On The Persistence of Cointegration in Pairs Trading[1]

By Matthew Clegg[2]

Abstract: An exploratory study is conducted to assess the persistence of cointegration among U.S. equities. In other words, if a pair of equities is found to be cointegrated in one period, is it likely that it will be found to be cointegrated in the subsequent period? An examination is performed of pairs formed from constituents of the S&P 500 during each of the calendar years 2002-2012, comprising over 860,000 pairs in total. The evidence does not support the hypothesis that cointegration is a persistent property.

Introduction

Pairs trading is a market neutral strategy for profiting from short-term divergences in securities prices. The idea behind pairs trading is simple. Two securities are identified whose prices tend to move together. When an unusually large gap opens between their prices, the cheaper security is bought long and the more expensive security is sold short. When the gap reverts to its historical levels, the position is closed and a profit is collected.

Cointegration [1] offers a statistical framework that seems well-suited to pairs trading. Engle and Granger gave a two-step procedure for determining whether or not two series are cointegrated and for identifying the cointegrating relations. If P_t and Q_t are the price series of two securities, a check is made to ensure that P_t and Q_t are each integrated. If they are, then the first step is to use ordinary least squares to find α and β such that

$$\log P_t = \alpha + \beta \log Q_t + R_t, \qquad (1)$$

where R_t is the residual series[3]. In the second step, ordinary least squares is then used to find ρ such that

---

[1] This paper is an expanded version of a talk that was presented at the R/Finance 2014 conference in May 2014.

[2] E-mail: matthewcleggphd@gmail.com

[3] Some practitioners will use ordinary prices rather than logged prices. In this study, the choice seems to be immaterial.

Electronic copy available at: http://ssrn.com/abstract=2491201

$$R_t = \rho R_{t-1} + \varepsilon_t, \qquad (2)$$

where ε_t is a series of innovations that is presumed to be independent and identically distributed and to have zero mean. The value of ρ that is fitted in the above equation will normally be between zero and one. This is called the coefficient of mean reversion. A test is performed to check whether or not $\rho = 1$. If $\rho = 1$, then the residual series contains a unit root. In this case, the pair is not cointegrated and must be discarded. However, if $\rho < 1$, then the pair is cointegrated and may be considered for trading.

The connection between cointegration and pairs trading has spawned a significant literature. Many published papers follow a common recipe. First, a pairs trading algorithm is articulated. Next, a universe of securities is identified and a study period is defined. The study period is broken down into a sequence of formation periods, each of which is followed by a trading period. In each formation period, all possible pairs of securities are considered, and those that are identified as cointegrated are considered for trading. In the corresponding trading period, a simulated backtest is conducted to assess the performance of the trading algorithm applied to the cointegrated pairs. Finally, the results are tabulated and reported. Often, unusually high positive trading returns are found.

In the context of pairs trading, the cointegration test distinguishes between two hypotheses:

$\mathcal{H}_0$ During the formation period, the price series of the two securities were not cointegrated.

$\mathcal{H}_1$ During the formation period, the price series of the two securities were cointegrated.

Rejection of $\mathcal{H}_0$ is not sufficient to guarantee success in pairs trading, though. This is because cointegration-based pairs trading algorithms seem to rely upon a different hypothesis, namely:

$\mathcal{H}_2$ During the trading period, the price series of the two securities will be cointegrated.

In fact, these algorithms don't just rely upon this assumption but upon the stronger assumption:

$\mathcal{H}_3$ During the trading period, the price series of the two securities will be cointegrated, and the parameter estimates obtained in the formation period will be valid for the trading period.

A number of published papers seem to implicitly assume that satisfaction of $\mathcal{H}_1$ implies satisfaction of $\mathcal{H}_2$ or even $\mathcal{H}_3$. However, this assumption may not be justified. The purpose of this study is to empirically examine the hypothesis that $\mathcal{H}_1$ implies $\mathcal{H}_2$ or that $\mathcal{H}_1$ implies $\mathcal{H}_3$.

In other words, the fundamental question that this paper addresses is, "Given that a pair is found to be cointegrated in one period, with what level of confidence can it be asserted that this

Electronic copy available at: http://ssrn.com/abstract=2491201

pair will continue to be cointegrated in the subsequent period?" How persistent is the property of cointegration?

There have been numerous recent studies that have examined cointegration-based pairs trading and reported back-testing results. See for example [2], [3], [4], [5] and the references therein. Several recent books have been published about pairs trading, such as [6], [7] and [8]. A number of recent student theses and dissertations that describe pairs trading and cointegration have also appeared. These include [9], [10], [11], [12] and [13].

The methodology for selecting pairs for trading varies considerably. In some studies, the length of the formation period can be as little as 60 days, while in other studies it can be several years. Some studies form portfolios consisting of all qualifying pairs, while others choose only the top 5, 20 or 50. Some use the Dickey-Fuller test for testing for cointegration, while others use alternate tests. What effects do these decisions have on the outcomes that are reported? A secondary goal of this paper is to attempt to shed light on how these choices may affect a pairs trading strategy.

The remainder of this paper is organized as follows. An initial assessment of pairs from the year 2002 is conducted to help frame the problem. Next, a discussion of how to improve the effectiveness of the cointegration test ensues. The data set used for the remainder of the paper is next described. It consists of more than 860,000 pairs formed from securities belonging to the S&P 500 over an eleven year period. Next, tests are performed on the persistence of cointegration and the results are presented. Finally, conclusions are given.

A Preliminary Example

To get an initial reading on the frequency of cointegration, an experiment was conducted. The S&P 500 constituents list was downloaded from the Standard & Poor's web site on August 13, 2013. The adjusted closing prices of these securities for the years 2002 and 2003 were downloaded from Yahoo. Of these 500 securities, 162 of them had a complete set of observations for 2002 and 2003. From this data set were eliminated all securities whose logged price series were identified by the augmented Dickey-Fuller test as being stationary, using a p-value of 0.05. Three such series were identified. Using the Engle-Granger procedure, all pairs of stocks among the remaining 12,561 pairs were identified whose logged price series were found to be cointegrated in the year 2002. The results of this analysis are summarized in the following table.

	N	%	$\overline{\rho}$	$sd(\overline{\rho})$
$p = 0.05$	649	5.17%	0.914	0.037
$p = 0.01$	106	0.84%	0.895	0.045
Top 50	50	0.40%	0.886	0.049
Top 20	20	0.16%	0.866	0.039

Pairs Identified as Cointegrated out of 159 Securities in 2002

The first column reports the number of pairs found to be cointegrated. The percentage of the total population is given in the second column. The third column reports the mean of the coefficients of mean reversion for the pairs found to be cointegrated, and the fourth column reports the standard deviation.

Data source: Adjusted closing prices reported by Yahoo! for securities that were constituents of the S&P 500 as of August 13, 2013.

Of the 12,561 pairs that were examined, 649 (5.17%) had a p-value from the augmented Dickey-Fuller test less than or equal to 0.05, which represents an excess of 21 pairs above the number that would be predicted if there were no cointegration relations among them.

This could be viewed as evidence that at least some of the pairs were cointegrated. How strong is this evidence? If the cointegration tests are independent of one another, it can be expected that the number of false positives at the 5% level out of 12,561 pairs will be distributed according to a Poisson distribution with mean 628 (0.05 x 12,561). The probability of observing 649 or more false positives would then be 0.195, which is not statistically significant. The value reported for $\bar{\rho}$ is the mean value of the coefficient of mean-reversion. The half-life of a process with a mean-reversion coefficient of 0.914 is 7.7 days.

The second row of this table reports the figures for when the cutoff p-value is taken to be 0.01. The number of qualifying pairs is slightly below the number that would be expected if no cointegration relations existed among the securities.

The third row of the table reports the figures for the 50 pairs with the smallest ADF statistic. In other words, these are the 50 pairs having the smallest p-values. Similarly, the fourth row of the table reports the figures for the 20 pairs with the smallest ADF statistic. Thus, these are the 20 pairs for which the evidence of mean reversion appears to be the most extreme, at least when viewed through the lens of the augmented Dickey-Fuller test.

For each pair that was identified as being potentially mean-reverting, the Engle Granger procedure was then performed for the following year and its ADF statistic was calculated. In other words, a check was performed to see if the property of being cointegrated persisted into the following year. The following table summarizes the results that were obtained, where the critical value is set to $p = 0.05$.

	N(I1)	N(CI)	%	ρ	$sd(\overline{\rho})$
$p = 0.05$	507	22	4.34%	0.943	0.036
$p = 0.01$	75	1	1.33%	0.962	--
Top 50	33	0	0.00%	--	--
Top 20	11	0	0.00%	--	--

Pairs that Continued to Exhibit Cointegration in Second Year (2003)

Of those pairs found to be cointegrated in 2002, the first column reports the number of pairs for which both components were again integrated in 2003. The second column reports the number of pairs from this subgroup that were also cointegrated, and the third column reports this number as a fraction of the number of number of integrated pairs. The fourth and fifth columns report the mean and standard deviation of the coefficient of mean reversion for the cointegrated pairs.

Of the 649 pairs that were identified as being cointegrated in 2002, there were 507 pairs for which both securities were identified as being integrated in 2003. Of these, only 22 (or 4.34%) were identified as cointegrated. When $p = 0.05$, it would be expected that 5% of the pairs would be falsely identified as cointegrated. Thus, there is no evidence that cointegration persisted into the second year. Even less favorable results were obtained when setting $p = 0.01$. Of the 106 pairs identified as cointegrated in 2002, only one was identified as cointegrated in 2003. Of the 50 pairs with the most extreme p-values in 2002, none were found to be cointegrated in 2003.

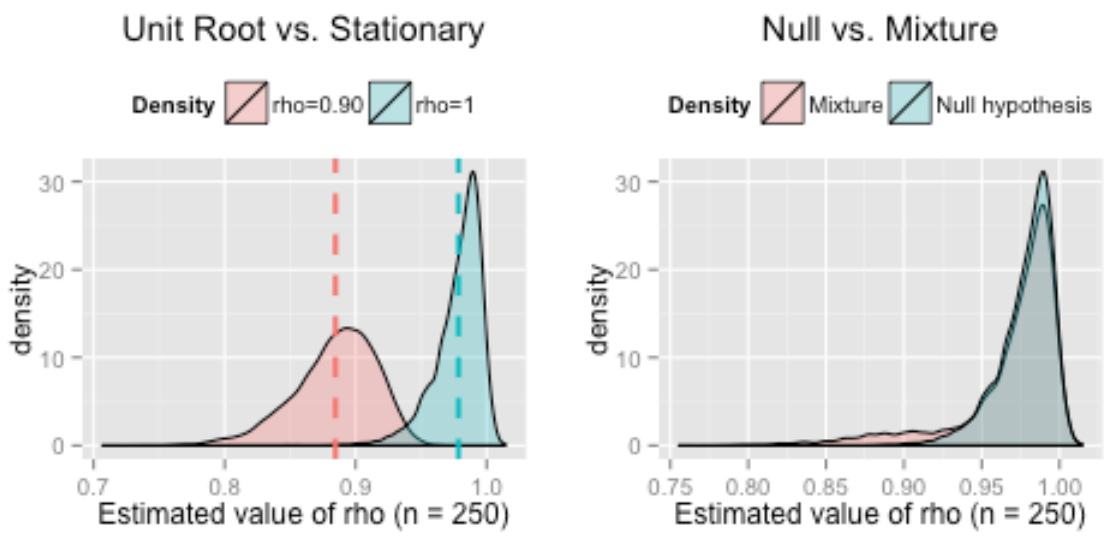
Although this data appears to show low levels of persistence of cointegration, it actually presents an overly optimistic portrayal of the true picture that would be faced by a trading algorithm. That is because the test for cointegration in the trading period was performed by estimating the parameters α and β from the prices that occurred in this period. These are results obtained from using in-sample parameter estimates. A pairs trading algorithm would not have this data available to it. It would be constrained to estimate α and β from data available during the formation period and to use these estimated values during the trading period.

INCREASING THE POWER OF THE COINTEGRATION TEST

In the previous section, it was shown that the standard procedure for identifying cointegrating pairs failed to identify a significant number of pairs in 2002 (or perhaps any pairs) for which cointegration persisted into the second year. Perhaps a significant number of cointegrated pairs were present in the data, but the cointegration test simply failed to identify them. In this case, one should search for a more powerful cointegration test. Therefore, in this section, various alternative cointegration tests are examined.

To test for cointegration, a check is performed to see if the mean reversion parameter ρ is significantly different from one. Although ρ is estimated using a simple linear regression, the standard error of the linear regression is not the correct estimator of the confidence bound for

ρ. This is because the distribution of the estimates of ρ when $\rho = 1$ will be non-normal. The following diagram illustrates the situation.



Comparison of Densities of the Estimates of ρ for Mean Reverting Processes to Unit Root (Random Walk) Processes.
The left-hand figure displays the simulated density of $\hat{\rho}$ for a mean reverting process with $\rho = 0.9$ versus the simulated density of a unit root process. The right-hand figure displays the density of a mixture, where 90% of the samples are random walks and 10% are mean reverting with $\rho = 0.9$. The density of the null hypothesis (random walk) is displayed for comparison. The vertical dashed lines show the locations of the means of the respective distributions.

In the left-hand figure, two density curves are shown. The red density curve shows the distribution of estimates $\hat{\rho}$ of ρ for randomly generated series when the true value is $\rho = 0.9$. As can be seen, the distribution is approximately normal. The blue density curve shows the distribution of $\hat{\rho}$ when $\rho = 1$. It has a high peak to the left of 1 and a heavy left tail.

The right-hand figure illustrates the situation that is expected to occur when searching for cointegrated pairs. The blue curve again represents the distribution of $\hat{\rho}$ when $\rho = 1$. The red curve represents a mixture distribution. In this distribution, 10% of the samples were generated from series where $\rho = 0.9$, while the remaining 90% of the samples were generated from series where $\rho = 1$. The heavy red tail to the left represents the area where samples are most likely to have come from mean-reverting processes. Thus, an ideal test statistic would clearly separate the cointegrated pairs in the red area from the random walks in the blue area.

Dickey and Fuller [14] were the first to characterize the distribution of $\hat{\rho}$ when $\rho = 1$ and to give confidence intervals for it. However, researchers realized relatively quickly that the (augmented) Dickey-Fuller test lacks power, which is to say that it often fails to reject the null

hypothesis when the null hypothesis is false. This is especially true when ρ is close to one. Many alternative tests have since appeared in the literature, a number of which have been implemented in the R statistical programming language.

Consequently, a comparison was made of various unit root tests available in R. The following table summarizes the unit root tests that were considered:

Test	Function	R Library	Description and Reference
ADF	`adf.test(X)`	`tseries[15]`	Augmented Dickey Fuller test [14]
PP	`pp.test(X)`	`tseries`	Phillips-Perron test [16]
JO-E	`ca.jo(X, type="eigen")`	`urca[17]`	Johansen's eigenvalue test [18]
JO-T	`ca.jo(X, type="trace")`	`urca`	Johansen's trace test [18]
ERS-P	`ur.ers(X, type="P-test")`	`urca`	Elliott, Rothenberg and Stock Feasible point optimal test [19]
ERS-D	`ur.ers(X, type="DF-GLS")`	`urca`	Elliott, Rothenberg and Stock Detrended ADF-type test [19]
SP-R	`ur.sp(X, type="rho")`	`urca`	ρ statistic of [20]
HURST	`aggvarFit(X)`	`fArma[21]`	Hurst exponent [22]
BVR	`bvr.test(X)`	`egcm[23]`	Breitung's variance ratio [24]
PGFF	`pgff.test(X)`	`egcm`	Weighted symmetric estimator [25]

Cointegration Tests Considered in This Study

The power of each test was calculated as a function of the length n of the sequence and the coefficient of mean reversion ρ. For each value of n and ρ, 1,000 random pairs of cointegrated sequences were generated having those parameters. Each cointegrated pair was generated according to the following specification, where α and β were each chosen uniformly at random from the interval [-10,10]:

$$\begin{aligned} X_t &= X_{t-1} + \varepsilon_t, & \varepsilon_t \sim N(0,1) \\ Y_t &= \alpha + \beta X_t + R_t & \\ R_t &= \rho R_{t-1} + \delta_t, & \delta_t \sim N(0,1) \end{aligned} \qquad (3)$$

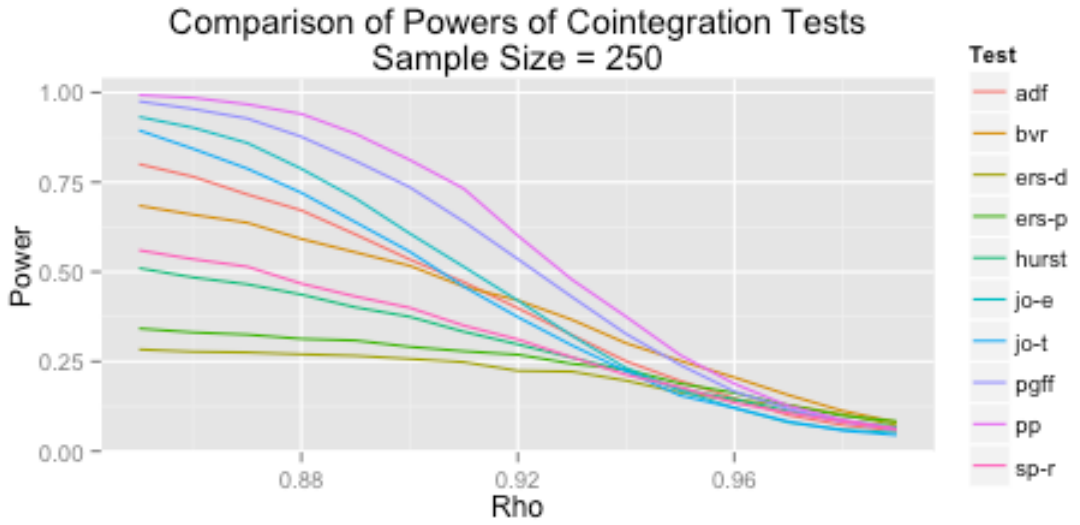
A fit was found to the randomly generated pair using the Engle Granger procedure, and the unit root test was applied to the residual series of the fitted model. (For the Johansen tests, the cointegration test was performed directly on the pair of random cointegrated vectors.) A record was then made of the percentage of cases that the test rejected the null hypothesis at $p = 0.05$. The results are given in the Appendix.

When the number of observations is less than 250, none of the tests performed especially well. A minimum of 250 observations is recommended for cointegration testing, and it would be ideal to have 500 observations or more.

Several of the tests outperform the augmented Dickey Fuller (ADF) test. These include PP, JO-E, JO-T, BVR ($n \leq 125$), and PGFF ($n \geq 250$). In some cases, the outperformance is dramatic. Perhaps the best overall performance was given by the Phillips-Perron test (PP).

In terms of non-parametric tests, Breitung's variance ratio (BVR) seemed to outperform the Hurst exponent as implemented by the R function `aggvarFit()`, however both of these tests were inferior to ADF. It should be noted that the package `fArma` contains several other methods for estimating the Hurst exponent, and these have not been included in this study. Also, the R package `vrtest` contains a number of other variance ratio tests, and these were not included either.

The graph below focuses on sample sizes of 250 (approximately one year of daily trading data). The best performing test was PP until about $\rho = 0.95$, after which BVR offered slightly better power.

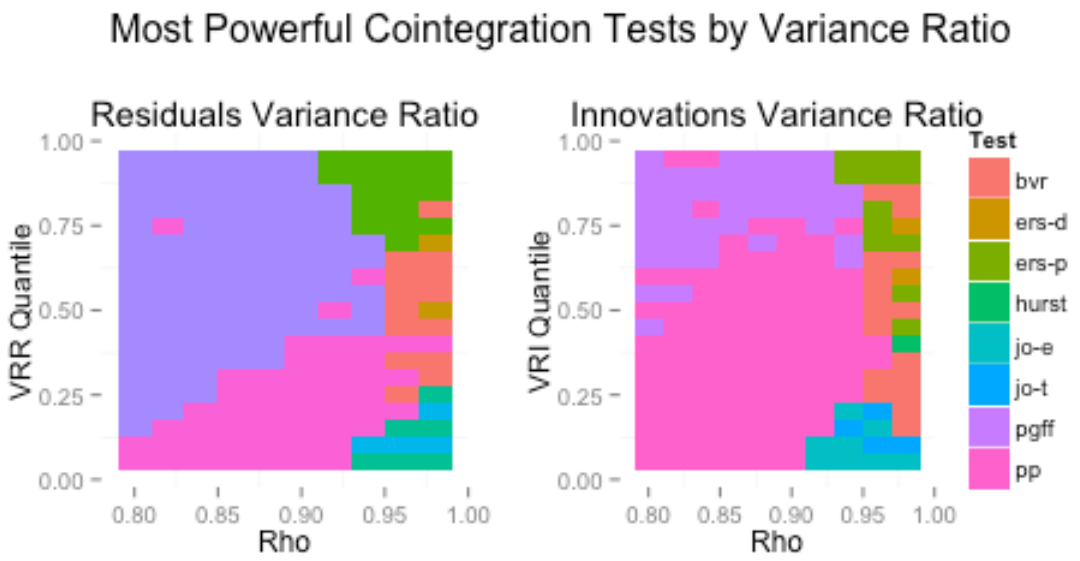


In the preceding discussion the variances of ε_t and δ_t where both fixed at 1. Subsequent investigations showed that if these variances are allowed to vary with respect to each other, then it can affect the power the unit root test. Using the notation of equation (3), the following variance ratios are defined:

$$VRR = \frac{\text{var}\,(R_t)}{\text{var}\,(\varepsilon_t)} = \frac{\text{var}\,(\delta_t)}{(1-\rho^2)\text{var}\,(\varepsilon_t)}$$
$$VRI = \frac{\text{var}\,(\delta_t)}{\text{var}\,(\varepsilon_t)}$$

Thus, the first quantity, VRR, is the variance ratio of the residuals, and the second quantity, VRI, is the variance ratio of the innovations. Reasonable ranges for these quantities were obtained by examining the distribution of values when performing the Engle-Granger procedure on pairs

of securities from the S&P 500 over the years 2002-2011. For various values of VRR (respectively VRI), and for various values of ρ, a number of random cointegrated pairs were generated and tested with respect to each of the unit root tests under study. The sequence length was fixed at $n = 250$. The unit root test having the highest power for that value of VRR (respectively VRI) and ρ was determined. The results obtained are shown in the following graphs.



As can be seen, the largest areas in each graph are from the Phillips-Perron test and the PGFF test, comprising more than 75% of the area in each graph. In the left graph, the PGFF test fares better, while in the right graph, the Phillips-Perron test fares better. When the variance ratio is at about the 75th percentile or above, and ρ is above about 0.9, then the tests of Elliott, Rothenberg and Stock perform well. When the variance ratio is at about the 20th percentile or below, and ρ is above 0.9, then the Johansen tests perform well. The ADF tests and the SP-R tests were never optimal. These results suggest that an improved cointegration test may be obtained by combining the various unit root tests in some way. The best method for doing so is left for future research.

DATA

The S&P 500 constituents list was downloaded from the Standard & Poor's web site on August 13, 2013. For each of these securities, the daily adjusted closing prices were downloaded from Yahoo for the period January 1, 2001 through December 31, 2012. For many of the securities, only a partial series was available.

For each year from 2002 – 2012, a dataset was constructed containing the daily adjusted closing prices for all of those stocks for which a complete set of observations was available in that year.

In other words, if a series contained any missing observations in a given year, it was not included in the dataset for that year. This gives rise to a set $S_1, S_2, \ldots, S_{N(y)}$, where $N(y)$ is the number of complete series that were available in year y. For each pair of stocks (S_i, S_j) with $1 \leq i < j \leq N(y)$, various statistics were tabulated with respect to that pair, such as the Dickey-Fuller statistic and the PGFF statistic. These are reported and discussed in the next section.

Two-year intervals were also considered. That is to say, for each year from 2001 – 2011, a dataset was constructed containing the daily adjusted closing prices for all of those stocks for which a complete set of observations was available for the two year period. Then, pairs were formed and statistics were computed on the pairs.

While this dataset is probably adequate for a preliminary analysis, it contains problems that would need to be addressed in a more comprehensive study. First, there is the issue of survivorship bias. The study only included securities that survived through August 13, 2013 and which were included as components of the S&P 500. Next, there is the issue of the completeness of the data. There were many days for which a closing quote was not available for a particular security. A possible approach might have been to interpolate the missing values, however this could possibly have introduced a new source of bias into the data, so the decision was made to simply exclude securities for which missing values were present in a given year.

Even given a perfectly complete data set, there are other factors that could potentially bias the results of a study of the persistence of cointegration. It is well known that the log returns of securities prices are not normally distributed, and this may affect tests for cointegration. Also, securities prices exhibit volatility clustering, and previous authors have reported that this can bias the results of a cointegration test [26]. In this study, results are obtained from computing various statistics on various populations of pairs, but each population of pairs is formed by drawing from a limited pool of securities. Therefore, statistics on the population of pairs will exhibit a dependence structure. These factors should be kept in mind when considering the results presented in the subsequent sections.

How Persistent Is Cointegration among US Equities?

An examination is now performed of securities prices over the period 2002 – 2012. For each year, the number of cointegrated pairs is estimated using an ADF test with $p = 0.05$. For each pair that was identified as cointegrated, a check was performed to see if it remained cointegrated in the following year, again using an ADF test with $p = 0.05$. In-sample estimates were used to compute the parameters of the cointegration relations. The results are reported in the following table.

	Formation Period			Trading Period		
	N(I1)	N(CI)	$\bar{\rho}$	N(I1)	N(CI)	$\bar{\rho}$
2002	12,561	649	0.91	507	22	0.94
		(5.17%)	(0.037)		(4.3%)	(0.033)
2003	66,795	3,540	0.93	3,007	142	0.93
		(5.30%)	(0.038)		(4.7%)	(0.035)
2004	88,410	4,588	0.94	4,240	183	0.93
		(5.2%)	(0.034)		(4.3%)	(0.036)
2005	93,961	3,685	0.93	3,426	118	0.93
		(3.9%)	(0.037)		(3.4%)	(0.033)
2006	97,903	3,850	0.94	3,595	142	0.92
		(3.9%)	(0.036)		(4.0%)	(0.046)
2007	102,378	3,507	0.92	3,439	253	0.90
		(3.4%)	(0.044)		(7.4%)	(0.041)
2008	108,811	7,897	0.91	5,627	363	0.92
		(7.3%)	(0.044)		(6.5%)	(0.039)
2009	78,210	5,756	0.92	5,662	253	0.94
		(7.4%)	(0.042)		(4.5%)	(0.032)
2010	113,050	4,882	0.93	4,636	233	0.92
		(4.3%)	(0.035)		(5.0%)	(0.036)
2011	107,880	4,802	0.93	4,565	208	0.93
		(4.5%)	(0.037)		(4.6%)	(0.030)

Number of Pairs Found to Exhibit Cointegration in the Formation Period and Trading Period for each year from 2002-2011 using ADF test.

The formation period is one year, and the trading period is the following year. The number of pairs for which each security is found to be integrated is listed in the column N(I1). Of these, the number of pairs found to be cointegrated is listed in the column N(CI). A pair is deemed to be cointegrated if the p-value of the ADF test is no more than 0.05. For each pair that is found to be cointegrated in the formation period, a check is performed to see if it remains cointegrated in the trading period. Thus, the right-hand panel counts those pairs that were found to be cointegrated in both periods. In-sample parameter estimates are used in both periods.

The table is divided into two panels. The left panel reports the statistics for pairs that were found to be cointegrated among the entire population of pairs in that year. The right panel reports statistics in the following year for those pairs that were identified as cointegrated in the first year.

The columns labeled N(I1) report the number of pairs for which both securities were found to be integrated, and the number N(CI) reports the number of pairs out of this set for which a cointegrating relation was found. The percentage of pairs found to be cointegrated is reported in parentheses below N(CI). The columns labeled $\bar{\rho}$ report the average coefficient of mean reversion among these pairs. The standard deviation is reported in parentheses below $\bar{\rho}$.

On average, 5.03% of pairs were found to be cointegrated in any given year, which is nearly identical to the amount that would be expected under the null hypothesis. The average

coefficient of mean reversion was found to be 0.925. Of those pairs that were found to be cointegrated in the formation period and for which both securities remained integrated in the trading period, 4.86% were found to be cointegrated in the trading period. The average coefficient of mean reversion was found to be 0.927.

The persistence of cointegration is now examined. It is hypothesized that the presence of cointegration in the formation year is predictive of the presence of cointegration in the trading year. In other words, let $CI(y)$ denote the set of pairs of securities (P_t, Q_t) such that both P_t and Q_t are $I(1)$ in year y and such that the pair (P_t, Q_t) is cointegrated in year y. If cointegration is persistent, then it is expected that

$$Pr\big(x \in CI(y+1) \big| x \in CI(y)\big) > Pr(x \in CI(y+1)).$$

On the other hand, if being cointegrated in year y is independent of being cointegrated in year $y+1$, then it is to be expected that these two quantities will be unrelated. A χ^2-test can be used to test the null hypothesis that being cointegrated in period y is independent of being cointegrated in period $y+1$.

Another approach to testing persistence is to examine the coefficient of mean reversion ρ. Let $\rho_{i,y}$ be the coefficient of mean reversion found for security i in year y. A regression is performed:

$$\rho_{i,y+1} = a_0 + a_1 \rho_{i,y}.$$

The null hypothesis is that there is no predictive association between $\rho_{i,y+1}$ and $\rho_{i,y}$. If a_1 is found to be positive and significantly different from zero, then this hypothesis is rejected.

The following table reports the results that were obtained:

y	$\Pr(x \in CI(y+1) \mid x \in CI(y))$	$Pr(x \in CI(y+1))$	$\hat{a}_1$	$R^2(a_1)$
2002	3.39%	3.64%	-0.0127	0.0002
2003	4.01%	4.70% *	0.0380 ***	0.0016
2004	3.99%	3.63%	0.0529 ***	0.0028
2005	3.20%	3.63%	-0.0160 ***	0.0003
2006	3.69%	3.21%	0.0102 **	0.0001
2007	7.21%	7.07%	0.0933 ***	0.0061
2008	4.60%	5.05% *	0.1340 ***	0.0230
2009	4.40%	4.23%	0.0227 ***	0.0010
2010	4.77%	4.09% **	0.0082 **	0.0001
2011	4.33%	3.80% *	0.0053 **	0.0000

Persistence of cointegration as measured using the ADF statistic.
The first column measures the probability that a pair would be found to be cointegrated in year two given that it was found to be cointegrated in year one. The second column measures the unconditional probability that a pair would be found to be cointegrated in year two. A χ^2-test of goodness of fit is used to test whether the distribution from column one is different from that in column two. Significance levels are 0.10 (*), 0.05(**) and 0.01(***). The third column reports the regression coefficient when regressing the coefficient of mean reversion from year two on that of year one. The fourth column reports the proportion of explained variance (R^2) associated with the regression.

In six of the ten years, the χ^2-test was unable to reject the null hypothesis that the conditional probability of being cointegrated in year two is the same as the unconditional probability. In the four years where the null hypothesis could be rejected at the 10% level, there were two years in which the unconditional probability was higher the conditional probability and two years in which it was lower. Thus, there appears to be little evidence in favor of the hypothesis that being cointegrated in a given year is predictive of being cointegrated in the following year, at least when the augmented Dickey-Fuller test is used.

In all of the years except one, the coefficient a_1 was found to be significantly different from zero. However, in two of the ten years, a negative correlation was found. And in all years except one, the proportion of explained variance R^2 was below 1%. This is a typical situation when working with large data sets when no actual relation is present in the data. Although the coefficients of the linear model are reported as being significant, the signs fluctuate and the amount of explained variance is negligible. Thus, the evidence provided by the linear fit also fails to suggest that being cointegrated in a given year is predictive of being cointegrated in the following year.

Note that the percentages of cointegrated pairs in this table are different from and generally lower than the percentages reported in the previous table. This is because the denominators are different. In the latter table, the percentage is calculated in proportion to all pairs, not just those pairs where both series are I(1).

It is possible that the negative results obtained so far reflect the low power of the ADF test. In the simulation studies, the Phillips-Perron (PP) test seemed to perform quite well. Therefore, the above analysis was repeated using the PP test in place of the ADF test. The following table summarizes the results that were obtained:

	Formation Period			Trading Period		
	N(I1)	**N(CI)**	$\overline{\rho}$	**N(I1)**	**N(CI)**	$\overline{\rho}$
2002	11,935	825	0.91	789	28	0.90
		(6.9%)	(0.036)		(3.6%)	(0.039)
2003	83,028	3,035	0.91	2,558	108	0.92
		(3.7%)	(0.038)		(4.2%)	(0.040)
2004	84,255	3,143	0.92	2,461	79	0.92
		(3.7%)	(0.038)		(3.2%)	(0.037)
2005	87,153	2,715	0.91	2,385	76	0.92
		(3.1%)	(0.039)		(3.2%)	(0.038)
2006	93,961	3,065	0.92	2,401	173	0.91
		(3.3%)	(0.038)		(7.2%)	(0.042)
2007	85,491	4,609	0.91	4,030	468	0.90
		(5.4%)	(0.040)		(11.6%)	(0.039)
2008	100,128	11,723	0.89	8,428	681	0.89
		(11.7%)	(0.041)		(8.1%)	(0.034)
2009	77,421	5,413	0.90	5,059	213	0.92
		(7.0%)	(0.036)		(4.2%)	(0.033)
2010	104,653	3,479	0.92	2,807	140	0.91
		(3.3%)	(0.037)		(5.0%)	(0.036)
2011	91,806	4,452	0.91	3,817	136	0.91
		(4.9%)	(0.036)		(3.6%)	(0.035)

Number of Pairs Found to Exhibit Cointegration in the Formation Period and Trading Period for each year from 2002-2011 using the Phillips-Perron test.
The formation period is one year, and the trading period is the following year. The number of pairs for which each security is found to be integrated is listed in the column N(I1). Of these, the number of pairs found to be cointegrated is listed in the column N(CI). A pair is deemed to be cointegrated if the p-value of the PP test is no more than 0.05. For each pair that is found to be cointegrated in the formation period, a check is performed to see if it remains cointegrated in the trading period. Thus, the right-hand panel counts those pairs that were found to be cointegrated in both periods. In-sample parameter estimates are used in both periods.

The average rate of cointegration was 5.29%, as opposed to 5.03% when the ADF test was used. The persistence of cointegration is again examined. The results are reported in the following table.

y	$\mathbf{Pr}\,(x \in CI(y+1) \mid x \in CI(y))$	$Pr(x \in CI(y+1))$	
2002	3.39%	3.12%	
2003	3.56%	3.20%	
2004	2.51%	2.68%	
2005	2.80%	2.89%	
2006	5.64%	4.22%	***
2007	10.15%	10.50%	
2008	5.81%	4.75%	***
2009	3.93%	3.01%	***
2010	4.02%	3.79%	
2011	3.05%	2.90%	

Persistence of cointegration as measured using the PP statistic. The first column measures the probability that a pair would be found to be cointegrated in year two given that it was found to be cointegrated in year one. The second column measures the unconditional probability that a pair would be found to be cointegrated in year two. A χ^2-test of goodness of fit is used to test whether the distribution from column one is different from that in column two. Significance levels are 0.10 (*), 0.05(**) and 0.01(***).

In three of the ten years, the conditional probabilities are found to be lower than the unconditional probabilities. In those years where the conditional probability is higher, the difference is only found to be statistically significant in three of them. The average of the conditional probabilities is 4.48%, while the average of the unconditional probabilities is 4.11%. The evidence for persistence of cointegration seems at best to be weak.

A number of variations of these experiments were tried. In place of the PP test, the JO-E, ERS-P, BVR and PGFF tests were tried. Tests were performed where the p-value for accepting a pair as cointegrated was set to 0.01. The locfdr package of [27] and the fdrtool package of [28] were used to try to determine a cutoff based upon false discovery rates, and tests were performed based upon these suggested cutoff values. Tests were performed on the unlogged price series, and two-year intervals (both logged and unlogged) were examined as well. The results were not qualitatively different from what has been presented so far[4].

Of all of the different permutations that were tried, the most tantalizing results were obtained with the PGFF test at $p = 0.05$:

[4] Details available from the author

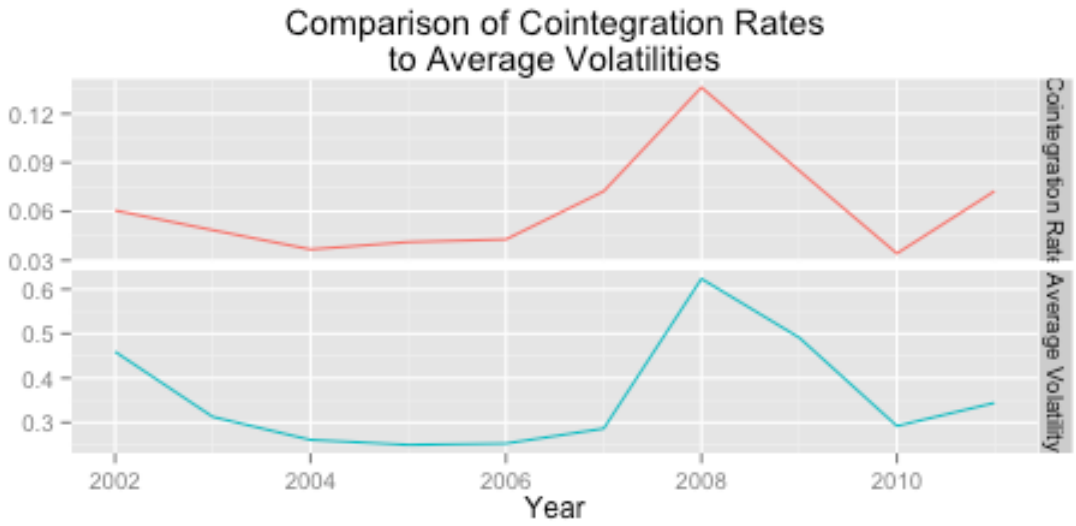
y	$\Pr(x \in CI(y+1) \mid x \in CI(y))$	$Pr(x \in CI(y+1))$	
2002	4.56%	4.85%	
2003	4.32%	3.67%	**
2004	5.64%	4.12%	***
2005	5.91%	4.28%	***
2006	8.04%	7.24%	**
2007	17.40%	13.61%	***
2008	10.01%	8.50%	***
2009	4.24%	3.42%	***
2010	7.71%	7.25%	
2011	3.87%	3.40%	**

Persistence of cointegration as measured using the PGFF statistic. The first column measures the probability that a pair would be found to be cointegrated in year two given that it was found to be cointegrated in year one. The second column measures the unconditional probability that a pair would be found to be cointegrated in year two. A χ^2-test of goodness of fit is used to test whether the distribution from column one is different from that in column two. Significance levels are 0.10 (*), 0.05(**) and 0.01(***).

In this experiment, the conditional probability of cointegration was greater than the unconditional probability in all years except one. In eight of the ten years, the difference was found to be statistically significant.

Assume for the moment that this is not just the result of data mining, and consider a hypothetical investor who chooses 100 pairs in which to invest. The mean rate of cointegration in the trading period among the unconditional data was 6.0%, while the mean rate of cointegration in the trading period among the conditional data was 7.2%. Thus, if the pairs were chosen by the investor at random, then on average, the performance of 94 of those pairs would have been consistent with a random walk, and the remaining 6 would have behaved as if they were cointegrated. On the other hand, if the investor had chosen the portfolio of pairs by first screening for evidence of cointegration using the PGFF test, then on average, the performance of 93 of those pairs would have been consistent with a random walk, and the remaining 7 would have behaved as if they were cointegrated.

One interesting feature of the data is that higher than expected numbers of cointegrated pairs are found in the formation periods, and these numbers reach especially large values in the years surrounding the financial crisis. One possible explanation can be found from examining the following graph.



This graph compares the cointegration rate using the PP test for each year to the average volatility of the stocks in the dataset in that year. As can be seen, the two curves are very similar. In fact, a linear regression of the cointegration rates on the average volatilities yields an adjusted R^2 value of 0.75. It may be the case that cointegration tests are affected by changes in volatility. This would be consistent with the findings of [26], who examined how cointegration tests can be impacted by GARCH effects.

Is There Short-Term Persistence of Cointegration?

In the preceding section, it was found that if a pair is identified as cointegrated in a given year, there is little basis to believe that the pair will again be identified as cointegrated in the following year. This leaves open the question, "How quickly does the cointegration relationship fall apart?" Perhaps cointegration is a transient property that dissipates rapidly, say within a few months.

In order to answer this question, it is necessary to have a method for relatively quickly identifying when the cointegration relationship ceases to hold. Recall that the original specification of the model is as follows:

$$\begin{aligned} \log P_t &= \alpha + \beta \log Q_t + R_t, \\ R_t &= \rho R_{t-1} + \varepsilon_t. \end{aligned} \qquad (4)$$

Under this specification, it is assumed that the innovations ε_t are independent and identically distributed with mean zero.

Suppose that price series P_t and Q_t have been given and that values for α, β and ρ have been determined. From this data, it is possible to determine the corresponding sequence of innovations ε_t as follows:

$$R_t = \log P_t - \alpha - \beta \log Q_t \qquad (5)$$
$$\varepsilon_t = R_t - \rho R_{t-1}.$$

If the cointegration model holds and the relations have been specified correctly, then the values of ε_t determined in this way will be independent and identically distributed with mean zero. Therefore, a test can be formulated to check this assumption. If the test fails in a given period, it is taken as evidence that the cointegration relationship no longer holds.

This approach depends critically upon the assumption that the parameters α, β and ρ have been estimated accurately. Unfortunately, this gives rise to some complications. One complication concerns estimation bias. Engle and Granger established that the parameter estimates obtained from the two-step procedure are consistent. Although the estimates for α and β will be unbiased, the estimate for ρ will not be. A second complication stems from the fact that the parameter estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\rho}$ will inherently contain some estimation error, and this estimation error will affect the estimations of the series of innovations $\{\hat{\varepsilon}_t\}$. Thus, a test such as the student's t-test would not be accurate.

It might be possible to reformulate the t-test in a way that overcomes these problems, however the approach taken here is to perform a qualitative comparison. Two different hypotheses, A and B, are formulated. The question is then posed, "Does the data more close resemble hypothesis A or hypothesis B?" No critical values are given. It is left to the reader to judge the strength of the evidence for the two hypotheses.

The following test procedure was implemented. For each of the ten years under study, a random sample of 1,000 pairs was chosen from among those identified as being cointegrated when using the Phillips-Perron test at confidence level $p = 0.05$. A bias correction was computed for $\hat{\rho}$, and the corresponding sequence of innovations ε_t was computed for the following year.

For each 20-day period within the trading period, the mean μ_i and standard deviation σ_i was formed from the innovations ε_t occurring within that month. From this value, a Z-score was computed as

$$Z_i = \frac{\mu_i}{\sigma_i/\sqrt{n}}$$

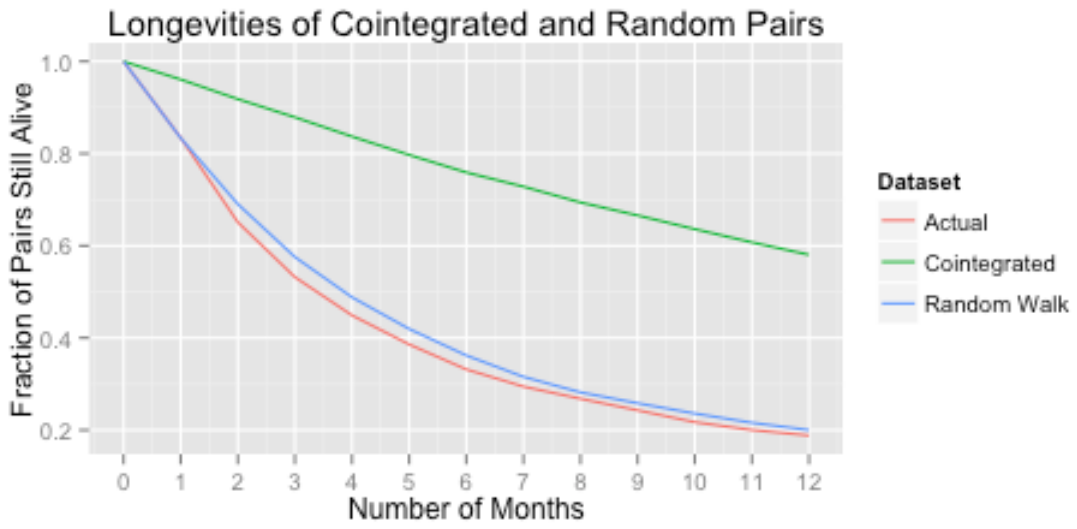
If the absolute value of the Z-score exceeded 2.5, then the pair was deemed to have died in month i. Otherwise, it was deemed to be still alive.

The *longevity* of the pair was then defined as the maximum number of consecutive months that it remained alive, starting from $i = 1$. By performing this procedure on each of the 10,000 randomly chosen pairs, a distribution of longevities was obtained.

Longevity distributions were obtained analogously for the two comparison groups. The first comparison group was formed from random pairs of cointegrated series. That is to say, random values for α, β and ρ were chosen, and a pair of cointegrated series of length 490 was generated according to these parameters. Then, parameter estimates $\hat{\alpha}, \hat{\beta}$ and $\hat{\rho}$ were computed from the first 250 observations in each series. Based on these parameter estimates, the longevity of the pair was computed from the latter 240 observations. This procedure was performed 10,000 times. This is referred to as the Cointegrated data set.

The second comparison group was formed from pairs of random walks. That is to say, two independent unit root series were constructed of length 490, subject to the constraint that the Engle Granger procedure identified the first 250 elements of the series as being cointegrated. Parameter estimates $\hat{\alpha}, \hat{\beta}$ and $\hat{\rho}$ were computed from the first 250 observations in each series. Based on these parameter estimates, the longevity of the pair was computed from the latter 240 observations. This procedure was performed 10,000 times. This is referred to as the Random Walk data set.

Thus, the question that is posed is, "Does the longevity curve of the actual data more closely resemble that of cointegrated pairs or pairs of random walks?" The following graph depicts the results that were obtained.



The longevity curve for the actual series is nearly identical to that of the Random Walk data set. If cointegration were a short-term phenomenon that persisted only for a few months and then

decayed, it would be expected that the longevity curve for the actual data would more closely resemble that of the Cointegrated data set, or at least that it would be significantly above the longevity curve for the Random Walk dataset in the first few months. Thus, the evidence does not support the hypothesis that cointegration is a short-term phenomenon that dissipates over time.

CONCLUSIONS

This paper has considered the question of whether or not cointegration is a persistent property. In the case of pairs of U.S. equities belonging to the S&P 500 during the period 2002-2012, it appears that the answer is no. If two securities appear to be cointegrated in one year, no useful information is gained as to whether or not they would have been cointegrated in the following year.

In the end, cointegration may not be the most suitable tool for identifying candidate pairs for trading. What are the chances that two randomly chosen time series will be cointegrated? In the case of securities, a necessary condition that two securities series be cointegrated is that the factor loadings of one security must be an exact multiple of the factor loadings of the other. This is already a restrictive condition. Even if this condition is satisfied, though, it is not enough to guarantee that the two series will be cointegrated. An additional requirement is that the idiosyncratic risk components of both securities must be free of unit roots. In other words, all firm-specific news must have no more than a temporary effect on the firm's security price. From this vantage point, using pairs to hedge away unit roots seems unrealistic.

Despite the negative results presented here, a number of authors have reported that cointegration-based pairs trading yields abnormal positive returns. What explains this apparent contradiction? Some authors have tested pairs trading with different markets and universes of securities. For example, [3] consider pairs trading in the Brazilian stock market, and [12] considers pairs trading applied to ETF's. It is possible that these markets have different characteristics than the U.S. equities market, resulting in significant numbers of identifiable cointegrated pairs. For example, numerous pairs of ETF's can be found where both securities in the pair track the same underlying index or commodity price. These will necessarily be cointegrated.

In the course of the analysis presented here, several other observations were made. A problem in identifying cointegrated series is that unit root tests tend to have low power. Among the unit root tests that were tried, the Phillips-Perron test seemed to perform well. It was found that the power of the cointegration test is dependent upon the ratio of the variance of the residuals to the variance of the original series.

Any pairs trading strategy that makes use of data mining to identify cointegrated pairs is likely to generate large numbers of false positives. Therefore, the pairs trading algorithm should be designed to perform reasonably for securities that are not actually cointegrated. Where possible, an attempt should be made to quantify the false positive rate.

Finally, estimates of α and β will be consistent and unbiased, however estimates of ρ can exhibit significant bias for smaller sample sizes. It is advisable to incorporate a bias correction for the estimate of ρ.

An R package has been submitted containing code for performing the BVR and PGFF unit root tests and for fitting cointegration models using the two-step Engle-Granger procedure. The name of the package is "`egcm`".

BIBLIOGRAPHY

[1] R. F. Engle and C. W. Granger, "Co-integration and error correction: representation, estimation, and testing," *Econometrica*, pp. 251–276, 1987.

[2] M. Avellaneda and J.-H. Lee, "Statistical arbitrage in the US equities market," *Quant. Finance*, vol. 10, no. 7, pp. 761–782, 2010.

[3] J. Caldeira and G. Moura, "Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy," *Rev. Braz. Finance*, vol. 11, no. 1, pp. 49–80, 2013.

[4] B. Do, R. Faff, and K. Hamza, "A new approach to modeling and estimation for pairs trading," in *Proceedings of 2006 Financial Management Association European Conference*, 2006.

[5] A. Galenko, E. Popova, and I. Popova, "Trading in the Presence of Cointegration," *J. Altern. Investments*, vol. 15, no. 1, pp. 85–97, Summer 2012.

[6] G. Vidyamurthy, *Pairs trading quantitative methods and analysis*. Hoboken, N.J.: J. Wiley, 2004.

[7] D. S. Ehrman, *The handbook of pairs trading: strategies using equities, options, and futures*. John Wiley & Sons, 2006.

[8] M. Whistler, *Trading Pairs: capturing profits and hedging risk with statistical arbitrage strategies*. John Wiley & Sons, 2004.

[9] M. Harlacher, "Cointegration Based Statistical Arbitrage," Department of Mathematics, Swiss Federal Institute of Technology, Zurich,Switzerland, 2012.

[10] H. Puspaningrum, "Pairs trading using cointegration approach," School of Mathematics and Applied Statistics, Univeristy of Wollongong, Wollongong, Australia, 2012.

[11] A. D. Schmidt, "Pairs trading: a cointegration approach," University of Sydney, Sydney, Australia, 2008.

[12] M. Sipilä, "Algorithmic Pairs Trading: Empirical Investigation of Exchange Traded Funds," Department of Finance, Aalto University, Otaniemi, Finland, 2013.

[13] M. Yakop, "A comparative analysis of pairs trading," Faculty of Economics and Business, University of Amsterdam, Amsterdam, The Netherlands, 2011.

[14] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *J. Am. Stat. Assoc.*, vol. 74, no. 366a, pp. 427–431, 1979.

[15] A. Trapletti and K. Hornik, *tseries: Time Series Analysis and Computational Finance. R package version 0.10-32*. 2013.

[16] P. C. Phillips and P. Perron, "Testing for a unit root in time series regression," *Biometrika*, vol. 75, no. 2, pp. 335–346, 1988.

[17] B. Pfaff, *Analysis of integrated and cointegrated time series with R*, Second. Springer, 2008.

[18] S. Johansen and K. Juselius, "Maximum likelihood estimation and inference on cointegration—with applications to the demand for money," *Oxf. Bull. Econ. Stat.*, vol. 52, no. 2, pp. 169–210, 1990.

[19] B. E. Elliot, T. J. Rothenberg, and J. H. Stock, "Efficient tests of the unit root hypothesis," *Econometrica*, vol. 64, no. 8, pp. 13–36, 1996.

[20] P. Schmidt and P. C. Phillips, "LM Tests for a Unit Root in the Presence of Deterministic Trends," *Oxf. Bull. Econ. Stat.*, vol. 54, no. 3, pp. 257–287, 1992.

[21] D. Wurtz and many others and see the SOURCE file, *fArma: ARMA Time Series Modelling. R package*. 2013.
[22] M. S. Taqqu, V. Teverovsky, and W. Willinger, “Estimators for long-range dependence: an empirical study,” *Fractals*, vol. 3, no. 04, pp. 785–798, 1995.
[23] M. Clegg, *egcm: Engle-Granger cointegration models. R package*. 2014.
[24] J. Breitung, “Nonparametric tests for unit roots and cointegration,” *J. Econ.*, vol. 108, no. 2, pp. 343–363, 2002.
[25] S. G. Pantula, G. Gonzalez-Farias, and W. A. Fuller, “A comparison of unit-root test criteria,” *J. Bus. Econ. Stat.*, vol. 12, no. 4, pp. 449–459, 1994.
[26] T.-H. Lee and Y. Tse, “Cointegration tests with conditional heteroskedasticity,” *J. Econ.*, vol. 73, no. 2, pp. 401–410, Aug. 1996.
[27] B. Efron, *Local false discovery rates*. Division of Biostatistics, Stanford University, 2005.
[28] K. Strimmer, “fdrtool: a versatile R package for estimating local and tail area-based false discovery rates,” *Bioinformatics*, vol. 24, no. 12, pp. 1461–1462, Jun. 2008.

Appendix: Comparison of Cointegration Tests

N	$\rho =$	0.80	0.90	0.92	0.94	0.96	0.98
60	ADF	0.17	0.08	**0.08**	0.06	0.05	0.05
	PP	**0.22**	0.09	**0.08**	0.06	**0.06**	0.06
	JO-E	0.17	0.07	0.07	0.06	**0.06**	0.05
	JO-T	0.18	0.08	0.07	0.06	0.05	0.04
	ERS-P	0.06	0.04	0.04	0.04	0.04	0.04
	ERS-D	0.07	0.06	0.06	0.04	**0.06**	0.06
	SP-R	0.09	0.07	0.05	0.05	0.05	0.05
	HURST	0.07	0.04	0.03	0.02	0.02	0.02
	BVR	**0.22**	**0.10**	**0.08**	**0.07**	**0.06**	**0.07**
	PGFF	0.09	0.04	0.03	0.03	0.02	0.03
125	ADF	0.50	0.19	0.15	0.10	0.08	0.06
	PP	**0.84**	**0.25**	0.16	0.12	0.08	**0.07**
	JO-E	0.61	0.17	0.12	0.08	0.06	0.04
	JO-T	0.56	0.18	0.13	0.09	0.06	0.04
	ERS-P	0.21	0.11	0.10	0.08	0.06	0.06
	ERS-D	0.19	0.12	0.11	0.10	0.07	**0.07**
	SP-R	0.32	0.16	0.15	0.12	0.09	0.08
	HURST	0.27	0.14	0.10	0.08	0.06	0.04
	BVR	0.50	**0.25**	**0.17**	**0.14**	**0.10**	**0.07**
	PGFF	0.48	0.10	0.06	0.04	0.03	0.03
250	ADF	0.92	0.54	0.40	0.26	0.16	0.08
	PP	**1.00**	**0.82**	**0.61**	**0.36**	0.18	0.09
	JO-E	**1.00**	0.62	0.41	0.23	0.12	0.06
	JO-T	0.99	0.56	0.39	0.22	0.12	0.06
	ERS-P	0.37	0.30	0.26	0.22	0.16	0.10
	ERS-D	0.28	0.26	0.23	0.20	0.14	0.10
	SP-R	0.65	0.39	0.31	0.21	0.18	0.08
	HURST	0.60	0.39	0.30	0.22	0.14	0.08
	BVR	0.80	0.53	0.43	0.30	**0.21**	**0.11**
	PGFF	**1.00**	0.73	0.54	0.33	0.18	0.08
500	ADF	**1.00**	0.98	0.93	0.78	0.48	0.16
	PP	**1.00**	**1.00**	**1.00**	**0.95**	**0.63**	0.20
	JO-E	**1.00**	**1.00**	0.97	0.77	0.38	0.11
	JO-T	**1.00**	0.99	0.92	0.71	0.35	0.10
	ERS-P	0.54	0.45	0.44	0.40	0.34	**0.21**
	ERS-D	0.42	0.41	0.39	0.37	0.33	0.21
	SP-R	0.95	0.81	0.74	0.60	0.38	0.17
	HURST	0.90	0.72	0.64	0.52	0.35	0.17
	BVR	0.95	0.80	0.71	0.60	0.43	**0.21**
	PGFF	**1.00**	**1.00**	**1.00**	0.94	0.62	**0.21**
1000	ADF	**1.00**	**1.00**	**1.00**	**1.00**	0.96	0.49
	PP	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.59
	JO-E	**1.00**	**1.00**	**1.00**	**1.00**	0.96	0.36
	JO-T	**1.00**	**1.00**	**1.00**	**1.00**	0.94	0.36
	ERS-P	0.77	0.66	0.65	0.61	0.58	0.47
	ERS-D	0.58	0.57	0.58	0.57	0.54	0.44
	SP-R	**1.00**	0.99	0.98	0.95	0.83	0.46
	HURST	0.99	0.95	0.91	0.81	0.65	0.31

	BVR	0.99	0.95	0.92	0.84	0.72	0.45
	PGFF	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**0.66**

Comparison of the Powers of Various Cointegration Tests.

Each column represents a different value of the mean reversion coefficient $\boldsymbol{\rho}$, and each group of rows represents the length of the vector that is tested for cointegration. For each cell, 4,000 random cointegrated pairs were generated with the specified length and mean reversion parameter, and the frequency of rejection of the unit root hypothesis was recorded. The best value within each group is shown in bold. The mnemonics for the various unit root tests are as follows: ADF = augmented Dickey-Fuller test, PP = Phillips-Perron test, JO-E = Johansen's eigenvalue test, JO-T Johansen's trace test, ERS-P = Elliott, Rothenberg and Stock point optimal test, ERS-D = Elliott, Rothenberg and Stock detrended ADF-style test, SP-R = Schmidt-Phillips $\boldsymbol{\rho}$ statistic, HURST = Hurst exponent calculated through aggregated variance method, BVR = Breitung's variance ratio, PGFF = Pantula, Gonzalez-Farias and Fuller weighted symmetric estimator.