J Econ Finan (2012) 36:463–480
DOI 10.1007/s12197-010-9136-9

On the cointegration of international stock indices

Richard Fu · Marco Pagani

Published online: 30 April 2010

Abstract This paper investigates the cointegration relationship among a group of international stock indices in light of new developments of econometric methods. Kasa (1992) first documented strong evidence for cointegration relations among five national stock indices, which suggests that there exists a common trend among those stock indices. Using Johansen multivariate cointegration test, we find that his findings are persistent in a sample of longer periods and more countries. In order to investigate whether these results are driven by statistical biases related to the sample size, we apply to our tests the Johansen's small sample correction factor. The results still point toward the existence of a cointegration relationship but the evidence becomes much weaker. We next examine the empirical patterns emerged from different lag specifications and argue that Kasa's findings are more likely due to the size distortion in extreme long lag VAR models. Indeed, when we employ a newly developed non-parametric test that does not require estimation VAR models, the null hypothesis of no cointegration cannot be rejected for the original sample of Kasa's five-country stock indices from 1974 to 1990, nor for the extended period from 1970 to 2003.

Keywords Cointegration · Market Efficiency · Johansen's Test · Non-parametric Cointegration Test

JEL Classification G14 · G15

R. Fu (✉)
Department of Accounting and Finance, College of Business, San José State University,
One Washington Square, San Jose, CA 95192-0066, USA
e-mail: fu_r@cob.sjsu.edu

M. Pagani
Department of Accounting and Finance, College of Business, San José State University,
One Washington Square, San Jose, CA 95192-0066, USA

1 Introduction

Since Engle and Granger's (1987) seminal paper, the economic literature on cointegration has grown exponentially. Cointegration tests and error correction models are widely used in the finance literature to examine relationships among asset returns. It is Engle and Granger's (1987) paper that first formalized the concept of cointegration and developed the bivariate cointegration test. However, even if there is no pairwise cointegration, one cannot conclude that there is no cointegration among the multiple time series. Johansen (1988)'s methodology based on multivariate vector autoregressive (VAR) model has provided researcher necessary tools for multivariate cointegration tests and error corrections.

An important application of cointegration tests in finance area is to test market efficiency by examining the long-run relationship among time series of prices of different assets. For example, if the prices in two markets are cointegrated then it would be possible to forecast one from the other, which suggests that markets are either not efficient, or individual markets are efficient but segmented, or market indices are driven by time-varying risk factors. Kasa (1992) investigated national stock indices of five developed countries from 1974 to 1990. He used Johansen's (1991) likelihood ratio test to examine the cointegration relationship among the five national stock indices. He reported a pronounced cointegration relationship among five national stock indices comprising the US, UK, Japanese, Germany and Canadian markets. In particular, his results demonstrated much stronger evidence in favor of cointegration with long-lag models than for short ones. Finally, Kasa showed that in long-lag specifications, there are four cointegration vectors among the five indices and argued that there was one common stochastic trend exists among the five markets.

Kasa's findings are quite puzzling for two reasons. First, the existence of a cointegration relationship runs counter to the efficient market theory. It implies that in the long run, the stock indices among different countries are closely linked, even though they may temporarily diverge in the short run. As shown in the literature, researchers need only examine past returns of stock indices, which are completely public knowledge, to detect potential cointegration relationships. The existence of the cointegration among these five country stock indices violates the weak form market efficiency unless it can be explained by time varying expected return or complete market segmentation.[1] Second, if there is one underlying common stochastic trend among major countries' stock indices, as argued by Kasa, the benefits of international diversification would be greatly diminished. On the contrary, diversification benefits from international investments are well documented in empirical studies over past decades.[2]

[1] Since the 1980's, most developed countries have opened their capital markets to foreign fund investors. Furthermore, the liberation of capital markets has also occurred in several emerging markets in the 1990's. Nevertheless, we show that Kasa's findings are persistent even with 1990's data included, Hence, the market segmentation is unlikely to be a significant factor explaining the existence of cointegration among national stock indices.

[2] See, for example, Grubel (1968); French and Poterba (1991); Eun and Resnick (1994).

Richards (1995) examined the statistical basis for the rejection of the null hypothesis of no cointegration among five national stock markets in Kasa's (1992) paper. He concluded that the finding of cointegration relationship is due to the failure to adjust asymptotic critical values to take into account the small number of degrees of freedom embedded in the Johansen's methodology, especially in a long lag VAR model. He then used two other methodologies similar to those of Fama and French (1988) and DeBondt and Thaler (1985) to investigate whether national stock market returns demonstrate weaker form of long-horizon relative predictability. He found that the national stock indices include a common world component as well as two country-specific components and argued that there are comovements among national stock market returns but no cointegration.[3]

In this paper, we first show that Kasa's results are persistent if we use the same methodology in a sample for an extended period from 1970 up to 2003, as well as during different sub-periods. Moreover, using Kasa's methodology, we show the existence of cointegration even among emerging markets, though somewhat weaker than the cointegration among developed countries. Yet as in Kasa (1992), the findings of cointegration largely depend on using long-lag VAR model specification in the Johansen's test.

Next, we investigate whether the findings of cointegration are driven by size distortions of the Johansen's test, as documented in previous econometric simulation studies. We employ the Johansen 's (2002) small sample correction for the likelihood ratio tests. After the correction for small sample, we find that the cointegration in our sample still holds, yet it is much weaker than without correction. Moreover, the findings of cointegration critically depend on long-lag VAR models. In fact, after taking sufficiently long lag model, the null hypothesis of no cointegration is always rejected, as in the tests without small sample correction. These results suggest that Richards's (1995) argument that Kasa's results are due to small sample sizes is not likely to be a sufficient explanation of Kasa's apparent findings of cointegration.

In the econometric literature, various simulation studies document that the performance of the Johansen's test is sensitive to the misspecification of VAR model, which may lead to over-rejection of the null hypothesis of no cointegration.[4] Both Kasa's study and ours using Johansens' test show that long-lag VAR model tend to produce stronger results of cointegration, which might be a spurious finding due to the likelihood of the misspecification of the VAR model. In order to avoid this problem, we implement a new non-parametric cointegration rank test developed by Shintani (2001) to test the cointegration relations in our sample. The distinctive feature of Shintani's test is that it does not need to pre-specify the lag length as the VAR model requires. Using this non-parametric test enables us to avoid the likely size distortion of the Johansen's test due to the VAR misspecifcation, which results in over-rejection of no-cointegration null. The empirical results from Shintani's non-

[3] Recent papers focus on comovements among international stock returns as, for example, in Bekaert et al. (2009) These studies essentially examine *correlations* among returns across countries and industries, which is conceptually different from *cointegration* relationship investigated in this paper.

[4] See, for example, Boswijk and Franses (1992); Bewley and Yang (1995); Richards (1995); Ho and Sorensen (1996); Haug (1996).

parametric test show that we cannot reject the null hypothesis of no cointegration at conventional significance level. In conclusion, our results are consistent with the argument that Kasa's findings are likely due to statistical misspecifications caused by size distortions in the VAR type cointegration tests. Kasa's conclusion of pronounced cointegration relationship based upon long-lag VAR model in the Johansen's test are not supported by the results of non-parametric test, which is not subject to the size distortion due to the VAR misspecification.

The rest of the paper is organized as follows: Section 2 reviews cointegration tests and previous empirical findings. Section 3 presents our empirical results of Johansen's original cointegration test on a expanded sample, Johansen's (2002) test of small sample correction, and Shintani's (2001) non-parametric cointegration rank test. Section 4 concludes the paper.

2 Literature review

2.1 Background on cointegration tests

Consider two time series integrated with order one, denoted I(1). By the definition of I(1), taking the first difference of I(1) series yields stationary I(0) time series. If there exists a linear combination of these two time series that is an I(0) series, then these two series are said to be cointegrated. I(1) series are widely observed in the economics and finance literature. For example, the random walk with drift is an I(1) process, which display a tendency to "wander", characterized by the innovation term that follows a normal distribution. However, when two I(1) series are cointegrated, they tend to "wander together", as if there exists some common trend corralling them together. More generally, if for a set of times series integrated of order d [henceforth denoted as I (d)], there exists a linear combination of the component variables such that it is of I(d–b) where $0<b\leq d$, then these time series are said to be of cointegrated of order (d,b).

The econometric literature has developed two broad approaches for parametric tests of cointegration. The Engle and Granger (1987) method is based on testing the stationarity of single-equation estimates of the equilibrium errors by a variant of Augmented Dickey-Fuller (ADF) test. The second approach, due to Johansen (1988, 1991) and Stock and Watson (1988), is based on the VAR approach. The idea is to detect the implied restrictions in an otherwise unrestricted VAR among a set of variables that is truly cointegrated.

In the finance literature, the processes of asset prices in logarithm are approximately random walks, or I(1) series. Hence, performing cointegration tests of I(1) variables is equivalent to examining whether there is a linear combination of these variables that is an I(0) series, among which the most common is the series of white noise. The existence of such linear combination means that these variables are cointegrated. Engle and Granger's test procedure regresses one variable y_t on the other x_t by using OLS and tests the corresponding residuals for nonstationarity using ADF tests. This procedure is easy to understand and implement. However, it can be only used to test pairwise cointegration and it is not applicable for detecting cointegration among multiple variables.

Johansen (1988) and Johansen and Juselius (1990) developed a maximum likelihood ratio test that allows testing for multiple cointegration vectors in a multivariate setting. The test is formulated in VAR model for multiple variables:

$$\Delta X_t = \mu + \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_k \Delta X_{t-k} + \varepsilon_t \quad (1)$$

where ε_t is assumed to be an i.i.d Gaussian process.

If the time series to be tested have unit roots, then their first difference terms are I (0) processes. In order to make the above equation established on both sides, Π must be of reduced rank. If Π is of full rank, it would be contradictory to the original specification of I(1) for the time series variables X_t. Π could be a null matrix, in which case the above equation becomes a standard VAR in first differences, and there are cointegration relations among those variables. If Π has a rank r between 0 and n, then Π can be factored as $\Pi=\alpha\beta'$, where α and β are n x r matrices. In order to make both sides of the equation to be I(0) variables, ΠX_{t-1} must be stationary, i.e. I(0). Then, the r columns of β are cointegration vectors. The following statistics are Johansen's trace test statistic and maximum eigenvalue statistics for null hypothesis of no cointegration H_0: $r=0$:

$$\begin{aligned} Trace &= -T\sum_{i=1}^{n}\ln\left(1-\widehat{\lambda}_i\right) \\ \lambda_{\max} &= -T\ln\left(1-\widehat{\lambda}_1\right) \\ S_{ij} &= T^{-1}\sum_{t=1}^{T} R_{it}R'_{jt}(i,j=0,1) \end{aligned} \quad (2)$$

where λ_i s are the eigenvalues of the matrix $S_{11}{}^{-1}S_{10}S_{00}{}^{-1}S_{01}$ in a decreasing sequence, and S_{ij}.are the covariance matrices of the residuals vectors.

In practice, the Johansen's test starts with the null hypothesis of cointegration rank $r=0$. If the null is rejected then test $r\leq 1$. The procedure goes on if null of smaller rank is rejected until $r\leq n-1$. The statistics must be adjusted to test different hypotheses. Generally, in testing $r\leq$q, trace statistic and maximum eigenvalue statistics are as follows:

$$\begin{aligned} Trace &= -T\sum_{i=q+1}^{n}\ln\left(1-\widehat{\lambda}_i\right) \\ \lambda_{\max} &= -T\ln\left(1-\widehat{\lambda}_{q+1}\right) \end{aligned} \quad (3)$$

The Johansen's test has gained wide popularity since its invention. However, later research on the methodology has revealed some of its flaws and limitations. Simulation studies show that the small sample properties of the trace test are different from the asymptotic properties (Cheung and Lai 1993; Gregory 1994; Toda 1995; Haug 1996; Ho and Sorensen 1996; Gonzalo and Pitarakis 1999). Moreover, Gonzalo and Lee (1998) show analytically and numerically a few important situations where the Johansen Likelihood Ratio test tends to find spurious cointegration with probability approaching one asymptotically. In summary, the results of recent econometric studies on the Johansen tests suggest that researchers need to take great care to avoid finding spurious cointegration relations.

The flaws of the Johansen tests lie in the fact that the procedure requires the assumption of a correctly specified finite order VAR process, as shown in equation (1). The aforementioned simulation studies show that the performance of the test is sensitive to VAR model misspecifications. One way to avoid the problem is to utilize statistical tests that do not require VAR models. Shintani (2001) developed a non-parametric cointegration rank test that does not required estimation of a VAR model. The test statistic is based on the inverse of the sample covariance matrix. The procedure has the advantage of simple computation and implementation. It uses the rank degeneracy in the sample covariance matrix rather than the degeneracy in the first order VAR coefficient estimates, and the sample moment is standardized with the nonparametric long-run variance estimation rather than lags in VAR models. Hence, it is a fully nonparametric cointegration test without the dependence on VAR model specification.

2.2 Cointegration and asset pricing

Granger (1986) first used cointegration methods to test the efficient market hypothesis. If the prices in two markets are cointegrated, then it implies the possibility to forecast one from the other. In an efficient market, prices ought not to be predictable based on previous prices. Thus the efficient market hypothesis implies the absence of cointegration. However, one should be cautious to use the cointegration tests to test the market efficiency. The existence of cointegration does not necessarily imply market inefficiency. This long-run relationship among asset prices might be explained by time-varying risk factors or market segmentation. In fact, one can consider the absence of cointegration as the necessary but not sufficient condition to market efficiency.

In the area of international financial market, the cointegration may exist due to market segmentation caused by the national borders. Individual markets might be efficient while national boundaries and regulations keep them from fully integrated. Second, long range predictability is not equal to the market inefficiency. The best-known instance is the predictability of dividend yield on long term stock return (Fama and French 1988, 1989; Conrad and Kaul 1988; Poterba and Summers 1988). The central theme of such debate is whether the long range predictability reflects time-varying risk factors. Therefore, the cointegration test may be used to test Market Efficiency Hypothesis, but researchers must be very careful to interpret the empirical results.

Kasa (1992) employed Johansen's methodology to test the cointegration relationship among five developed countries' national stock indices. Using price and dividend data from the Capital International indices constructed by Morgan Stanley (MSCI) for U.S, U.K, Canada, Japan and Germany from 1974 to 1990, the author showed that the null hypotheses of no cointegration are strongly rejected with a VAR specification of taking lag of 15 on monthly data. However, the results appear to be sensitive to the lag length in the VAR model. Cointegration relations are much weaker using VAR models than with shorter lags. Using quarterly data, Kasa detected a pronounced cointegration relation in a VAR model with a lag of 10 quarters. Furthermore, the tests reject the null hypothesis of the cointegration rank, $r \leq 3$ in favor of the hyphothesis that $r=4$. Thus, Kasa argues that there is a common stochastic trend underlying these five stock indices since the cointegration rank is four.

Kasa's findings of cointegration among five national stock indices are somewhat controversial. The strong evidence of cointegration relationship seem to largely hinge upon the lag length assumed in the VAR model: little evidence of cointegration in short lag models but strong evidence in long lag models. Kasa argues that long lag models reflect the long run cointegration relationship. Another justification for the long lag model is to make the error term more consistent with Gaussian/i.i.d. assumption which the Johansen's test relied upon. In the empirical literature, random walk proves to be a fairly satisfactory working model. However, the distortion introduced by extra lags may well offset the benefits of making the innovation closer to normality.[5]

Richards (1995) challenged the findings of cointegration among national stock indices examining the statistical basis of Kasa's (1992) empirical methods. He conducted simulation on the Johansen's test with different lags to investigate possible size distortion of long lag models. In his simulation, systems of five randomized, non-cointegrated series were generated by cumulating the innovations drawn, without replacement, from the innovations in the log real stock prices used by Kasa. His simulation shows that the size distortion is significant in that the null hypothesis of no cointegration is considerably overejected even if there is no cointegration relation in the simulated data. The problem is particularly acute in VAR models with long lags. Overall, Richards (1995) study suggests that Kasa's findings may be due to failure to adjust asymptotic critical values to take account of the small number of degrees of freedom that remain in the Johansen's test.

Although our study employs the Johansen's small sample correction, our evidence still rejects the null hypothesis of no cointegration (see section 3.2). Furthermore, even when we increase the time span of the analysis from 1970 to 2003, in order to avoid possible market segmentations, the data nevertheless show cointegration relationships. Therefore, it is unlikely that Richards' contribution can exhaustively explain Kasa's finding.

3 Empirical results

3.1 Empirical analysis using Johansen's likelihood ratio test

The data of national stock market indices are from MSCI indices. These indices are end-of-month value weighted indices of a large sample of firms in each national

[5] In Richards (1995), p.639, footnote 7, the author claims that " the Akaike and Schwartz-Bayes information criteria both suggest only one lag in the VAR. Using the Sims likelihood-ratio test with a five percent significance level, one cannot reject successive reductions from $j=10$ until the restriction of $j=3$ is rejected against $j=4$, though this rejection may be spurious; the restriction of $j=1$ is not rejected against $j=4$ at conventional significance levels". However, in our empirical test, the Akaike and Schwartz-Bayes information criteria reported by E-view favors the long lag specification against the short lag. Furthermore, the two criterion statistics are monotonically increasing with regard to lag length, which render the two criteria useless in finding the optimal lag length.

market.[6] For each country, MSCI indices include local currency, US dollar denominated and U.S. CPI adjusted indices. In this paper, dollar denominated and U.S. CPI adjusted indices are used instead of those denominated in local currencies, thus excluding the complication of exchange rate changes. Like most other papers, all stock indices are taken in their natural logarithm form.

We first use Johansen's method to examine the cointegration relationship among five major developed countries' national stock indices as in Kasa (1992). These five countries are: United States, Japan, United Kingdom, Germany, and Canada. Kasa (1992), performed the cointegration tests on monthly and quarterly data in VAR models of various lag lengths. He reports the results with lag of two,VAR (2), and lag of fifteen, VAR (15), specifications. In the VAR (2) specification, the trace test cannot reject the null hypothesis of no cointegration, but the maximum likelihood test shows the existence of one cointegration vector among the five national stock indices. Using the VAR (15) specification, both tests strongly reject the null of no cointegration relation. The quarterly VAR (10) specification rejects the null hypotheses up till $r \leq 3$. Kasa interprets the findings as consistent with the existence of four cointegration vectors, hence, a single stochastic trend.

We replicate Kasa's cointegration tests with E-view 3.0. First, we perform modified Dickey-Fuller test proposed by Elliot et al. (1996) to test for unit roots. As in most empirical studies, our results show little evidence against the null hypothesis of unit root processes.[7] Table 1 reports the results of Johansen's cointegration tests. The four panels (Panel A–Panel D) report the trace statistics of monthly and quarterly data with different lag specifications. We perform the Johansen's test that allows for intercepts and drift terms, which is consistent with Kasa's method. Further, it is quite natural to include a drift term as stock indices have a clear upward trend over time.

For monthly data, the null hypothesis of no cointegration cannot be rejected in VAR with lag of two, but can be rejected in VAR with lag of fifteen. For quarterly data, the null hypothesis is rejected in VAR with lag of two at 5% significance level. In VAR with lag of eight, the no cointegration null hypothesis is strongly rejected at 1% significance level. Moreover, the hypotheses of cointegration rank up to three are all strongly rejected as well, which leads to the conclusion that cointegration rank is four among the five stock indices. In essence, the overall results are very similar to Kasa's findings.[8]

Kasa suggests that the empirical results indicate the existence of a long-run cointegration relationship among the five developed countries' national stock indices. Furthermore, based upon the statistics ($r=4$) from ten lag model of quarterly data, he concludes that there is a common stochastic trend in the long run. However, over-paramaterization in long-lag models might be a serious econometric problem. For example, there are only 67 observations from the first quarter of 1974 to the third quarter of 1990 for each country's quarterly data. In VAR specification

[6] The MSCI indices are widely used in the international finance research with the advantages of being constructed on a consistent basis across countries, regions and by netting out cross-listed securities.

[7] We do not report our test results for brevity and the results are available upon request.

[8] E-view uses the critical values tabulated by Osterwald-Lenum (1992), instead of those of Johansen and Juselius (1990) that is used in Kasa (1992). Consequently, the statistics and critical values are slightly different from those of Kasa's.

Table 1 Johansen's test on five countries stock indices

Eigenvalue	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value	Hypothesized No. of Cointegration Rank
Panel A: K=2 monthly data: 1974:01~1990:08				
0.200137	77.29363	87.31	96.58	None
0.078274	33.07733	62.99	70.05	At most 1
0.035198	16.93881	42.44	48.45	At most 2
0.031470	9.843911	25.32	30.45	At most 3
0.017584	3.512671	12.25	16.26	At most 4
Panel B: K=15 monthly data: 1974:01~1990:08				
0.290241	117.3568	87.31	96.58	None **
0.110901	53.93339	62.99	70.05	At most 1
0.070329	32.18715	42.44	48.45	At most 2
0.062223	18.69611	25.32	30.45	At most 3
0.036147	6.811096	12.25	16.26	At most 4
Panel C: K=2, quarterly data: 1974:01~1990:03				
0.499512	89.45828	87.31	96.58	None *
0.256666	44.46714	62.99	70.05	At most 1
0.173412	25.18751	42.44	48.45	At most 2
0.122705	12.80830	25.32	30.45	At most 3
0.063999	4.299028	12.25	16.26	At most 4
Panel D: K=8, quarterly data: 1974:01~1990:03				
0.851405	307.9960	87.31	96.58	None **
0.830129	197.4171	62.99	70.05	At most 1 **
0.596437	94.59937	42.44	48.45	At most 2 **
0.440795	41.96879	25.32	30.45	At most 3 **
0.132692	8.256926	12.25	16.26	At most 4

Johansen's cointegration test on five developed countries stock indices: U.S., U.K., Japan, Germany and Canada. All data are from MSCI in real U.S. dollar term and taken natural log. K denotes the number of lag specified in the VAR system. The tests are performed by E-view 3.0. Critical values are from Michael Osterwald-Lenum (1992)

*(**) denotes rejection of the null hypothesis at 5%(1%) significance level

with ten lags, there are fifty parameters of lagged differences that need to be estimated. In order to examine the relationship between the lag length and the detecting capability of cointegration, we run the Johansen test on the quarterly data from the period of 1970 to 2003 with the same specification. Thus, we have much longer time series in our sample. First, we run Chow's test (1960) to detect if there is any structure change between the subperiods in Kasa's study and the extended subperiod after 1990. Chow's statistics show that the p-values of the five index returns are well above 0.4. Thus, we cannot reject the null that there is no structural change between these two subperiods. We proceed to the Johansen's test on the extended time series and report in Table 2 the cointegration rank of various specification of lag length in the model.

From Table 2 panel A, it is clear that the null of no cointegration cannot be rejected until taking lag of ten. If Kasa's interpretation is correct, we should be able to detect cointegration when k is eight. In fact, in Kasa's quarterly dataset that has nearly two-thirds of the samples in our dataset, null hypothesis of no cointegration is rejected when k is two. If the long-run cointegration does exist, its detection should not depend on the lag length chosen when we employ a longer period of data samples. In the VAR (14) model where cointegration rank is four, there are only a hundred and twenty observations against seventy parameters to be estimated. The emerging pattern is that over-parameterization of long lag model coupled with limited samples may cause over-rejection the null hypothesis.

We then conduct the tests on a VAR system of three countries' stock indices: U.S., U. K. and Canada. Among the five countries in Kasa's original study, it can be argued that U.S., U.K., and Canada are more interconnected economically with each other. If the five countries stock indices are cointegrated and driven by one common stochastic trend, we would expect to detect a cointegration relation among the stock indices of the three countries.

Table 2 Johansen test with different lag specifications

Lag length	Cointegration rank	Lag length	Cointegration rank	Lag length	Cointegration rank
Panel A: quarterly data of U.S., U.K., Japan, Canada, and Germany, 1970:01~2003:01					
$K=1$	$r=0$	$k=7$	$r=0$	$k=13$	$r=3$
$K=2$	$r=0$	$k=8$	$r=0$	$k=14$	$r=4$
$K=3$	$r=0$	$k=9$	$r=0$	$k=15$	$r=5$
$K=4$	$r=0$	$k=10$	$r=0$	$k=16$	$r=5$
$K=5$	$r=0$	$k=11$	$r=1$	$k=17$	$r=5$
$K=6$	$r=0$	$k=12$	$r=2$	$k=18$	$r=5$
Panel B: quarterly data of U.S., U.K., and Canada, 1970:01~2003:01					
$K=1$	$r=0$	$k=9$	$r=0$	$k=17$	$r=1$
$K=2$	$r=0$	$k=10$	$r=0$	$k=18$	$r=1$
$K=3$	$r=0$	$k=11$	$r=0$	$k=19$	$r=2$
$K=4$	$r=0$	$k=12$	$r=1$	$k=20$	$r=2$
$K=5$	$r=0$	$k=13$	$r=2$	$k=21$	$r=2$
$K=6$	$r=0$	$k=14$	$r=2$	$k=22$	$r=3$
$K=7$	$r=0$	$k=15$	$r=0$	$k=23$	$r=3$
$K=8$	$r=0$	$k=16$	$r=1$	$k=24$	$r=2$
Panel C: quarterly data of U.S., U.K., Japan, Canada, and Germany, Australia, Hong Kong, Germany, Italy, and Singapore, 1970:01~2003:01					
$K=1$	$r=0$	$k=5$	$r=6$	$k=9$	$r=10$
$K=2$	$r=1$	$k=6$	$r=6$	$k=10$	$r=10$
$K=3$	$r=2$	$k=7$	$r=8$	$k=11$	$r=10$
$K=4$	$r=3$	$k=8$	$r=10$		

Johansen's cointegration test on multivariate systems of different country stock indices. All data from MSCI in real U.S. dollar term and taken natural log. K denotes the number of lag specified in the VAR system; r denotes the number of cointegration rank determined from Johansen's cointegration test.

Table 2 panel B reports the lag length and cointegration rank for U.S. U.K and Canada quarterly stock indices from 1970 to 2003. The results show that the no cointegration hypothesis cannot be rejected until the VAR (11) specification. We would expect to observe a much stronger cointegration relationship between the three “Anglo-Saxon” markets than with other markets, given similar economical and legal systems. However, the empirical results show weaker evidence in favor of cointegration relation among these three “cousin” markets. It is conceivable that the over-parameterization of the long lag model, coupled with limited sample size lead to the rejection of no-cointegration null.

Next, we perform the Johansen test on a sample with ten countries’ stock indices. In addition to the original five countries, we add Australia, Hong Kong, Germany, Italy and Singapore. As the VAR model now has a dimension of ten, the over-parameterization problem would become more serious for long lag model (one more lag adds ten additional coefficients for each country). If over-parameterization were the cause for the rejection of null in long lag model, we would expect the statistical problems to be more acute in the ten-country system, hence, the results should show much stronger evidence against no cointegration null. Table 2 panel C reports the Johansen’s test for the ten-country system. From the cointegration rank and the lag length, we clearly notice that the data level off when $k=2$, and the model detects eight cointegration vectors at lag length seven. Overall, it is highly surprising and counter to economic intuition that we find much stronger cointegration relationship when we add some more “peripheral” countries in the model. Of course this type of result casts some doubts on the accuracy of the statistical tests performed and warrants further enquiry in the following subsection[9].

3.2 Small sample correction for the Johansen’s test

As we previously mentioned, simulation studies have shown that small sample properties of the trace test are much different from asymptotic ones. Hence, the first robustness check for Kasa’s results is to test whether the small sample size is the main cause of finding the cointegration relations.

Johansen (2002) addresses the small sample problem for his likelihood ratio test in the VAR model. The distribution of the likelihood ratio test statistic depends on the sample size T and the number of VAR parameters under the null hypothesis. If T goes to infinity, the dependence on VAR parameters disappears. If the parameters are close to the boundary where the cointegration properties change, the approximation can be very poor. Johansen (2002) proposed a correction factor to the trace statistic to improve the finite sample properties. The idea is similar to the Bartlett correction (Bartlett 1937). Bartlett suggested finding the expectation of the likelihood ratio test statistic and then correcting it to have the same mean as the limit distribution.

[9] Interestingly, if we take sufficiently long lag specifications, the null hypothesis that $r\leq(n-1)$ can be rejected like other possible values of r for all the datasets with which we perform the Johansen’s test. It means that the rank of the matrix Π is n, the full rank, which is even contradictory to the assertion that the variables are I(1) in the first place.

Likewise, Johansen introduced a correction factor to the likelihood ratio test that improves the finite sample properties.[10]

Johansen also demonstrated some special cases where explicit form solutions can be obtained. In one special case for the test of no cointegration in models with two lags, a nominal 5% size test has a rejection probability of 83% using asymptotic values with a sample size of 50. The correction brings down the rejection rate back to 4%. In another case, for the test of rank one in models with one lag, the correction factor reduces the rejection probability from 29.9% to around 9% with a sample size of 50. In Kasa's paper, the quarterly data sample has a size of 69, and each lag reduces the sample size by one. The likelihood ratio tests on the quarterly data may well be subject to the small sample problem. Johansen shows that small sample properties causes the over rejection problem. It is likely that Kasa's empirical results are driven by over-rejection of the Johansen's test on small sample data.

We implement the Johansen's correction factor to address the possible small sample bias. We use Matlab to estimate the model parameters under the null hypothesis and calculate Johansen's correction factor. In this study, the approximation tables in Johansen (2002) are used to simplify the programming task. Table 3 reports the trace statistics adjusted by the correction factor for Kasa's quarterly data. The evidence for cointegration is much weaker after correcting for the small sample. Without correction, for all k, the null of no cointegration can be rejected. After the correction, for k=2 and 3 the null of no cointegration cannot be rejected. Furthermore, for k=4, the null of no cointegration cannot be rejected at the 1% level. However, the basic pattern of Kasa's finding is still present: the test tends to reject the null hypothesis with long lag VAR specification. Overall, our results indicate that the small sample correction makes the rejection of the null occur in longer lag than without the correction. We also use the correction factor on the dataset with longer period and different countries. Although we do not report here these results, the basic statistical patterns are similar.

3.3 Empirical analysis using a non-parametric cointegration rank test

The previous subsection shows that the null hypothesis of no cointegration is rejected after correcting for small sample size, although the results are much weaker. The key issue seems to be the validity of taking long lag VAR specification in investigating the cointegration relationship. If we can use a test that does not require VAR model specification, we might be able to shed some new light on this important issue. Shintani's (2001) non-paramentric cointegration rank test satisfies this requirement and helps us further investigate the long run cointegration among international stock indices. In this subsection we implement the non-parametric test on the original five countries with extended sample period from 1970 to 2003.

Shintani (2001) proposed a non-parametric cointegration rank test which does not require the assumption of VAR models. The test exploits the well-known fact that the number of cointegration vectors is identical to the degree of degeneration in the space spanned by the sample moment matrix in the limit. This idea of using the

[10] For technical details see Appendix A.

Table 3 Cointegration test with Johansen's small sample correction

Lag length	Trace statistic	Correction factor	Corrected statistic	5% critical value	1% critical value
k=2	89.46	1.14	78.19	87.31	96.58
k=3	93.30	1.14	81.91	87.31	96.58
k=4	108.19	1.13	95.62*	87.31	96.58
k=5	117.47	1.15	102.5**	87.31	96.58
k=6	146.06	1.16	126.08**	87.31	96.58
k=7	154.98	1.17	132.14**	87.31	96.58
k=8	229.97	1.27	181.61**	87.31	96.58

This table reports the small sample correction for Johansen test on the quarterly data for Kasa's five countries stock indices from 1974 to 1990. The trace statistic and correction factors are for testing H_0: r= 0, i.e. no cointegration vector exists. The parameter values required from estimating correction factors are from Johansen et al. (2002)

*(**) denotes rejection of the null hypothesis at 5%(1%) significance level

degeneracy of the covariance matrix was first employed by Phillips and Ouliaris (1990). Unlike the VAR-based tests, the sample moment is standardized using the nonparametric long-run variance estimation rather than lags in VAR models. Using the properties of nonparametric spectrum estimation, the test statistic based on the eigenvalues follows a familiar matrix unit root distribution in the limit under the null hypothesis of a specified cointegration rank.

We perform the Shintani's non-parametric test on the five countries' national stock indices with both monthly and quarterly data. We use Matlab to compute covariance matrices and their eigenvalues. The critical values are tabulated by Phillips and Ouliaris (1990). Since only critical values up to the system of five-dimension Brownian motions are available, we perform our empirical tests on the national stock indices of five countries. In the test, we first detrend the time series of national stock indices time series. We use the Parzen kernel and the Quadratic Spectral (QS) kernel to estimate the long run covariance matrix of the first difference. The Parzen kernel and QS kernel are widely used in the spectrum analysis and satisfies the Kernel condition assumptions in Shintani's method (Shintani 2001, p.342).[11]

We use Andrews' (1991) method to choose the bandwidth K. Andrews employed an asymptotic truncated mean squared error optimality criterion for heteroskedasticity and autocorrelation (HAC) consistent estimation of covariance matrices of parameter. His methodology is general enough to accommodate loose error assumptions. Andrews (1991) shows that Monte Carlo simulations indicate that the difference between the kernels are not large, but the Bartlett kernel is somewhat inferior to the others. He also shows that the automatic bandwidth parameters for HAC estimators perform well in most cases. In this paper, we use Andrews' automatic bandwidth parameter for the Parzen kernel and QS kernel. The computations show that optimal bandwidth parameters equal four and one for Parzen and QS kernels, respectively.

[11] Besides the Parzen kernel, the Bartlett kernel, Tukey-Hanning kernel, and Quadratic Spectral kernel are widely used in spectrum analysis as well. For the estimation with different kernels, we refer to Andrews' (1991) paper for the details of the econometric technicalities.

We apply Shintani's test in four sets of data. Two sets contain the five-country indices as in the original Kasa's study from 1974 to 1990, with monthly and quarterly frequencies. The other two sets contain the same five-country indices but from 1970 to 2003, with monthly and quarterly frequencies as well. The test results are reported in Table 4. The tests show that p-value are not significant even at the 15% significance level (critical values are from Phillips and Ouliaris (1990), Table IVa). The p-value of using both Parzen kernel and QS kernel are very close and cannot reject the null hypothesis of no cointegration. The results are similar on both monthly and quarterly data originally investigated by Kasa. The empirical finding of no cointegration is also robust to extended dataset from 1970 to 2003 (Table 4, Panel A and B). The empirical results of the non-parametric tests support the view that Kasa's results may be largely caused by the model misspefication with extremely long lag VAR. In fact, when we use the empirical tests that need not specify VAR model, we cannot reject the null hypothesis of no cointegration among the national stock indices.

Table 4 Test of cointegration among five national stock indices by Shintani non-parametric test

Rank	Parzen kernel	QS kernel	5% critical value	15% critical value
Panel A: monthly data from 1970:01 to 2003:03				
S=0	159.58	161.37	210.29	186.43
S=1	96.9	98.57	155.80	135.27
S=2	45.93	46.46	109.74	92.76
S=3	15.02	15.48	71.28	56.77
S=4	3.80	3.80	40.82	30.01
Panel B: quarterly data from 1970:01 to 2003:01				
S=0	153.83	152.89	210.29	186.43
S=1	98.23	97.37	155.80	135.27
S=2	46.42	46.14	109.74	92.76
S=3	17.00	16.84	71.28	56.77
S=4	3.70	3.69	40.82	30.01
Panel C: monthly data from 1974:01 to 1990:08				
S=0	172.52	170.83	210.29	186.43
S=1	83.21	82.54	155.80	135.27
S=2	46.84	46.63	109.74	92.76
S=3	23.38	23.18	71.28	56.77
S=4	6.42	6.27	40.82	30.01
Panel D: quarterly data from 1974:01 to 1990:03				
S=0	155.10	152.70	210.29	186.43
S=1	78.69	77.48	155.80	135.27
S=2	42.33	42.11	109.74	92.76
S=3	21.28	21.06	71.28	56.77
S=4	6.97	6.84	40.82	30.01

This table reports the shintani test on five national stock indices: U.S., U.K. Japan, Canada, and Germany. Andews (1991)'s methods is used to decide automatic optimal bandwidth for the Parzen and QS kernels. The critical values are from Phillips and Ouliaris(1990), Table IV a

4 Conclusion

This paper investigates the issue first raised by Kasa (1992) on long run cointegration relationships among national stock indices. The extended studies using Kasa's methodology suggest that his finding on cointegration relationship may be attributed to the small sample problem and the long lag VAR specification. By employing Johansen's small sample correction, our study shows that the cointegration relationship is much weaker than without correction. Furthermore, in order to avoid the problem of misspecification of VAR model, we perform a recently developed non-parametric cointegration rank test. The test cannot reject the null hypothesis of no cointegration, both on Kasa's data and the data over an extended period from 1970 to 2003. Therefore, our study shows the evidence consistent with the view that Kasa's anomalous finding is most likely due to the size distortion of Johansen test performed on datasets with small sample sizes and using long-lag VAR models.

Appendix

Appendix A. The implementation of Johansen's small sample correction on cointegration test

Johansen (2002) proposed a following correction factor to the trace statistic to improve the finite sample properties.

$$\frac{f(n-r,n_d)}{f(T,n-r,n_d)}\frac{-2\log LR}{\left(1+T^{-1}b\left(\widehat{\theta}\right)\right)}=\frac{-2\log LR}{a(T,n-r,n_d)\left(1+T^{-1}b\left(\widehat{\theta}\right)\right)}$$
$$f(n-r,n_d)=\lim_{T\to\infty}f(T,n-r,n_d)$$
$$a(T,n-r,n_d)=\frac{f(T,n-r,n_d)}{f(n-r,n_d)}$$
$$b(\theta)=c_1(1+h(n_b,n_d))+(n_bc_2+2(c_3+n_dc_1))\frac{g(n_b,n_d)}{n_b^2}-2tr\left[\psi'\Sigma^{-1}\sum_{j=0}^{\infty}\psi_j\right]k(n_b,n_d,j) \tag{A.1}$$

in which, T is the sample size; n is the number of variables in the VAR; r is the rank in the null hypothesis; n_b is n–r; n_d is the parameter to denote the deterministic term. The detailed definitions of all the terms above, we refer to Johansen (2002), page 1933 to 1938.

We implement the above Johansen's correction factor to adjust the probable small sample problem. Johansen (2002) provided approximation for the *a, h, g* and *k* functions in the correction formula. In Johansen et al. (2002), more detailed simulation tables for functions *a, h, g* and *k* are provided. Johansen (2002) also suggests *k* function can be approximated to zero. Parameters *c1, c2* and *c3* in the *b* function are related to the model parameters estimated under the null hypothesis. We use Matlab to estimate the model parameters under the null hypothesis and calculate Johansen's correction factor. In this study, the approximation tables in Johansen

(2002) are used to simplify the programming task. The approximations for functions a, h and g in the case of determined trend (n_d=1) are listed as follows:

The approximation of $a(T,n_b,n_d)$ is of the form $1+a1(n_d)\, n_b/T+a2(n_d)(n_b/T)^2+ a3(n_d)\,(n_b/T)^3+b(n_d)/T$. The approximation of $h(n_b,n_d)$ is of the form $h1(n_d)/n_b+ h2(n_d)/(n_b)^2+h3(n_d)/(n_b)^3$. The approximation of $g(n_b,n_d)$ is $g0(n_d)+g1(n_d)/n_b+ g2(n_d)/(n_b)^2+g2(n_d)/(n_b)^3$.

Appendix B. The implementation of shintani's nonparametric cointegration test

Shintani's test starts with an n-vector process generated by

$$z_t = z_{t-1} + \xi \tag{B.1}$$

where ξ_t is a n-vector stationary innovation sequence.

If z_t has r cointegration vectors, then we can rotate z_t into I(0)/I(1) subsystems with matrix H as:

$$\begin{aligned} \underline{z}_t = H'z_t = \begin{bmatrix} H_1'z_t \\ H_2'z_t \end{bmatrix} = \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} I(0) \\ I(1) \end{bmatrix} \\ H = [H_1 : H_2] \end{aligned} \tag{B.2}$$

The transformed system can be rewritten as in Phillips and Solo (1992, p.985):

$$\begin{aligned} z_{1t} &= H'C^*(L)\varepsilon_t = u_{1t} \\ \Delta z_{2t} &= H_2'C(L)\varepsilon_t = u_{2t} \end{aligned} \tag{B.3}$$

in which ε_t is white noise.

For the technical detail of the error conditions as maintained assumptions we refer to Shintani (2001) p. 340. As pointed by Engle and Granger (1987, p.260), when both I(0) and I(1) elements appear in the system, the sample covariance matrix of z_t degenerates in the limit to:

$$T^{-1}M_{\underline{zz}} = T^{-2}\sum_{t=1}^{T}\begin{bmatrix} z_{1t}z_{1t}' & z_{1t}z_{2t}' \\ z_{2t}z_{1t}' & z_{2t}z_{2t}' \end{bmatrix} \xrightarrow{d} \begin{bmatrix} 0 & 0 \\ 0 & \int_0^1 B_2B_2' \end{bmatrix} \tag{B.4}$$

Here B_2 is $n–r$ dimension standard Brownian motion.

Similarly, for the sample covariance matrix of Δz_t

$$\widehat{\Omega}_{\Delta\underline{z}\Delta\underline{z}} \xrightarrow{p} \Omega_{\Delta\underline{z}\Delta\underline{z}} = \begin{bmatrix} \Omega_{\Delta u1\Delta u2} & \Omega_{u2\Delta u1} \\ \Omega_{\Delta u1u2} & \Omega_{u2u2} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \Omega_{22} \end{bmatrix} \tag{B.5}$$

Shintani employs the kernel estimator to estimate the covariance matrix of Δz_t

$$\widehat{\Omega}_{\Delta\underline{z}\Delta\underline{z}} = \sum_{j=-T+1}^{T-1} w(j/K)\widehat{\Gamma}_{\Delta z\Delta z}(j), \quad \widehat{\Gamma}_{\Delta z\Delta z}(j) = T^{-1}\sum_{1\le t,t+j\le T}\Delta z_{t+j}\Delta z_t \tag{B.6}$$

Where w( ) is a kernel function, K is a bandwidth parameter.

For testing the hypotheses of the cointegration rank of the n variable system,

H_0: r=s vs. H_A: r>s,

Let $\lambda_i(A)$ be the ith largest eigenvalue of a matrix A. The test statistic is:

$$\begin{aligned} P(n,s) &= T \sum_{i=s+1}^{n} \hat{\lambda}_i \\ \hat{\lambda}_i &= \lambda_i\left(\widehat{\Omega}_{\Delta \underline{z} \Delta \underline{z}} M_{zz}^{-1}\right) \end{aligned} \tag{B.7}$$

where λ_i is ith largest eigenvalue of the matrix in the parenthesis. Hence, in order to obtain the statistics, we need to estimate the covariance matrix of Δz_t and the inverse of the sample covariance matrix of $z_{t.}$ We then find out the eigenvalues of the product of the two covariance matrix, add them up and multiply the sample size T to obtain the statistic.

Assuming certain technical conditions, Shintani (2001, theorem 3.1) shows that under the null hypothesis,

$$P(n,s) \xrightarrow{d} tr\left\{\left(\int_0^1 W_{n-r} W_{n-r}'\right)^{-1}\right\} \tag{B.8}$$

W_{n-r} is standard (n–r) dimension Wiener process. The integrals of Wiener processes are readily available.

In implementing the test, we use the critical values tabulated by Phillips and Ouliaris (1990). We use the Parzen kernel Quadratic Spectral (QS) kernel to estimate the long run covariance matrix of the first difference. The Parzen kernel and QS kernel are widely used in the spectrum analysis and satisfies the Kernel condition assumptions in Shintani's method (Shintani 2001, p.342). We use Andrews' (1991) method to choose the bandwidth K. Andrews employed an asymptotic truncated mean squared error optimality criterion for heteroskedasticity and autocorrelation (HAC) consistent estimation of covariance matrices of parameter.

References

Andrews DWK (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59(3):817–858

Bartlett MS (1937) Properties of sufficiency and statistical tests. Proc R Soc Lond A 160:268–282

Bekaert G, Hodrick RJ, Zhang X (2009) International stock return comovements. J Finance 64(6):2591–2626

Bewley R, Yang M (1995) Testing for cointegration: the effects of mis-specifiying the lag length. Math Comput Simul 39(3–4):251–255

Boswijk P, Franses PH (1992) Dynamic specification and cointegegration. Oxf Bull Econ Stat 54(3):369–381

Cheung YW, Lai KS (1993) Finite-sample sizes of Johansen's likelihood ratio tests for cointegration. Oxf Bull Econ Stat 55(3):313–328

Chow GC (1960) Tests of equality between sets of coefficients in two linear regressions. Econometrica 52:211–222

Conrad J, Kaul G (1988) Time-variation in expected returns. J Bus 61(4):409–425

DeBondt WFM, Thaler RH (1985) Does the stock market overreact? J Finance 40:793–805

Elliot G, Rothenberg TJ, Stock JH (1996) Efficient tests for an autoregressive unit root. Econometrica 64:813–836

Engle RF, Granger CWJ (1987) Cointegration and error correction: presentation, estimation, and testing. Econometrica 55(2):251–276

Eun CS, Resnick BG (1994) International diversification of investment portfolios: U.S. and Japanese perspective. Manage Sci 40(1):140–161

Fama EF, French KR (1988) Permanent and temporary components of stock prices. Journal of Political Economy 96:246–273

Fama EF, French KR (1989) Business conditions and expected returns on stocks and bonds. J Financ Econ 25:26–49

French KR, Poterba J (1991) Investor diversification and international equity markets. Am Econ Rev 81 (2):222–226

Gonzalo J, Lee TW (1998) Pitfalls in testing for long run relationships. J Econometrics 86:129–154

Gonzalo J, Pitarakis J (1999) Dimensionality effect in cointegration tests. In: Engle RF, White H (eds) Cointegration, Causality and Forecasting. A Festschrift in Honour of C.W.J. Granger. Oxford.

Granger CWJ (1986) Developments in the study of cointegrated economic variables. Oxf Bull Econ Stat 12:347–360

Gregory AW (1994) Testing for cointegration in linear quadratic models. J Bus Econ Statistics 12:347–360

Grubel H (1968) Internationally diversified portfolios: welfare gains and capital flows. Am Econ Rev 58:1299–1344

Haug A (1996) Tests for cointegration: a Monte Carlo comparison. J Econometrics 71(1–2):89–115

Ho MS, Sorensen BE (1996) Finding cointegration rank in high dimensional systems using the Johansen test: an illustration using data based on Monte Carlo simulations. Rev Econ Stat 78(4):726–732

Johansen S (1988) Statistical analysis of cointegration vectors. J Econ Dyn Control 12:231–254

Johansen S (1991) Estimation and hypothesis testing of cointegrated vectors in Gaussian vector autoregressive models. Econometrica 59(6):1551–1580

Johansen S (2002) A small sample correction for the test of cointegration rank in the vector autoregressive model. Econometrica 70(5):1929–1961

Johansen S, Juselius K (1990) Maximum likelihood estimation and inference on cointegration with applications to the demand for money. Oxf Bull Econ Stat 52:169–210

Johansen S, Hansen H, Fachin S (2002) A simulation study of some functionals of random walk, Working paper

Kasa K (1992) Common stochastic trends in international stock markets. J Monetary Econo 29:95–124

Osterwald-Lenum M (1992) A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics. Oxf Bull Econ Stat 54(3):461–72

Phillips PCB, Ouliaris S (1990) Asymptotic properties of residual based tests for cointegration. Econometrica 58(1):165–193

Phillips PCB, Solo V (1992) Asymptotics for linear processes. Ann Stat 20(2):971–1001

Poterba JM, Summers LH (1988) Mean reversion in stock prices: evidence and implications. J Financ Econ 22:27–59

Richards AJ (1995) Comovements in national stock market returns: evidence of predictability, but not cointegration. J Monetary Econ 36:631–654

Shintani M (2001) A simple cointegration rank test without vector autoregression. J Econometrics 105:337–362

Stock JH, Watson M (1988) Testing for common trends. J Am Stat Associations 83(404):1097–1107

Toda HY (1995) Finite sample performance of likelihood ratio tests for cointegrating rank in vector autoregressions. Econometric Theory 11:1015–1032