3M-1Y
SQ	Y(1%)	Y(1%)	Y(5%)	Y(1%)	Y(5%)	Y(10%)

The null of no-cointegration is rejected in all cases, at least at 10% level. Hence the data appears to be segmentally cointegrated.

To see this more clearly, we estimate the change point using the method proposed by BLS (1998). The model is an unrestricted error correction model with four dependent variables, allowing a one time structural change in the coefficients and in the variance covariance matrix. The date of the break that maximizes the likelihood is 1982:12, and varying the lag length between 1 and 4 does not affect the result. Now, applying Johansen's test using the data after 1982:12, the null hypothesis of no cointegration is rejected at 1% level for all pairs. Hence we see strong evidence in favour of EH.

Some comments are in order. First, our estimated change point is close to the major monetary policy regime change in 1982. The period between September 1979 and October 1982 is known as the period with a *non-borrowed reserves operating procedure*, during which the Fed was not directly targeting the interest rate. The Fed switched to *interest rate targeting* in fall 1982. Secondly, the test of EH tells quite different stories depending on whether we account for the regime change. This highlights the importance of accounting for structural changes when testing for such hypotheses. Finally, Our result provides support for Hansen (2003) and other related research by showing the relevance of using the year 1982 as the boundary of two regimes.

6.2 Linkage between European Interest Rates

European countries have gone a long way toward an integrated market. One way to look at the degree of financial markets integration is through the so called *covered interest rate parity (CIP)*. CIP states that in an efficient market, currencies with high interest rates are expected to depreciate, and currencies with low interest rates are expected to appreciate. This tendency will continue until there is no room for arbitrage. More specifically, we would have

$$r_{dt} = r_{ft} + f_t - s_t$$

where r_{dt} (r_{ft}) denotes the yield on a domestic (foreign) asset, f_t is the forward exchange rate and s_t is the spot rate (all expressed in logarithm form). The forward premium $f_t - s_t$ can be decomposed into a risk premium (RP) plus expected change in the exchange rates ($E(\Delta s)$), so that

$$r_{dt} = r_{ft} + RP - E(\Delta s)$$

Then if the expected change in the exchange rate is stationary and the risk premium is time-invariant, there would be a close co-movement between the domestic and foreign interest rates. If we believe these are I(1), then they should be cointegrated. For European interests data, cointegration is only found to a limited degree. Various reasons have been offered to explain the failure to find cointegration, some of which are: 1) Market are segmented and regulations such as capital controls prevent arbitrage; 2) Costs have been unaccounted for, such as differences in political risks across countries. 3) Either the change in exchange rates or risk premium is non-stationary, which may be caused by regime changes or outlying events. To us, the third explanation is most intuitively appealing. To see whether this is the case, we use the tools developed above to search for segmented cointegration in the sample.

The data consists of $6-$month $LIBOR$ rates taken from IFS. We shall not include France because its interest rate has been subjected to constant attack in the early 80s, which would complicate the analysis. Hence, we are left with the data of Germany (GM), Netherlands (NL), Switzerland (CH) and United Kingdom (UK). Note that Switzerland is not a member of EU, but we still include it in the analysis. The data covers the same period as before, 1978 : 1 to 1997 : 6, which is plotted in the figure 8.

To make our analysis more informative, we pursue the following strategy. First, we analyze the property of a tri-variate system, which includes GM, NL and CH. Then, we use UK series as extrapolation. This helps us to alleviate the small sample problem, as well as to show the generality of our result. We start by applying Johansen's test to the tri-variate system, using the full sample. The lag length is chosen by AIC, which is equal to 5. The value of the statistic is 31.54, while the the 10% critical value is 32.00. In contrast, the SQ and WQ test reject the null at 10% level, which shows evidence toward segmented cointegration. Also, a closer look reveals that $SupQ^2$ rejects at 5% but $SupQ^1$ does not reject even at 10% level, which altogether suggests the existence of multiple regime changes. We proceed to estimate the model allowing a maximum of two changes with the lag length being 5 as suggested by AIC. The result is as follows: a) If allowing only one break, its estimate is 1982 : 7[5]; b) If allowing two breaks, the first break is at 1982 : 6 and the second is at 1987 : 12[6].

Now, we apply Johansen's trace test on all possible pairs, using the partitioned sample. The results are presented in Table 12.

Table 12. Test for the null of no cointegration based on the partitioned sample

	Regimes		
Pairs	1978:1-1982:6	1982:7-1987:12	1988:1-1997:6
GM-NL	N	Y(1%)	Y(1%)
GM-CH	N	Y(10%)	N
NL-CH	N	Y(1%)	N

Three facts emerge from table 12. First, there appears to be three different regimes, with different degrees of cointegration. Second, within the second regime all series appears to be stationary, with unit root rejected at least at 10% level. This is a feature that does not hold for the other segments. Third, for the third regime, the connection between Germany and the Netherlands seems to remain but is weakened for pairs involving Switzerland

The above conclusion are strengthened if we include UK in the analysis. Taking the estimated breaks as given, for UK, the series appears to be stationary in the second regime, with the unit root rejected at 10% level. In the third regime, it is cointegrated with Germany

[5] When the lag length varies between 1 and 5, the estimated break varies between 1982:7 and 1983:6.

[6] When the lag length varies between 1 and 4, the first break is 1982:7 and the second does not change.

and also Netherlands with no cointegration rejected at 1% level. This altogether suggests a regime specific co-movement.

Interestingly, our classification of regimes accords with major historical events, as stated in Zhou (2003). Although established in 1979, the EMS did not prompt significant monetary discipline until the early 1980s. In our sample, this manifests itself as no cointegration in the first regime. There was a interest exchange rate realignment in March 1983, after which the EMS entered a period characterized by enhanced exchange rate stability. This corresponds to our second regime, in which all series move closely together. For the third regime, the interest rates of between EMS members still move together, but the links with Switzerland weaken. Similar observations are made in Zhou (2003), among others.

Finally, notice that the regimes identified by our procedure are very much in line with the regimes analyzed by other authors, who chose boundaries of regimes based on the knowledge of economic history or subjective methods. For example, Centeno and Mello (1999) study the linkage between European interest rates during the 1985-1994 period. They use 1990:10 as the date of regime change, which is close to our second change point. Zhou (2003) uses daily data and a longer span, 1979-1998. She partitions the sample into three regimes: 1979:3-1983:3, 1983:4-1990:12 and 1991:1-1998:7. Again these are very much in line with our results.

7 Conclusion

In this paper, we propose a procedure to search for regime-sensitive cointegration and to determine the location of the changes endogenously. Among our conclusions, two important ones are the importance of testing against multiple-break alternatives and the relevance of accounting for structural instability when analyzing cointegration. Our analysis could be extended in various directions. One is to extend the procedure to trending series. Another is to apply the data correction procedure to other non-parametric tests and to study whether a data-dependent lag length could further improve the performance.

References

[1] Bai, J., Lumsdaine, R.L., and Stock, J.H. (1998), "Testing for and Dating Common Breaks in Multivariate Time Series" *Review of Economic Studies,* **65**, 395-432.

[2] Bai, J. and Perron, P. (1998), "Estimating and Testing Linear Models with Multiple Structural Changes", *Econometrica* **66**, 47-78.

[3] Bai, J., and Perron, P. (2003), "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics, **18**, 1-22.*

[4] Breitung, J. (2002), "Nonparametric tests for unit roots and cointegration", *Journal of Econometrics*, 108, 343-363.

[5] Centeno, M., and Mello, A.S. (1999), "How integrated are the money market and the bank loans market within the European Union", *Journal of International Money and Finance*, 18, 75-106.

[6] Hansen, P.H. (2003), "Structural Changes in the Cointegrated Vector Autoregressive Model", *Journal of Econometrics*, 114, 261-295.

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

[8] Johansen, S. (1995), "Likelihood-based Inference in Cointegrated Vector Autoregressive Models", *Oxford University Press.*

[9] Kim, J.Y. (2003), "Inference on Segmented Cointegration" , *Econometric Theory*, 19, 620-639.

[10] Kleimeier, S., and Sander, H. (2000), "regionalization versus Globalization in European Financial Market Integration: Evidence from Co-integration Analysis", *Journal of Banking and Finance*, 24, 1005-1043.

[11] Kwiatkowski, 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, 159-178.

[12] Leybourne, S., Kim, T.H., Smith,V., and Newbold, P. (2003), "Tests for a Change in Persistence against the Null of Difference-stationarity", *Econometrics Journal*, 6, 290-310.

[13] Nyblom, J. (1989), "Testing for the constancy of parameters over time", *Journal of the American Statistical Association*, 84, 223-230.

[14] Nyblom, J., and Harvey, A. (2000), "Tests of Common Stochastic Trends", *Econometric Theory*, 16, 176-199.

[15] Qu, Z., and Perron, P. (2004), "Estimating and testing structural changes in multivariate regressions", Unpublished Manuscript, *Economics Department, Boston University*.

[16] Siklos, P.L., and Granger, C.W.J. (1997), "Regime-sensitive Cointegration with an Application to Interest-rate Parity", *Macroeconomic Dynamics*, 1,640-657.

[17] Zhou, S. (2003), "Interest Rate Linkages within the European Monetary System: New Evidence Incorporating Long-run Trends", *Journal of International Money and Finance*, 22, 571-590.

8 Appendix

Proof of Proposition 1: For $\lambda_1 \in [\varepsilon, 1]$, we define

$$\bar{A}_{\lambda_1} = \frac{1}{T} \sum_{t=1}^{[T\lambda_1]} \hat{U}_t \hat{U}_t' \ , \ \ \bar{B}_{\lambda_1} = \frac{1}{T^2} \sum_{t=1}^{[T\lambda_1]} (\sum_{j=1}^{t} \hat{U}_j)(\sum_{j=1}^{t} \hat{U}_j)'$$

also,

$$\bar{Q}_{\lambda_1} = \left(\bar{B}_{\lambda_1}\right)^{-1/2} \bar{A}_{\lambda_1} \left(\bar{B}_{\lambda_1}\right)^{-1/2}$$

To prove the proposition, it is sufficient to show that for any admissible partition $\Lambda = (\lambda_1, ...\lambda_m)$ with $\lambda_i = T_i/T$, $i = 1, ..., m$, we have,

$$\sum_{i=1}^{n-r_0} \rho_i(T\bar{Q}_{\lambda_1}) \Rightarrow Tr \left(\lambda_1^2 \int_0^{\lambda_1} W_\lambda(s) W_\lambda(s)' ds \left[\int_0^{\lambda_1} V_\lambda(s) V_\lambda(s)' 1 ds \right]^{-1} \right) \tag{3}$$

Then the proof is completed applying continuous mapping theorem, using the fact that the $Sup(.)$ is a continuous function over a compact set.

To show (3), we need to solve the following eigenvalue problem

$$\left|\rho \bar{B}_{\lambda_1} - \bar{A}_{\lambda_1}\right| = 0$$

for $\lambda_1 \in [\varepsilon, 1]$. First, notice that the problem is rotation invariant. So, without lose of generality, we can assume $Y_t = (Y_{1t}, Y_{2t})$ with Y_{1t} being a r_0 dimensional transitory component and Y_{2t} being $n - r_0$ dimensional purely nonstationary component. Then, we could write

$$\bar{A}_{\lambda_1} = \begin{bmatrix} \bar{A}_{\lambda_1}^{11} & \bar{A}_{\lambda_1}^{12} \\ \bar{A}_{\lambda_1}^{21} & \bar{A}_{\lambda_1}^{22} \end{bmatrix}$$

and

$$\bar{B}_{\lambda_1} = \begin{bmatrix} \bar{B}_{\lambda_1}^{11} & \bar{B}_{\lambda_1}^{12} \\ \bar{B}_{\lambda_1}^{21} & \bar{B}_{\lambda_1}^{22} \end{bmatrix}$$

with the partition done accordingly. Second, define $D_T = \begin{pmatrix} I_{r_0} & 0 \\ 0 & T^{-1/2} I_{r_0} \end{pmatrix} \equiv (\beta, \beta_\perp)$, which serves as a normalization of the moment matrices.

The eigenvalues of $\left|\rho \bar{B}_{\lambda_1} - A_{\lambda_1}\right| = 0$ are the same as those of

$$\left|\rho D_T' \bar{B}_{\lambda_1} D_T - D_T' \bar{A}_{\lambda_1} D_T\right| = 0$$

The above problem has r_0 eigenvalues of order $O_p(1)$, with eigenvectors lying in the space spanned by the cointegrating vectors. It has other $n - r_0$ eigenvalues of order $O_p(T^{-1})$, which correspond to the eigenvectors lying in the space orthogonal to that spanned by the cointegrating vectors. We need to find the asymptotic distribution of the latter $n - r_0$ eigenvectors.

To this end, we follow the approach used in Johansen (1995, p.159). Define $S_{\lambda_1}(\rho) = \rho\bar{B}_{\lambda_1} - \bar{A}_{\lambda_1}$, we consider the following decomposition:

$$\begin{aligned}\left|D_T^{'}S_{\lambda_1}(\rho)D_T\right| &= \left|\begin{bmatrix}\beta^{'}S_{\lambda_1}(\rho)\beta & \beta'S_{\lambda_1}(\rho)\beta_\perp \\ \beta_\perp^{'}S_{\lambda_1}(\rho)\beta & \beta_\perp^{'}S_{\lambda_1}(\rho)\beta_\perp\end{bmatrix}\right| \\ &= \left|\beta^{'}S_{\lambda_1}(\rho)\beta\right|\left|\beta_\perp^{'}\left\{S_{\lambda_1}(\rho) - S_{\lambda_1}(\rho)\beta\left[\beta^{'}S_{\lambda_1}(\rho)\beta\right]^{-1}\beta'S_{\lambda_1}(\rho)\right\}\beta_\perp\right|\end{aligned}$$

Let $T \to \infty$, and $\rho \to 0$, such that $\eta = T\rho$ is fixed, then,

- $\beta^{'}S_{\lambda_1}(\rho)\beta = T^{-1}\eta\bar{B}_{\lambda_1}^{11} - \bar{A}_{\lambda_1}^{11} = -p\lim\bar{A}_{\lambda_1}^{11} + o_p(1)$, with $p\lim\bar{A}_{\lambda_1}^{11}$ being a non-random positive definite matrix. Hence $\left|\beta^{'}S(\rho)\beta\right|$ has no root when $T \to \infty$.
- $\beta_\perp^{'}S_{\lambda_1}(\rho)\beta = T^{-1/2}\left(\rho\bar{B}_{\lambda_1}^{21} - \bar{A}_{\lambda_1}^{21}\right)\beta = T^{-3/2}\eta\bar{B}_{\lambda_1}^{21} - T^{-1/2}\bar{A}_{\lambda_1}^{21}\beta = O_p\left(T^{-1/2}\right) = o_p(1)$.
- $\beta_\perp^{'}S_{\lambda_1}(\rho)\beta_\perp = T^{-1}\left(\rho\bar{B}_{\lambda_1}^{22} - \bar{A}_{\lambda_1}^{22}\right) = T^{-2}\eta\bar{B}_{\lambda_1}^{22} - T^{-1}\bar{A}_{\lambda_1}^{22} = O_p(1)$

Hence we have,

$$\beta_\perp^{'}\left\{S_{\lambda_1}(\rho) - S_{\lambda_1}(\rho)\beta\left[\beta^{'}S_{\lambda_1}(\rho)\beta\right]^{-1}\beta'S_{\lambda_1}(\rho)\right\}\beta_\perp = T^{-2}\eta\bar{B}_{\lambda_1}^{22} - T^{-1}\bar{A}_{\lambda_1}^{22} + o_p(1)$$

Notice that

$$\left|T^{-2}\eta\bar{B}_{\lambda_1}^{22} - T^{-1}\bar{A}_{\lambda_1}^{22}\right| = 0$$

has $n - r_0$ positive eigenvalues $\eta_{r_0+1}(\lambda_1), ..., \eta_n(\lambda_1)$, and we have

$$\sum_{i=r_0+1}^{n}\eta_i(\lambda_1) = Tr\left(\left(T^{-2}\bar{B}_{\lambda_1}^{22}\right)^{-1}T^{-1}\bar{A}_{\lambda_1}^{22}\right) + o_p(1)$$

Hence

$$\sum_{i=r_0+1}^{n}\eta_i(\lambda_1) \Rightarrow Tr\left(\lambda_1^2\int_0^{\lambda_1}W_{\lambda_1}(s)W_{\lambda_1}(s)'ds\left[\int_0^{\lambda_1}V_{\lambda_1}(s)V_{\lambda_1}(s)'1ds\right]^{-1}\right)$$

for $\lambda_1 \in [\varepsilon, 1]$. This completes the proof.

Proof of Proposition 2: Define $\hat{A}(L) = \hat{A}(1)^{-1}\hat{A}^*(L)$, which is a finite order polynomial of lag operators. Notice that since the first difference series are stationary, $\hat{A}(L)$ has no root on the unit circle, at least asymptotically. Hence $\left|\hat{A}(1)\right| \neq 0$ and $\hat{A}(1)^{-1}$ exists.

We have

$$H'\left(\widetilde{Y}_t-\alpha\right)=\begin{bmatrix}\beta'\left(Y_t-\alpha\right)\\ \gamma'\left(Y_t-\alpha\right)\end{bmatrix}+\begin{bmatrix}\beta'\hat{A}(L)\Delta Y_t\\ \gamma'\hat{A}(L)\Delta Y_t\end{bmatrix}$$

$T^{-1/2}\gamma'\hat{A}(L)\Delta Y_t \to^p 0$ follows from assumption 1. We also have

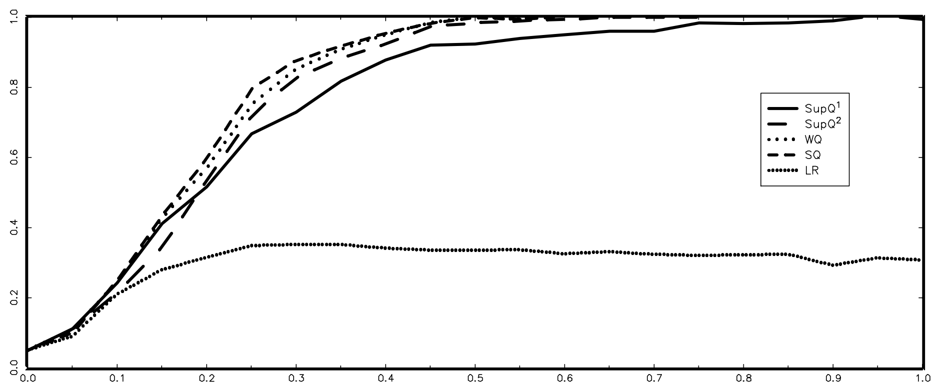
$$T^{-2}\sum_{t=1}^{[Ts]}\beta'\left(\hat{A}(L)\Delta Y_t\right)\left(\hat{A}(L)\Delta Y_t\right)'\beta\to 0$$

uniformly in s. This completes the proof.

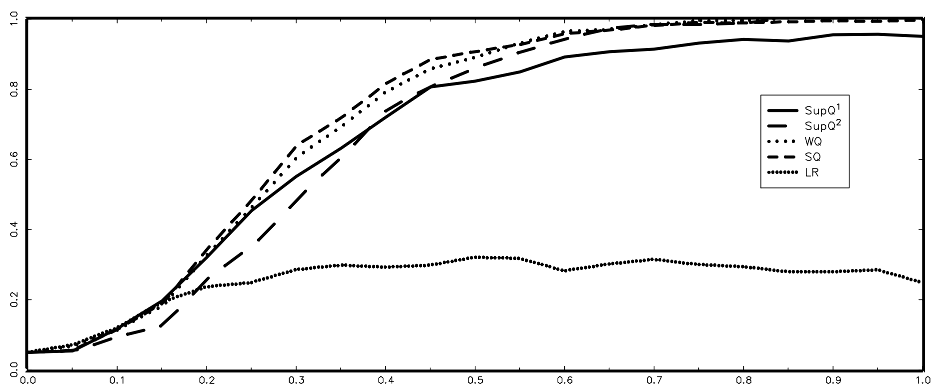
Figure 1: The power of the tests with a single change in the cointegrating rank

$(T=200,\ \varepsilon=0.2,\ C_t=[0.2T,T])$

a) n=1



b) n=2



c) n=3

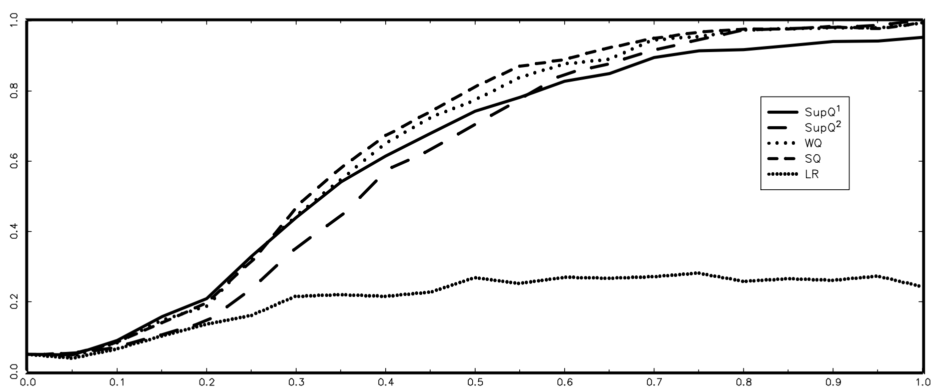
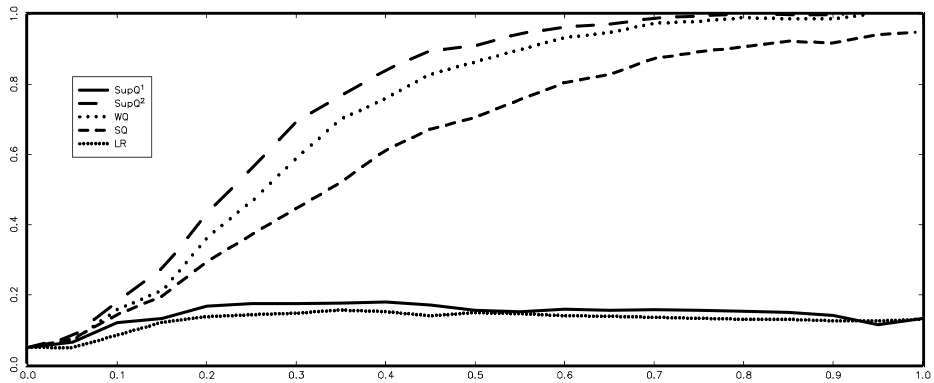


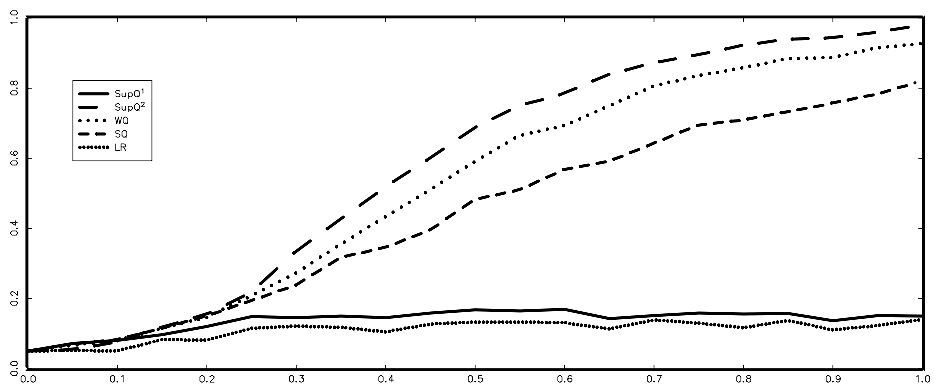
Figure 2: The power of the tests with multiple changes in the cointegrating rank

($T=200$, $\varepsilon=0.2$, $C_t=[0.2T,0.8T]$)

a) n=1



b) n=2



c) n=3

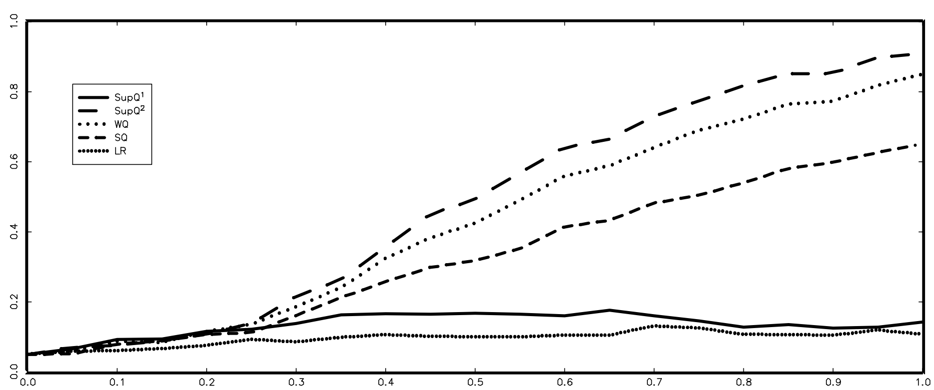
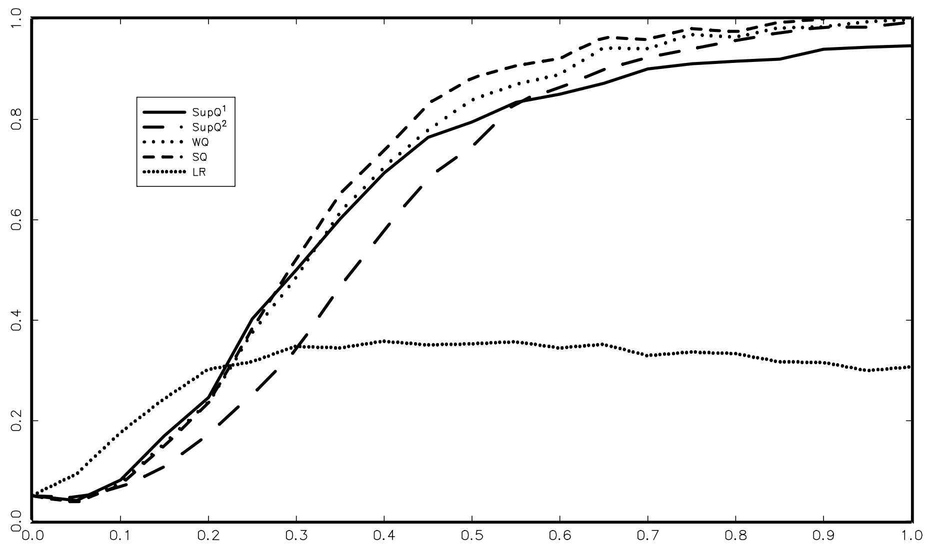


Figure 3: The power of the tests in the tri-variate VARMA(1,1) model

a) T=200, ε=0.2, C_t=[0.2T,T]



b) T=200, ε=0.2, C_t=[0.2T,0.8T]

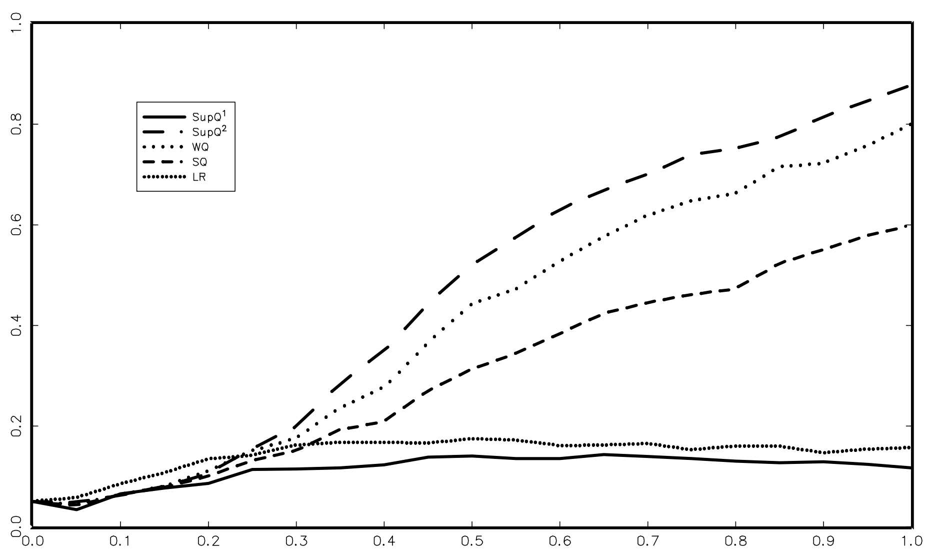


Figure 4: Change in the cointegrating vector: two typical realizations

Figure 5: The power of the tests with a single change in the cointegrating vector

(T=200, $\varepsilon=0.2$, $\beta_2-\beta_1=0.5$)

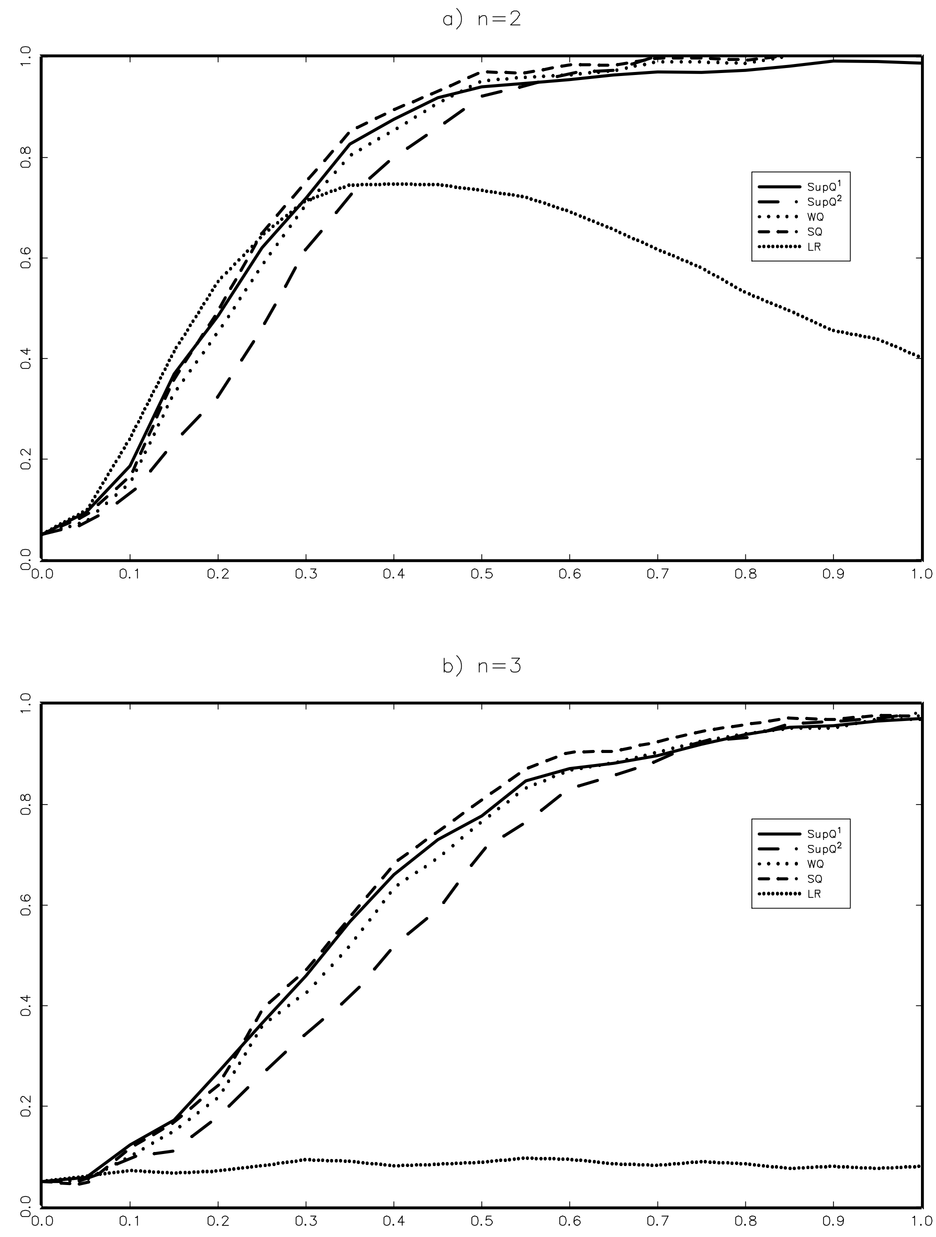


Figure 6: The power of the tests with a single change in the cointegrating vector

(T=200, $\varepsilon=0.2$, $\beta_2-\beta_1=1$)

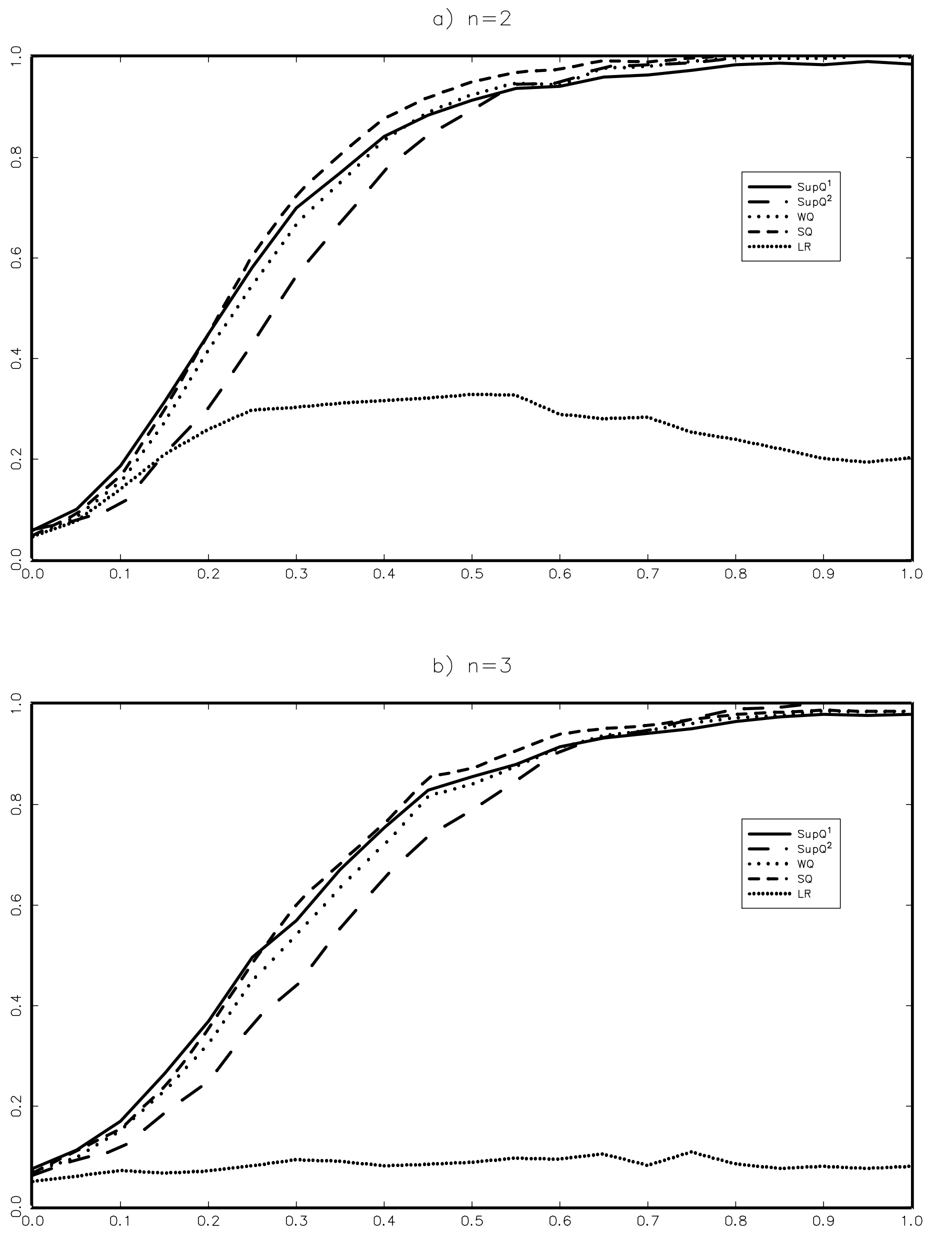


Figure 7: US dollar LIBOR rates of different maturities

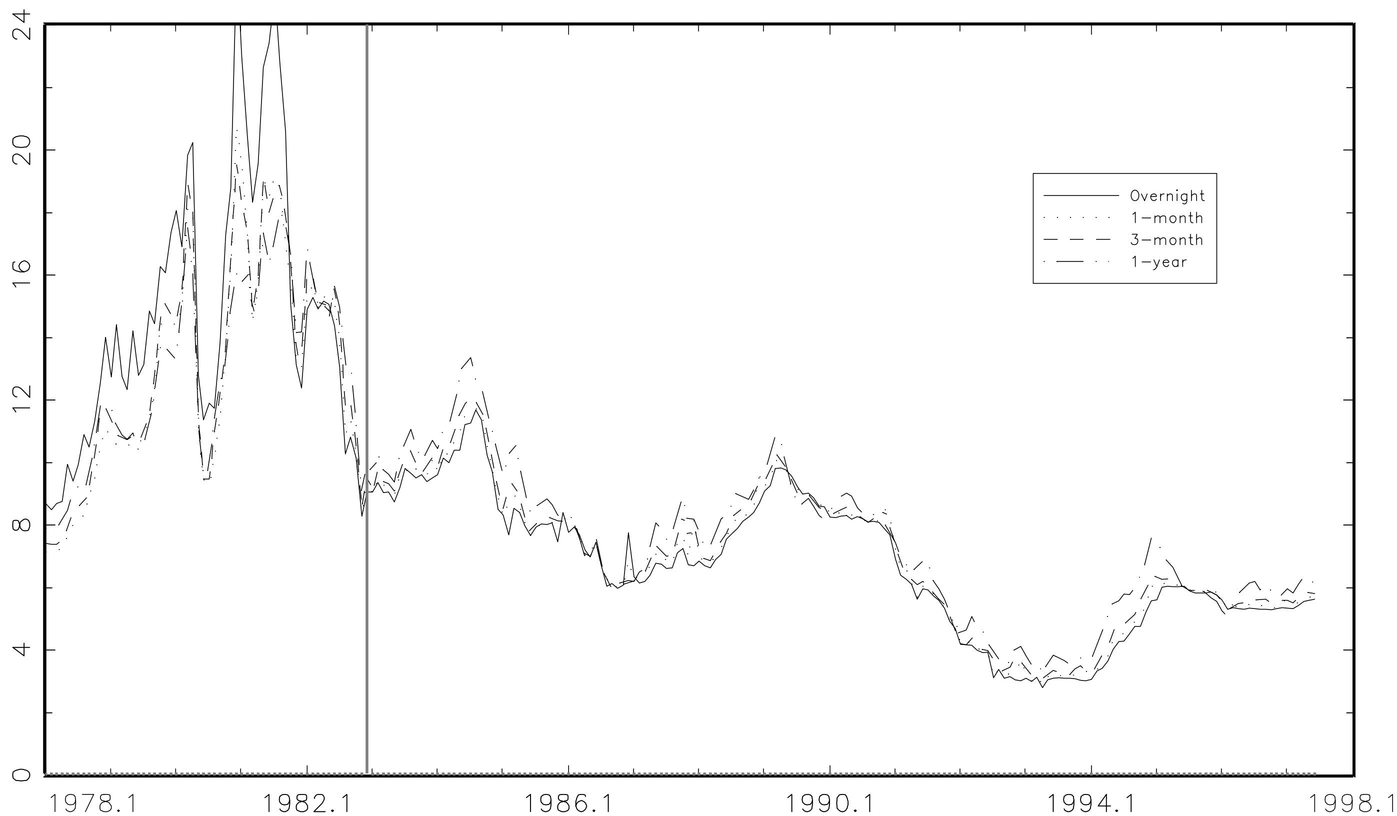


Figure 8: The 6-month LIBOR rate of some European Countries