ROBUST TESTS FOR COINTEGRATION WITH APPLICATION TO STATISTICAL ARBITRAGE TRADING STRATEGIES

A Paper
Submitted to the Graduate Faculty
Of the
North Dakota State University
of Agriculture and Applied Science

By

Thomas Alan Hanson

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Statistics

December 2010

Fargo, North Dakota

North Dakota State University
Graduate School

Title

A Robust Test for Cointegration with Application to

Statistical Arbitrage Trading Strategies

By

Thomas Alan Hanson

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

ABSTRACT

Hanson, Thomas Alan, M.S., Department of Statistics, College of Science and Mathematics, North Dakota State University, December 2010. Robust Tests for Cointegration with Application to Statistical Arbitrage Trading Strategies. Major Professor: Dr. Rhonda Magel.

This study proposes two new cointegration tests that employ rank-based and least absolute deviation techniques to create a robust version of the Engle-Granger cointegration test. Critical values are generated through a Monte Carlo simulation over a range of error distributions, and the performance of the tests is then compared against the Engle-Granger and Johansen tests. The robust procedures underperform slightly for normally distributed error terms but outperform for fatter-tailed distributions. This characteristic suggests the robust tests are more appropriate for many applications where departures from normality are common. One particular example discussed here is statistical arbitrage, a stock trading strategy based on cointegration and mean reversion. In a simple example, the rank-based procedure produces additional profits over the Engle-Granger procedure.

ACKNOWLEDGEMENTS

Writing this paper has been a process of frustration and fascination for me. It has provided an opportunity to combine my interests in statistics and finance and to hone my computer simulation skills, all of which I have enjoyed, even as I struggled. I would never have completed it or endured the challenges, however, were it not for the help of many wonderful people, some of whom I would like to acknowledge here.

I must begin by thanking Dr. Jeff Terpstra for his guidance in shaping early drafts of this work. Particular thanks must also go to Dr. Rhonda Magel for her assistance in completing the work, and for her patience and flexibility to work via email to finish the paper. I also thank my other professors during my time at NDSU as well as my committee.

I would also like to acknowledge the superb education and direction I was fortunate to receive from a number of inspiring and dedicated educators at previous institutions: Fulton Fahrner, Karen Molitor, and Marit Elliott at Melrose High School; Drs. Kathy Chadwick, Paul Wojick, and Nancy Paddleford at St. Olaf College; and Dr. Jeanne Buckeye at the University of St. Thomas. I would be a different person in a different place today were it not for the guidance I received from each of them.

During my time in Fargo, I enjoyed many hours of duplicate bridge with the members of the Fargo Bridge Club and the Fargo Country Club. Their memberships are composed of wonderfully warm and intelligent people, whom I would like to thank for providing a mentally stimulating and enjoyable distraction from the rigors of academia and the cold winters of North Dakota.

For many years, John Newstrom was a fantastic roommate and friend, and I appreciate the many laughs and adventures together. I only hope there are years of exploits still ahead of us, wherever else life takes us.

My family has been a constant source of support and inspiration throughout my life. No matter the choices I made, through challenges and victories, accomplishments and setbacks, I knew that I could always count on my parents Mary and Al Hanson to be there. I have also been blessed to have the encouragement of my grandparents, extended family, and my fellow runners and educators, Katie and Steve Lyman.

Last, I need to thank Jeff Child, who endured countless conversations about this paper and buoyed my spirits when I was most frustrated. He will no doubt celebrate the completion of this project nearly as much as I do. Throughout the writing process, he has been my inspiration and has provided more support and love than I can possibly thank him for here.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

Cointegration can be conceptually described as a long-run equilibrium relationship between two or more nonstationary time series. When considered independently, the time series each appear to wander in an unpredictable random walk, but there exists some linear combination of those same variables that is stationary and therefore follows a pattern of mean-reversion. This concept has widespread applications, particularly in economics and finance, in which theories predict stable equilibria among nonstationary time series such as bid and ask quotes of a given stock over time and across different markets.

To help fix this idea, consider the graphs in Figure A. The top row consists of time series plots of two variables, x and y, each constructed by the cumulative sum of 100 independent standard normal random variables. Both series were transposed vertically so that the first term is zero. Though they begin at the same point, these variables are otherwise completely independent of each other and are nonstationary. The third plot, a scatterplot of y versus x, makes the random and independent nature of their relationship visually clear. There is no evident relationship between the two time series. Finally, the fourth plot is a time series plot of the difference of these two variables. If the two series were cointegrated, this difference would be stationary and demonstrate mean reversion, which would manifest itself in multiple crossings of the x-axis. In this case, however, the difference series appears qualitatively similar to the random walks of each time series.

By contrast, Figure B shows the analogous four plots for two cointegrated series, c and d. As before, the data for time series c were generated by the cumulative sum of 100 independent standard random normal variables. However, in this case, the second series

was set equal to the first, plus a random disturbance term drawn from a standard normal distribution. The two series were once again transposed vertically to begin at the origin.

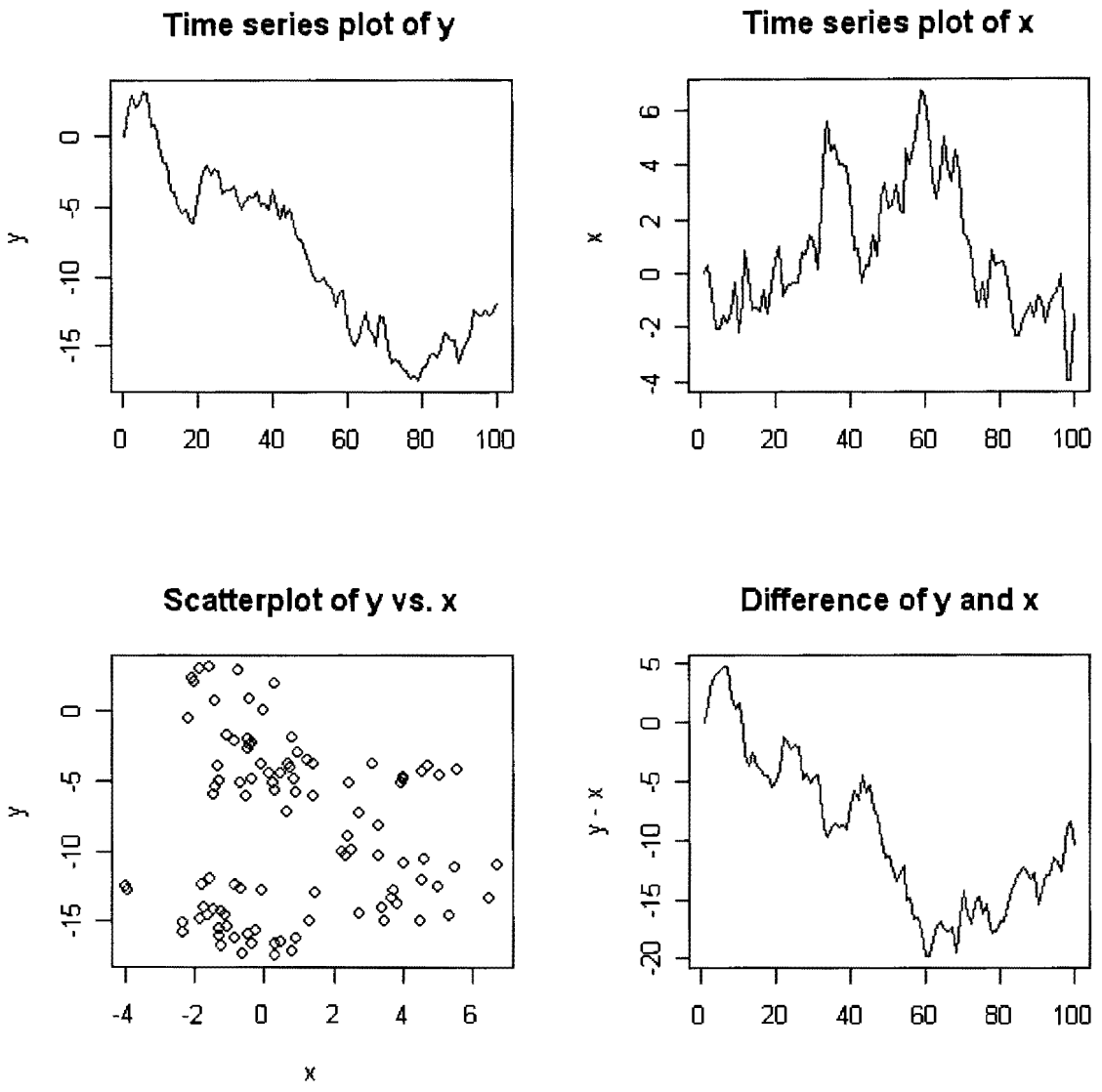

Figure A. Non-cointegrated time series

The cointegrated relationship between *c* and *d* is evident in all four graphs. The first two plots demonstrate that the series tend to vary contemporaneously; their local minima and maxima occur at approximately equal points. The scatterplot shows a strong linear relationship between the two variables, in stark contrast to the random scatterplot of Figure A. Finally, the difference series appears to be stationary; the mean and variance

both appear to be approximately constant. The fourth graph also exhibits strong mean-reversion with its multiple crossings of the x-axis.

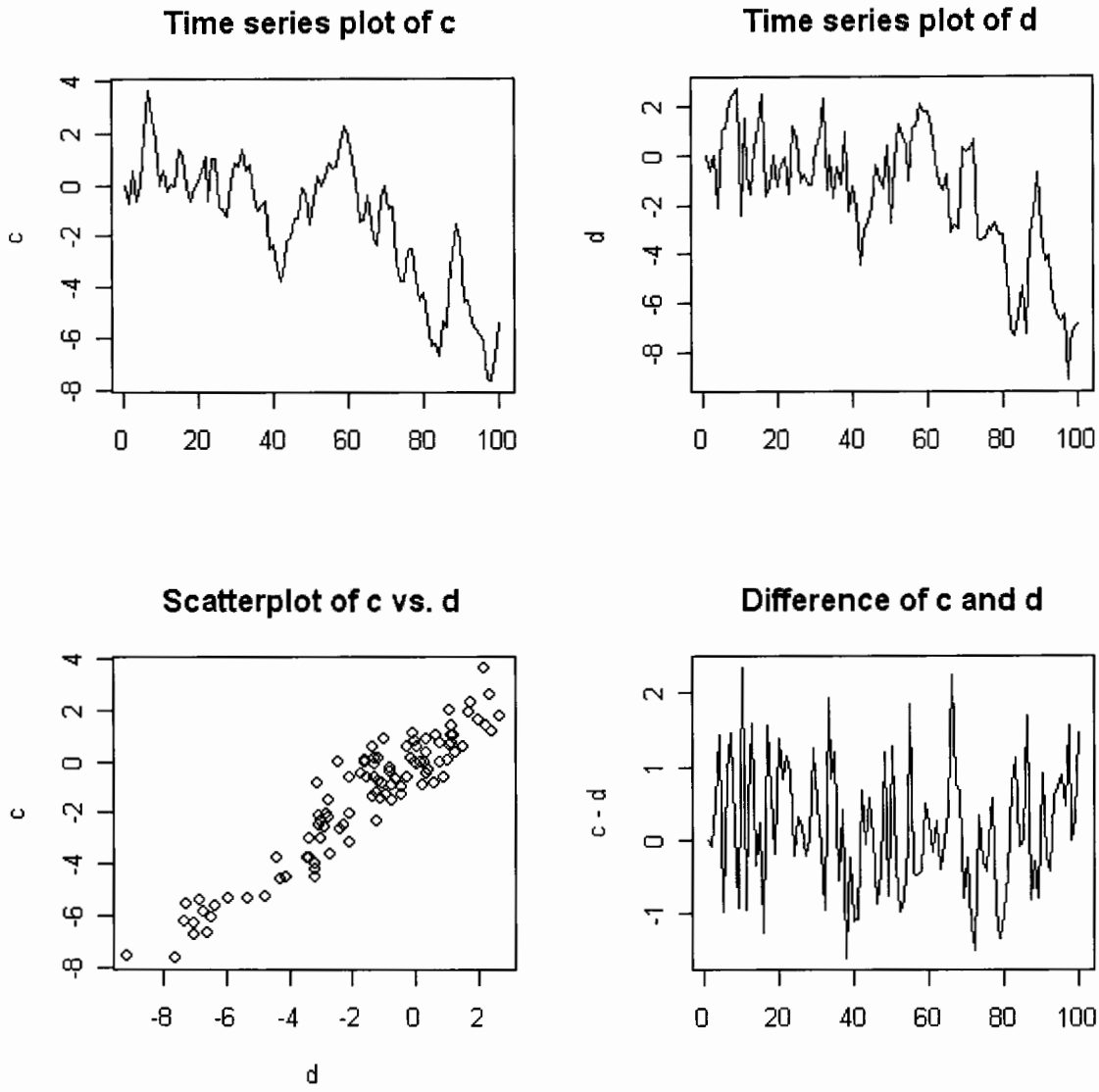

Figure B. Cointegrated time series

This is an illustration of bivariate cointegration, in which the linear combination of two nonstationary time series is stationary. While multivariate cointegration is well-defined, this investigation is limited to the bivariate case. The next section provides a theoretical overview of the motivation for this study.

Theoretical Underpinnings of the Study

The purpose of this paper is to propose robust cointegration tests based on the seminal work of Engle and Granger (1987), who provided the first cointegration test that serves as the initial baseline model for this paper. Their original test, in turn, relies on the work of Dickey and Fuller (1979) and their eponymous unit root test. There are numerous benefits to this overall procedure, not the least of which is its conceptual simplicity. The two proposed robust tests in this study are modifications to this basic testing procedure.

However, the Engle-Granger and Dickey-Fuller tests suffer from the flaw of notably weak power, particularly when the underlying distributions are non-Gaussian. To improve efficiency generally in the face of fat-tailed distributions, researchers frequently employ nonparametric methods. The two such methods employed in this study are weighted-Wilcoxon regression and least absolute deviation regression.

The Wilcoxon test is based on ranks and is known in the location problem of estimating the median to have 95% efficiency for the normal distribution. Hettmansperger and McKean (1998) provide a table that further compares Wilcoxon, least absolute deviation (LAD), and least squares estimates for a contaminated normal distribution ($\sigma = 3$). They demonstrate the superiority of the rank-based procedures for even 1% contamination, and for 15% contamination the Wilcoxon procedure is 50% more efficient. LAD estimation also surpasses the least squares estimate, but the crossover occurs at 10% contamination.

The issue of extreme, and perhaps even infinite, variance has been recognized as a problem for time series analysis in general for quite some time. Several approaches for dealing with such series have been tried over the years, including Granger and Orr (1972),

who suggested trimming the series before applying standard techniques. The issue has been particularly prevalent in the financial literature (e.g. Mandelbrot, 1963; Fama, 1965). Dealing with cases of extreme variance time series necessitates robust methods.

The driving motivation of this study is that Wilcoxon estimation sacrifices little when error distributions are normal and provides a significant improvement for fat-tailed distributions. LAD estimation underperforms even more for the case of normally distributed errors, but it may represent an improvement for extremely fat-tailed distributions. This paper proposes modifications to the Engle-Granger procedure and Dickey-Fuller test based on Wilcoxon and LAD procedures.

The format of the remainder of this paper is as follows. Chapter 2 reviews the relevant literature on the topics of robust regression, unit root tests, and cointegration tests. The development of the proposed robust cointegration tests as well as the simulation study methods are laid out in chapter 3; this chapter also hypothesizes the anticipated relationships between the manipulated parameters and the power of the tests in Monte Carlo simulations. The results of those simulations are presented in chapter 4. To illustrate one potential use of the new tests, chapter 5 provides an application in the form of a statistical arbitrage stock trading strategy, and chapter 6 concludes.

CHAPTER 2. LITERATURE REVIEW

The formal concept of cointegration is, in relative terms, a recent addition to the statistics and econometrics literature. It was Granger (1981) who formally introduced the idea and coined the term, while the first formal test for cointegration was presented by Engle and Granger (1987). Despite this recent pedigree, the origins of cointegration can be traced back at least as far as Yule (1926), who recognized the spurious regression problem that arises when least squares regressions are naively applied to trending time series. Furthermore, the research on cointegration has been intimately tied to the prior development of unit root testing, of which perhaps the most well-known test is the Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979, 1981).

After some basic definitions, this chapter reviews the literature on the interrelated topics of regression techniques, unit root testing, and cointegration. In doing so, it provides the background to motivate and develop the testing procedures that are used in this simulation study. First, robust regression techniques are presented, namely rank-based weighted-Wilcoxon (WW) regression and LAD regression. These two techniques are employed as robust alternatives to least squares in the proposed testing procedures. Second, the development of unit root testing is discussed, focusing on the Augmented Dickey-Fuller (ADF) test. Finally, a brief history of cointegration testing is presented. This study primarily utilizes the seminal works by Engle and Granger (1987) and Johansen (1988), while alternative cointegration tests are also discussed briefly.

Basic Concepts and Definitions

To motivate the concept of a time series with a unit root, consider the AR(1) process $y_t = \rho y_{t-1} + u_t$, where the error terms u_t are assumed to be normal, independent and

identically distributed with mean zero and unit variance. (The series is often simplified with the assumption or modification that $y_0 = 0$ as well.) There are three different cases for this model. First, when $|\rho|<1$, the process is stationary. Second, when $|\rho|=1$, the process is a non-stationary random walk; finally, when $|\rho|>1$, the process is a non-stationary, explosive model.

It is simple to demonstrate that an AR(1) process is non-stationary when $|\rho|=1$. Assuming $y_0 = 0$, we can rewrite the series as $y_t = \sum_{i=1}^{t} u_t$ (for $t = 1, 2, 3\ldots$) by repeated substitution. This implies $Var(y_t) = \sum_{i=1}^{t} \sigma^2 = t$. (Non-stationarity can be similarly expressed when $|\rho|>1$.) On the other hand, when $|\rho|<1$ the variance converges asymptotically to $\frac{1}{1-\rho^2}$. Similarly, the autocovariance function of an AR(1) depends on t only if the process has a unit root. If an AR(1) process also includes a constant drift parameter, the mean will also vary with t (Maddala and Kim, 1998).

The term "unit root" arises from rewriting an AR(p) process in terms of the lag operator as $(1-\alpha_1 L-\alpha_2 L^2-\ldots-\alpha_p L^p)y_t = u_t$ or, equivalently as $(1-\pi_1 L)(1-\pi_2 L)\ldots(1-\pi_p L)y_t = u_t$ where $\pi_1,\pi_2,\ldots\pi_p$ are the roots of the equation $m^p - \alpha_1 m^{p-1} - \ldots - \alpha_p = 0$. This equation is said to have a unit root if $m = 1$. For an AR(1) process, only one unit root is possible in which case the process is said to be integrated of order 1, denoted as I(1). An AR(p) process can potentially have multiple roots, say r instances, of $m = 1$, in which case it is integrated of order r and denoted I(r). For many applications and economic theories in particular, it is sufficient to consider I(1) processes.

Robust Regression Techniques

In reviewing the infamous priority dispute between Gauss and Legendre regarding the invention of least squares regression, Stigler (1981) described least squares regression as "the automobile of modern statistical analysis" (p. 465). He explained that while the method has limitations and occasional problems, it is still used for the bulk of statistical analyses. When approaching a new problem or constructing a new test, statisticians still often begin with the least squares approach, departing for more complex analysis only when necessary. In fact, this approach is perfectly valid in many situations, since least squares estimates correspond to maximum likelihood estimates when experimental errors are normally distributed. Therefore, Engle and Granger (1987) naturally employed least squares techniques in their original development of cointegration tests.

In many applications, however, the assumptions that underlie the Gauss-Markov theorem and least squares regression are violated: the error terms may be correlated or possess unequal variances, for example. Some researchers argue even further that in economic time series the variance may be infinite (e.g., Mandelbrot, 1963), which would invalidate the assumptions of even the classical proof of the Central Limit Theorem. Therefore, widening the potential application of the original Engle-Granger (1987) two-stage cointegration test to such time series requires a more robust estimation method. In creating new cointegration tests, this study employs the two alternative regression techniques of weighted-Wilcoxon (WW) and Least Absolute Deviation (LAD).

Weighted-Wilcoxon regression. The WW-estimate uses a rank-based approach to minimize the influence of outliers and departures from normality. Terpstra and McKean (2005) explained that a minimum of the following function provides a WW estimate of a

linear model, where b_{ij} denotes a weight used in the (i, j)th comparison and the residuals are denoted by $\varepsilon_i(\beta) = y_i - \beta_0 - x_i\beta_1$:

$$D_{WR}(\beta) = \sum_{1 \le i < j \le n} b_{ij} \left| \varepsilon_j(\beta) - \varepsilon_i(\beta) \right|$$

When $b_{ij} = 1$ for $i \neq j$ and zero otherwise (Wilcoxon weights), Hettmansperger (1984) demonstrated that this dispersion function reduces to a simpler form:

$$D_{WR}(\beta) = 2\sum_{i=1}^{n} \left(R(\varepsilon_i(\beta)) - \frac{n+1}{2} \right) \varepsilon_i(\beta)$$

Here $R(\varepsilon_i(\beta))$ denotes the rank of the residuals, which demonstrates explicitly why WW regression can be considered a rank-based procedure. Terpstra and McKean (2005) noted further that this function corresponds to Jaeckel's (1972) dispersion function with Wilcoxon scores. Hettmansperger and McKean (1998) commented that the dispersion function is invariant to location, so a reasonable and typical estimate of the intercept is given by $\hat{\beta}_0 = med\left\{\varepsilon_i\left(\hat{\beta}\right)\right\}$. Finally, Hettmansperger (1984) elucidated the robust property of the estimate by noting that the influence of outliers has a linear effect as opposed to the quadratic influence of least squares.

Tests for significance of the regression parameters can be obtained through the reduction in sums of squares test, Rao's scores test, or a Wald test (Hettmansperger and McKean, 1998). Terpstra and McKean (2005) provided details on each of these statistics, and more importantly, have made R code available that calculates the test statistics. Conceptually, the methodology is a straightforward standardization of the WW coefficient estimates. The calculation of the variance-covariance matrix, however, is complex and involves the estimation of nuisance parameters. Hettmansperger and McKean (1998) and

Terpstra and McKean (2005) provided the theoretical development and R code implementation of the procedure, respectively.

The proposed test statistic of this study is built through Monte Carlo simulation. Therefore, the asymptotic, distributional theory of the Wald test itself and the consequent *p*-values reported by the code of Terpstra and McKean (2005) are not of primary concern. The only value used in the proposed robust cointegration test is the standardized statistic that the R code labels TVAL. This statistic is the WW estimate of the slope coefficient, standardized by a measure of standard deviation. While it is thus analogous to the *t*-test of lest squares, the statistic does not follow a standard *t*-distribution.

Least absolute deviation regression. Historically, LAD regression actually predates the development of least squares by 50 years, and its potential efficiency advantage over least squares was known by Laplace (Wilcox, 1997). Once least squares estimators were shown to be maximum likelihood estimators when errors are normally distributed, however, LAD regression was rarely employed. Only the most basic ideas behind LAD estimates are covered here, because they are necessary for the later development of robust unit root and cointegration tests.

Consideration of the formulae for the least squares and WW estimates anticipates the dispersion function for LAD estimation. As a reminder, the least squares estimate is determined by minimizing the following expression, with notation consistent with that given above for the WW estimate:

$$D_{LS}(\beta) = \sum_{i=1}^{n} (\varepsilon_i(\beta))^2$$

This makes it notationally clear that least squares methods is a minimum of the sum of the squared residuals. Similarly, the WW estimator above is based on the sum of each residual

multiplied by its rank. The LAD estimator minimizes the sum of the absolute values of the residuals:

$$D_{LAD}(\beta) = \sum_{i=1}^{n} |\varepsilon_i(\beta)|$$

This lessens the influence of outliers even more and produces better estimates for leptokurtic distributions.

As with WW regression, the coefficients need to be standardized to calculate a test statistic. This is the statistic that is aggregated in the Monte Carlo simulation study for the proposed cointegration test. Koenker and Bassett (1978) proposed an estimate of the asymptotic covariance matrix that can be easily implemented in the R function SUMMARY.RQ. Once again, this statistic is analogous to the least squares *t*-statistic, though it does not share that distribution. Most importantly for the robust properties of the statistic, Koenker and Bassett (1978) observed that while LAD estimates underperform least squares estimates for Gaussian errors, they substantially outperform them for a wide range of other error distributions.

Both WW regression and LAD regression are robust alternatives to least squares, minimizing the influence of outliers. The need for such robust analysis has been acknowledged by a wide range of research, particularly in economics and finance (Mandelbrot, 1963; Fama, 1965). In his book on the topic, Wilcox (1997) summarizes the reasons for this need, including low power, poor coverage, and distortions that arise from incorrect variance estimates. All of these can lead to incorrect conclusions in the face of fat-tailed distributions. This motivates the use of alternative regression techniques to create the proposed robust cointegration tests of this study. Before describing those tests, the preliminary topic of unit root testing is introduced in the next section.

Unit Root Tests

Unit roots arise in non-stationary autoregressive processes and gained prominence following the work of Nelson and Plosser (1982). In that paper, the authors argued that most macroeconomic time series are I(1) processes, that is the series contained an autoregressive unit root. Because of this economic association, cointegration is often understood conceptually as a test for an equilibrium relationship between two non-stationary time series. Unit root tests play two important roles in cointegration testing. First, they are necessary as a pre-test to ensure that candidate time series are I(1). Second, they are often used directly in cointegration tests themselves, most notably as part of the Engle-Granger cointegration test.

The robust cointegration test proposed in this study seeks to improve on the power of existing tests, in part, by creating a more robust unit root test for use in the Engle-Granger methodology. The proposed unit root test is a modification of the Augmented Dickey-Fuller (ADF) test, and this section reviews the development of that statistical test. Following that, some alternative unit root tests are reviewed briefly, and the final two sections discuss previous attempts to make the ADF test more robust by using rank and LAD methods. These variations on the ADF test anticipate the unit root tests that are employed as part of the more robust cointegration test proposed in this simulation study.

Dickey-Fuller and ADF tests. The seminal test for the presence of a unit root is the Dickey-Fuller test statistic (Dickey and Fuller, 1979), which was later extended to the Augmented Dickey-Fuller test (Dickey and Fuller, 1981). This test remains one of the most popular and commonly utilized tests. The basic version is based on a least-squares regression of the equation $x_t = \mu + \rho x_{t-1} + \varepsilon_t$. This is reparameterized to the following

form: $\Delta x_t = \mu + (\rho - 1)x_{t-1} + \varepsilon_t$. This equation is then estimated with least squares, and the appropriate t-statistic is used to test the null hypothesis that $\rho = 1$.

This test statistic, however, has a nonstandard distribution because the series is non-stationary under the null hypothesis. If the standard critical values from a t-distribution are used, the result can be significant over-rejection of the null (Maddala and Kim,1998). Phillips (1987) describes the limiting distribution of the statistic in terms of Wiener processes. This differs from the original approach of Dickey and Fuller (1981), who calculated the appropriate critical values for the above model and others, such as those including a linear time trend.

One shortcoming of the original Dickey-Fuller test is that it includes the implicit assumption that the error term is not serially correlated. The ADF test overcomes this problem by expanding the previous equations to the following form:

$$x_t = \mu + \beta t + \rho x_{t-1} + \sum_{i=1}^{m} \lambda_i \Delta x_{t-i} + \varepsilon_t$$

$$\Delta x_t = \mu + \beta t + (\rho - 1)x_{t-1} + \sum_{i=1}^{m} \lambda_i \Delta x_{t-i} + \varepsilon_t$$

The lag length (m) in the equations above must be determined by the researcher. The number can be assumed *a priori* or determined in a general-to-specific search from a pre-determined maximum. Other possible methods for determining the appropriate lag length include the Akaike Information Criterion (AIC) or the Schwarz Criterion (SC). Once the lag length has been chosen, a least squares regression on the second equation leads to an F ratio to test the null hypothesis that $\beta = 0$ and $\rho = 1$ (or a t-statistic if no time trend is included). As above, the critical values are non-standard due to the non-stationary nature of the series under the null hypothesis.

The ADF test is generally presented in elementary textbooks of time series and econometrics; it is also widely available in computer programs. Therefore, it is likely the most widely used (Maddala and Kim, 1998). However, scholars have proposed many other unit root tests, several of which are discussed in the next section.

Alternative unit root tests. There are three primary reasons for the multiplicity of unit root tests. First, there is no uniformly most powerful test for the unit root hypothesis (Stock, 1994). The other two reasons lie with the size distortion and low power that have been a constant source of criticism for unit root tests. Some of the earliest Monte Carlo evidence of size distortion can be found in Schwert (1989), while DeJong, et al. (1992) complained about the low power of several tests, including the ADF test. Agiakloglou and Newbold (1992) make similar observations regarding the size and power of several unit root tests. Because of these problems, new tests are constantly being developed. This section highlights a few of the tests that have been proposed as replacements to the ADF test.

Two extensions to the basic ADF test have been proposed, the Said-Dickey test (Said and Dickey, 1984) and the Phillips-Perron (PP) test statistic, as developed by Phillips (1987) and Phillips and Perron (1988). The Said-Dickey procedure allows for both AR and moving average (MA) error distributions and therefore begins from a different assumption of the data-generating process than the ADF test. However, the final test statistic is based on a least squares regression and a *t*-test that follows the Dickey-Fuller distribution. Therefore, the test fails to provide a more robust estimate for fat-tailed error distributions.

While the PP test statistic is typically referred to as nonparametric (e.g., Banerjee, Dolado, Galbraith, and Hendry, 1993), this characterization misleadingly conceals the fact

that it is still based on the same least squares regression as the ADF test. The test simply modifies the ADF test statistic with nonparametric corrections that are meant to account for the effect of autocorrelated errors. Once again, the underlying least squares regression limits the robustness of the test. Furthermore, while the PP test has slightly higher power than the Said-Dickey test, both tests suffer from significant size distortions (Phillips and Perron, 1988). These tests offer only marginal improvements over the ADF test.

Maddala and Kim (1998) discussed a wide variety of other unit root tests. The diversity of available unit root tests is both vast and largely peripheral to the development of this study. Among them are the test proposed by Sargan and Bhargava (1983) that is based on the Durbin-Watson statistic, a variance ratio test based on the work of Cochrane (1988) and Lo and MacKinlay (1988), and tests based on instrumental variables as found in Hall (1989) and Choi (1992).

Tests that reverse the hypotheses and use stationarity as the null have also been proposed. The primary example is known as the KPSS test, developed by Kwiatkowski, Phillips, Schmidt, and Shin (1992). Maddala and Kim (1998) also reviewed related tests by Park (1990) and Leybourne and McCabe (1994). Each of these tests offers some improvement at the cost of complexity, but none has been found to be universally better in simulation studies.

Among the other candidates mentioned by Maddala and Kim (1998) in their overview of the literature are M and MM estimators (Lucas, 1995a; Lucas, 1995b), a likelihood based test using nonnormal errors (Rothenberg and Stock, 1997), and a test based on errors with a *t*-distribution (Hoek et al., 1995). Another proposed test is the range unit root test of Aparicio, Escribano, and Sipols (2006). For the most part, these tests have

not been widely employed nor have any comparison studies identified them as generally superior. Following their review of the literature on unit root tests, Maddala and Kim (1998) urged researchers to consider the proper role of unit root tests and summarized the current state of the art with two observations: first, tests beyond the ADF and PP are rarely employed by practitioners or researchers, and second, the commonly used versions of these two popular tests are so lacking in power that they do still need to be replaced or improved. The next two sections review unit root tests that aim to correct these problems by employing rank methodologies and LAD estimation, the same methodologies that will be employed in this investigation

Rank-based unit root tests. One method to introduce rank procedures into the ADF test is to transform the original time series into its ranks and then perform the standard test on the series of ranks. This test is sometimes referred to as the rank-ADF test and was introduced by Granger and Hallman (1991). They observed that the ADF test can perform poorly when the data generating process of the time series is nonlinear. In such a situation, their rank-ADF test performs better in terms of power. Since most time series are generally modeled as linear, at least in the short term, this advantage would appear to be modest. This perhaps explains why the test has been adopted and studied by so few practitioners.

Breitung and Gourieroux (1997) made a slight modification to the rank-ADF test by applying a rank transform to the first differences of the original data and then performing the standard ADF test. One benefit of this unit root test is that it holds even when variance of the data generating process is infinite, so long as the error distribution is symmetric (Janicki and Weron, 1994). This test is robust for structural breaks as well as additive and innovative outliers. However, in a comparison with the rank-ADF test, Fotopoulos and

Ahn (2001) show that the test has lower power and in some cases deteriorates below even that of the basic, parametric ADF test.

Hasan and Koenker (1997) exploit the relationship between quantile regression and rank statistics to give a new test procedure for the hypothesis that a series q_t contains a unit root. The test is based on the following statistic, where b_n is the regression rank score process of Gutenbrunner and Jureckova (1992) and q_{t-1}^{p} represents the projection of q_{t-1} onto the matrix $X_t = [1, \Delta q_{t-1}, \ldots, \Delta q_{t-p}]$:

$$S_n = \frac{(q_{t-1} - q_{t-1}^{p})b_n}{\left[\sum (q_{t-1} - q_{t-1}^{p})^2\right]^{1/2}}$$

With slight modification, they demonstrate that the statistic is asymptotically normal under the null hypothesis.

The results of a Monte Carlo simulation in Hasan and Koenker (1997) clearly demonstrate their test's superiority when the error distribution is nonnormal. For a time series of length 100, the ADF test is inferior for both Cauchy and t_3 distributions and performs only slightly better in terms of power for normal errors. Hasan (2001) extends the rank-based methods above to the situation where the error distribution has infinite variance, namely the Levy stable distribution with stability index α = (1.2, 1.0, 0.8, 0.5) and finds similar results. He observes that the Hasan-Koenker test actually performs better as the tails of the error distribution become fatter, in stark contrast to the ADF test. Koenker (1997) considers the loss of power under conditions of normality to be an acceptable risk to insure against the poor performance of the least-squares methods in situations with heavier tails.

However, there are at least four problems with the Hasan-Koenker approach. First, the test has significant size distortion (all reported power figures and comparisons in their papers are necessarily size adjusted power). Second, they impose an assumption of asymptotic normality on limited sample sizes, rather than generate critical values. Third, the test statistic requires the estimation of nuisance parameters. Last, the complex rank-based methods they use are not easily implemented in software packages. All of these shortcomings are addressed by the straightforward WW-based ADF test that is proposed in the next chapter.

LAD-based unit root tests. A unit root test based on LAD regression was developed by Herce (1996). He proposed several statistics based on modifications of the LAD regression estimate, $\hat{\beta}_{LAD}$, that he described as analogous to the Phillips-Perron adjustments to the ADF test. Of several candidates, the statistic he denoted $L_{\beta,\mu}$ possessed the best properties in terms of power, and he derived the distribution of the statistic as the product of independent Wiener processes and a standard normal random variable.

Herce (1996) provided the results of several simulation studies that demonstrated the superiority of his test statistic when the errors followed a double exponential distribution and the time series had length 100 and 200. This is unsurprising given that LAD regression is optimal for that distribution. The LAD-based test also outperformed in the case of contaminated normal errors, but significantly underperformed for normal errors.

The primary weakness of this LAD-based test is that the test has arbitrarily low power for normally distributed errors. Herce (1996) noted that the problem occurs because the LAD-based tests are inconsistent in that case, and the power can remain arbitrarily low even in the limit, as the length of the time series increases. Nuisance parameters also

remain a problem for implementation. Surmounting these problems is the motivation for the LAD-based unit root test described in the next chapter.

Cointegration Tests

The statistical concept of cointegration is grounded in the problem of the spurious regression that can occur when dealing with nonstationary time series. Hendry (1986) provided the long historical background of nonsense regressions, of which one example is Hooker (1901), who demonstrated a relationship between trade and the marriage rate. Granger and Newbold (1974) provided a rule of thumb for identifying such meaningless relationships, suggesting that any R^2 greater than the Durbin-Watson statistic should raise suspicions. Such problematic series require first differencing of the variables, unless they can be shown to be cointegrated.

Cointegration implies that a linear combination of nonstationary time series exists that is itself stationary. In other words, the two time series do not drift arbitrarily, but return toward an equilibrium relationship. This is sometimes referred to, especially in the field of finance, as mean-reversion of the relationship. A variety of tests have been proposed to identify cointegrated time series, but the bulk of analysis today is done with one of two tests. The first is one of the original tests proposed by Engle and Granger (1987) and the other is the test developed by Johansen (1988) that uses reduced rank regression. This section outlines these two tests and concludes by mentioning a few of the available alternative procedures.

Engle-Granger cointegration test. The first test for cointegration was developed by Engle and Granger (1987), and it validated the use of least squares with nonstationary series. Once it is established that the time series are I(1), a cointegrating regression is run,

which in the bivariate case used in this simulation study is a least squares regression of the form $y_t = \alpha \cdot x_t + z_t$, where z_t represents the error term. The estimate $\hat{\alpha}$ is often described as superconsistent because it converges to its true value at a rate of T, rather than the usual convergence rate of $T^{1/2}$ (Stock, 1987).

Once the least squares estimate is determined, the series of residuals $\hat{z}_t$ is calculated. If that series has a unit root, it implies a lack of mean reversion or equilibrium in the two original series. Therefore, the series would not be cointegrated. Conversely, if the series were cointegrated, the residuals would constitute a stationary series and not contain a unit root. Engle and Granger (1987) examine the potential use of a number of tests, and in the end argue in favor of an augmented Dickey-Fuller test as the most powerful.

It is imperative to notice that the standard ADF test critical values do not apply for the Engle-Granger cointegration test. The reason is that the series of residuals is itself based upon the cointegrating regression estimation. For this reason, Engle and Granger (1987) used Monte Carlo simulation to generate the appropriate critical values. Their work provides a control or baseline comparison for the least squares test that can be compared to the robust cointegration tests proposed in this study.

To conclude this section, several potential problems exist with the cointegrating regression need to be addressed. First, in limited samples the bias can be significant, as demonstrated by Banerjee et al. (1986). Second, the limiting distribution is non-normal, which creates the same problem as spurious regression: typical t and F statistics are not applicable (Stock, 1987). A third issue is normalization. If the regression were normalized to the series x_t, the regression would be of the form $x_t = \beta \cdot y_t + z_t$, and it is generally the case that $\hat{\beta} \approx 1/\hat{\alpha}$. However, Ng and Perron (1997) demonstrated that the least squares

estimator can have poor finite sample properties in one direction and not the other. Finally, any serial correlation or endogeneity (which can be modeled by non-zero covariance of the error terms) requires the estimation of nuisance parameters (Maddala and Kim, 1998). The next section examines one test that attempted to overcome these problems.

Johansen cointegration test. The most widely used cointegration test, according to Maddala and Kim (1998), was proposed by Johansen (1988) and further developed by Johansen and Juselius (1990). The primary motivation and benefit of the Johansen cointegration test is that it avoids the need for normalization. The procedure applies maximum likelihood estimation in a vector autoregression model, with the assumption that the error terms are normally distributed:

$$Y_t = A_1 Y_{t-1} + \cdots + A_k Y_{t-k} + U_t$$

Here Y_t is a vector of I(1) variables, so the first difference will be stationary. The model can be rewritten to capitalize on this property:

$$\Delta Y_t = B_1 Y_{t-1} + B_2 \Delta Y_{t-1} + \cdots + B_k \Delta Y_{t-k+1} + U_t$$

Since all of the differenced terms are stationary while Y_{t-1} is non-stationary, the matrix B_1 cannot be full rank. If it has rank r, then it can be rewritten as the product of two matrices AB, and BY are then the r cointegrated variables. In the bivariate case, there is at most one cointegrating constant.

The Johansen estimation procedure is called reduced rank regression and ends with maximizing the likelihood function by solving an eigenvalue problem. This results in two tests for cointegration known as the trace and maximum eigenvalue statistics. Each of these is based on the fact that if there are r cointegrating vectors, this implies that the $(n - r)$ smallest eigenvalues are zero. Johansen and Juselius (1990) favored the maximum

eigenvalue test, and critical values can be found in Osterwald-Lenum (1992). For consistency in this simulation study, however, the critical values are generated, as described in the next chapter.

Alternative cointegration tests. The search for effective cointegration tests has continued to challenge researchers, given the numerous potential problems and multiple settings for such tests. Of particular interest in this study, leptokurtic error distributions invalidate the assumptions that underlie most test procedures. Other problems can have a dramatic effect on the power of cointegration tests as well: nonlinearities, homoskedasticity, or structural breaks.

General robust methods can in some cases deal with these multiple problems. Koenker and Bassett (1978) observed that biased estimators do exist that are superior for non-Gaussian distributions, even beyond the case of infinite variance: "While least squares is obviously abysmal for distributions having infinite variance (having zero efficiency for the Cauchy for example) its gross inferiority to a variety of nonlinear estimators is by no means confined to distributions with infinite variance" (p. 35).

Both the Engle-Granger and Johansen tests perform poorly when faced with nonlinearities, homoskedasticity, or structural breaks. Therefore, a range of alternative tests has been proposed using various methods, including induced order statistics (Escribano, Santos, and Sipols, 2008), record counting correction (Escribano, Sipols, and Aparicio, 2006), and a nonparametric version of the Johansen test (Bierens, 1997). Furthermore, modifications to unit root testing allows for new cointegration tests, as can be found in the fully modified least squares of Phillips (1995) and range unit root tests (Aparicio, Escribano, and Sipols, 2006).

The range of new tests is impressive since the very concept of cointegration was defined less than thirty years ago. However, few of them have yet been used in applications or compared in published simulations. One particular advantage of this study is the direct application and comparison of the proposed robust cointegration tests. Furthermore, the proposed tests combine robust techniques with the simplicity of the Engle-Granger test.

Summary

This chapter has highlighted the relevant literature relating to three broad areas: robust regression techniques, unit root tests, and cointegration tests. Despite the variety of extant tests, low power is a persistent problem for cointegration tests, particularly when the errors are leptokurtic. Making robust modifications to the Engle-Granger procedure should improve performance for such fat-tailed distributions. The next chapter outlines such modifications as well as the simulation study methods that allow for comparison of the tests.

CHAPTER 3. TEST DEVELOPMENT AND SIMULATION STUDY METHOD

This investigation compares the properties of the Engle-Granger and Johansen cointegration tests with WW (rank)-based and LAD-based procedures. In order to estimate the size and power of the four tests, Monte Carlo simulations were designed and implemented in R code. According to the website of the R Project for Statistical Computing (http://www.r-project.org), "R is a language and environment for statistical computing and graphics." The open source code format allows for compilation of the work of several authors as well as a great deal of flexibility to manipulate existing functions. All simulations and graphics in this investigation were created in the R environment.

This chapter begins by explaining the modifications made to the ADF test to make it more robust. It then describes the procedures of the Engle-Granger cointegration test and the modifications made to it in order to create two proposed cointegration tests that are robust to fat-tailed errors. The implementation of the Johansen procedure is also outlined. Following that, the details of the simulation study are provided: the data-generation process used and the various parameters altered for the purpose of comparing the three cointegration tests.

ADF Test Modifications

Since the goal of this investigation is to replace least squares procedures with more robust techniques in cointegration testing, it is not sufficient to modify only the first step from the Engle-Granger procedure, the cointegrating regression. The unit root test must also differ from the classic ADF test, to remove its reliance on least squares estimation in favor of a more robust method. The R function used for all three unit root tests in this

study is MY.DF.TEST, which includes three versions based on least squares, Wilcoxon, and LAD regressions.

The first version, based on least squares regression, is identical to the original ADF test. The function LSFIT conducts a least squares regression of the first difference of the residuals on the residuals themselves. The test statistic of interest is the *t*-statistic for the null hypothesis that the slope is equal to zero. Rejection of the null hypothesis implies a stationary series. Conversely, non-rejection entails a non-stationary time series. However, in this simulation study, previously published critical values are not used to determine *p*-values or a rejection decision. Instead, the *t*-statistic is retained to generate critical values. This simulation technique is what allows a direct and valid comparison to the robust techniques by guaranteeing all tests share an empirical and not just nominal size.

The second version of the unit root test is an application of WW regression in the same general format as the ADF test. The difference in the residuals is regressed on the residuals via WW regression, using the function WWEST. The slope coefficient is standardized, and the distribution of this statistic is built through replication in the Monte Carlo simulation. The 5% critical value is then used in estimations of the test's size and power.

Similarly, LAD regression is used for the third and final version of MY.DF.TEST. Once again, the coefficient is estimated and standardized. This test calls the quantile regression function RQ, and the standard errors are calculated using the method of Koenker and Bassett (1978). Replication of the procedure then provides the 5% critical value of the test statistic.

Cointegration Tests

As discussed previously, the Engle-Granger test procedure consists of ordinary least squares regression followed by an augmented Dickey-Fuller test (which also involves least squares estimation). In the simulation study, the former calls the basic R function LM. The latter was carried out by the least squares version of the function MY.DF.TEST.

Replicating the Engle-Granger test with its least squares procedure fulfills two roles. First, it allows for comparison to the original results and thereby provides confidence in the validity of the simulation. Second, it provides a baseline for comparison with the proposed tests. The two proposed tests are robust variations on this seminal cointegration test.

The first robust cointegration test is based on WW regression procedures. First, the weighted Wilcoxon estimates are calculated using the function WWFIT (Terpstra and McKean, 2005; Terpstra, McKean, and Naranjo, 2001)[1]. The procedure employs a Wilcoxon weighting scheme. Analogous to the Engle-Granger method, these estimates supply residuals for the second step of the cointegration test, which is conducted by use of the WIL version of MY.DF.TEST.

The other proposed robust cointegration test uses LAD regression procedures throughout. It begins with a cointegrating regression using the function RQ. This step provides the residuals that are then tested for stationarity using the LAD version of MY.DF.TEST.

The implementation of the Johansen (1991) cointegration test comes from the contributed package Urca (Pfaff, 2006). This package includes the function CA.JO, which is employed to calculate the maximum eigenvalue version of the Johansen test. (The trace

[1] The R code for this function (and other necessary sub-functions) is made available by Terpstra and McKean at this website: http://www.stat.wmich.edu/mckean/HMC/Rcode/AppendixB/ww.r

statistic is not considered.) To mitigate any size distortions, the critical values of this test statistic are also generated through Monte Carlo simulation; however, no modifications are made to this test. It is included in this study solely for comparison.

The first three tests follow analogous procedures, utilizing different regression techniques for each step. The Johansen test is included for comparison because it is one of the more popular extant cointegration tests. The simulation proceeds by generating time series and applying each of the four tests independently within each replication. Comparison of their performance (in terms of power) is the primary contribution of this study. The remainder of this chapter discusses the various other parameters that are modified to allow for comparison under a multiplicity of conditions.

Simulation Study Parameters

Time series length. Much of the theory of cointegration tests is based on the validity of asymptotic distributions. However, practitioners frequently work with time series of limited length. One primary example is financial time series, where structural change can invalidate long-term time horizons, and the data are measured at discrete intervals. Therefore, this applied study focuses on finite series, specifically series with lengths $t = (50, 100, 250)$. The largest of these values corresponds approximately to the number of trading days on the New York Stock Exchange in a given year. This is of particular interest in the application to statistical arbitrage trading strategies, as discussed in Chapter 5.

Data generating process. The Monte Carlo simulations all employ the following data-generation process for the two time series x and y:

$$
\begin{aligned}
x_t + y_t &= u_t \\
x_t + \beta y_t &= e_t \quad \text{with} \quad |\rho| \le 1 \\
u_t &= u_{t-1} + \varepsilon_{1t} \qquad \beta = 2 \\
e_t &= \rho e_{t-1} + \varepsilon_{2t}
\end{aligned}
$$

In general, the cointegrating constant β can take on any real value, but it is convenient for the purpose of testing and comparison to limit the data-generation to this one particular value, which is also a common assumption in previous studies (e.g., Phillips and Hansen, 1990). The error vector $(\varepsilon_{1t}, \varepsilon_{2t})'$ consists of two independent and identically distributed stable random variables, each with an expected value of zero. The basic form of this DGP has been widely employed (e.g., Banerjee, Dolado, Hendry, and Smith, 1986; Engle and Granger, 1987; Phillips and Loretan, 1991; Kremers, Ericsson, and Dolado, 1992). Solving for x and y gives the following equations that are used to generate the simulated time series:

$$
\begin{aligned}
x_t &= \beta(\beta - 1)^{-1} u_t - (\beta - 1)^{-1} e_t \\
y_t &= -(\beta - 1)^{-1} u_t + (\beta - 1)^{-1} e_t
\end{aligned}
$$

When $|\rho| = 1$, both variables are I(1) and the two series are not cointegrated. In that case, the null hypothesis is false; this allows for estimation of the tests' size. When $|\rho| < 1$, both variables are still I(1) but they are cointegrated with cointegrating coefficient $\beta = -2$. Thus, varying the value of rho allows for estimation of power. It is more difficult for the tests to differentiate the near-cointegrated cases when $|\rho|$ is close to one; that is, power varies inversely with ρ. The simulations include a range of values for comparison among the three testing procedures: ρ = (0.8, 0.9, 0.95, 1).

In the simulations, time series of length $t + 20$ are generated by this data-generating process, and the first 20 values are then discarded. This methodology minimizes start-up

effects. Furthermore, it allows for the valid inclusion of a constant term in all regressions and tests.

Some authors have broadened the data-generating process to allow for correlated error structure. For example, Hansen and Phillips (1990) generate multivariate normal error terms. They view the degree of covariance as a measure of the endogeneity of the system. While covariance and endogenous variables may be important in many applications, their consideration adds to the complexity and lies outside the scope of this investigation. All error terms in the Monte Carlo simulations in this study are independent, and the next section addresses their other properties in more depth.

Error distributions. The data-generating process described above is quite general with regard to the distribution of the error terms. Engle and Granger (1987) considered only normally distributed errors in the development of their cointegration test. Most papers since then have made similarly strong restrictions by considering a narrow range of error distributions. The purpose of the proposed WW- and LAD-based methods is to demonstrate more robust tests that apply to a broader range of error distributions.

Stable distributions provide a family of probability distributions that can include the properties of both heavy tails and skewness. One theoretical reason for their use is that the normal distribution is a special case within this family. In fact, the normal, Cauchy, and Levy distributions are the only three special cases within this family for which a density function exists in closed form. In general, stable distributions are those distributions that maintain their shape under addition, leading to Nolan's (2003) definition of the stable law: if X, X_1, X_2, ...,X_n are independent and identically distributed stable random variables, then for all n, $X_1 + X_2 + \ldots + X_n \xrightarrow{d} c_n X + d_n$ for some constants $c_n > 0$ and d_n.

There are at least three reasons that stable distributions meeting this criterion may be chosen as a model for a given data set, as outlined in Nolan (2003). First, there may be theoretical reasons for expecting such a distribution in a physical process. Second, empirical investigations may demonstrate that a data set exhibits heavy tails and skewness. This motivation has a long history in finance and economics, with examples from Mandelbrot (1963) and Fama (1965) to Rachev and Mittnik (2000). Stock prices in particular are well known to exhibit leptokurtic, left-skewed distributions.

The final motivation for modeling with stable distributions is the Generalized Central Limit Theorem (GCLT). For comparison, the classical Central Limit Theorem states that for a random sample drawn from a distribution with some mean μ and positive (non-infinite) variance σ^2, $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} Z$, where Z is the standard normal distribution (Hogg, McKean, and Craig, 2005). The GCLT drops the assumption of finite variance and shows that the limiting distribution must be stable. That is, if $X, X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables, there exist constants $a_n > 0$, b_n, and a non-degenerate random variable Z such that $a_n(X_1 + \ldots + X_n) - b_n \xrightarrow{d} Z$ if and only if Z is stable (Nolan, 2003). Therefore, if an observed data series can be thought of as the sum of many small terms it may be modeled adequately by a stable distribution. This is precisely the case in many financial time series, including stock prices.

In general, Stable distributions are characterized by four parameters: a characteristic exponent $\alpha \in (0,2]$, a skewness parameter $\beta \in [-1,1]$, a scale parameter $\gamma \geq 0$, and a location parameter $\delta \in \Re$ (Nolan, 2003). In this investigation, the normal distribution ($\alpha = 2$, $\beta = 0$) is included as a control or baseline for comparison to the heavy tailed distributions.

The other distributions considered are chosen for inclusion due to their tail behaviors; that is, the values of α are such that $1 < \alpha < 2$, exhibiting a leptokurtic distribution.

In the Monte Carlo simulations, the particular value of $\alpha = 1.5$ is chosen as the midway point, while $\alpha = 1.7$ is selected due to Mandelbrot's (1963) well-known estimate of the parameter from the time series of cotton prices. In that same work, Mandelbrot notes that the data suggests that the series are not symmetric. Instead, β "takes a small negative value" (p. 405). This is the motivation for considering a skewness parameter of $\beta = -0.25$ in this study. The joint consideration of $\alpha = (1.5, 1.7)$ and $\beta = (0, -0.25)$ implies four stable distributions used to generate error terms in addition to the normal distribution. The power of the WW and LAD tests should increase as α decreases, and the power of all the tests should suffer from the skewness that is implied by a negative beta parameter.

The Engle-Granger data-generating process originally required an expected value of zero. Stable distributions, however, do not always have well-defined expectations. Specifically, the mean does not exist whenever $\alpha \leq 1$ (Nolan, 2003). In that case, the requirement could be more broadly interpreted to mean a location parameter (δ) equal to zero. Note that this is equivalent to the original assumption of a zero mean whenever $\alpha > 1$.

The final parameter is the scale parameter γ, which corresponds to variance for the normal distribution. In a financial application, this parameter would describe the volatility of a given price series. Unlike the other three parameters, which are equal in both time series within a given simulation, the ratio of scale parameters is varied so that it takes on the values $\gamma_1/\gamma_2 = (4, 2, 1)$. This ratio can be thought of as a signal-to-noise ratio, as in Hansen and Phillips (1990). Using the normal distribution in a Monte Carlo simulation study of cointegration, Banerjee et al. (1993) allowed the ratio of variances to cover a

similarly wide range. The authors measured the bias in estimating the cointegrating coefficient and found that it varied inversely with the ratio of variances. This logically implies that power should vary directly with the ratio γ_1/γ_2 where the scale parameters γ_1 and γ_2 are used to generate the errors ε_{1t} and ε_{2t}, respectively.

Lag length. The size and power of the EG and Rank tests are influenced by the number of lagged terms included in the second step unit root test (Schwert, 1989). This necessitates a choice from several potential guidelines for the number of lagged terms: an arbitrary fixed level; a function of the length of the time series, such as Schwert's proposal of $Int\left[12(T/100)^{1/4}\right]$; the Akaike or Schwarz Information Criteria; or a sequential rule. Examples of the last method are discussed in Hall (1994), in which the superior choice is demonstrated to be general to specific. In that method, the researcher begins with a large value, tests the significance of the last term, and then decreases the lag length until a significant statistic is found.

In a comparison study of these choices, Ng and Perron (1995) conclude that the general to specific method consistently chooses larger values for the appropriate lag length. This inclusion of more lags tends to decrease power slightly, but it lessens size distortions significantly (DeJong, Nankervis, Savin, and Whiteman, 1992). Therefore, once again the general to specific methodology is preferred for applied research.

The question of lag length is clearly non-trivial, and previous research has provided guidance for use in any given application. However, considerations of lag length are not the primary focus of this study. Furthermore, using the general to specific method would add unwieldy complexity and computing time to the simulations. Therefore, for the

purpose of this exploratory investigation, lag length is fixed at the prespecified level of l = 1 for all tests.

Critical Value Generation

In making comparisons among Monte Carlo simulation results, it is imperative to ensure approximately equal size for each test. Tests that are "liberal" or "conservative" in comparison to their nominal size also have distorted power. Naïve power comparisons that ignore this issue of size distortion can lead to incorrect conclusions. Lloyd (2005a, 2005b) provides several suggestions to ensure valid comparisons, including *post hoc* methodologies to calculate size-adjusted power. He argues most forcefully that no matter what method is chosen, steps must be taken to ensure the validity of power comparisons among proposed tests.

This investigation avoids the necessity for size-adjusted power by generating the critical values in the same manner as Engle and Granger (1987). The distributions of the test statistics are built through 10,000 repetitions of the test under the null hypothesis, and the 5% critical values are determined from the empirical distribution of the test statistic. These critical values are then used in the simulations to calculate size and power. In this way, all of the tests have approximately equal empirical size of 5% and not just a nominal size of 5% that can vary widely in application.

There are two primary advantages of this approach. First, it eliminates the need for *post hoc* size adjustments before comparing the tests. Second, the least squares results are analogous to the original results of Engle and Granger (1987). Similar results in that case provide assurance that the simulation code is correctly specified. All critical values used throughout the simulations were therefore generated using this empirical methodology.

Size and Power Estimation

Once the critical values have been generated, the simulations are run again under the null hypothesis of no cointegration. In that case the time series are built from independent random errors, and the tests are based upon the generated critical values. Theoretically, all tests should have an empirical size of 5% with a standard error of $\sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{10{,}000}}$, since the simulations are all run with 10,000 repetitions (Lloyd, 2005a). This standard error is dependent upon the calculated size, but for an expected size of 5% the standard error would be approximately just 0.2179%.

Power calculations are made by generating cointegrated series and applying the tests with the estimated critical values. As mentioned in the discussion of the data generating process, power is calculated for three values of rho: $\rho = (0.8, 0.9, 0.95)$.

Summary

The Monte Carlo simulations employ 10,000 replications on the parameter space $t = (50, 100, 250)$, $\gamma_1/\gamma_2 = (4, 2, 1)$, and $\rho = (0.8, 0.9, 0.95, 1)$. Five different error distributions are considered in the data generating process. In addition to the normal distribution, errors are generated using the four distributions implied by the parameter space $\alpha = (1.7, 1.5)$ and $\beta = (0, -0.25)$. Finally, each of the four cointegration tests (OLS, Rank, LAD, and JO) is applied to allow for comparison of the tests' size and power. This gives a total of 720 experiments. The R code utilized for these experiments appears in Appendix A.

CHAPTER 4. RESULTS

Results for the 720 simulation experiments are presented in tables 1, 2, and 3. In table 1, the error distributions are generated with $\gamma_1/\gamma_2 = 1$, while Tables 2 and 3 have this dispersion parameter ratio set to 2 and 4, respectively. In each table, Panel A employs normally distributed errors in the data generating process, Panel B uses a Levy distribution with fatter tails ($\alpha = 1.7$), and Panel D has even greater leptokurtosis ($\alpha = 1.5$). Meanwhile, Panels C and E utilize the same respective values of α and also employ a negative skew for its error distribution ($\beta = -0.25$).

In the tables, the four tests are labeled as follows: least absolute deviation-based test (LAD), weighted Wilcoxon rank-based (Rank), least squares Engle-Granger procedure (OLS), and the Johansen test (JO). The reported values are the means over 10,000 repetitions. Standard errors are given by the formula $\sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{10{,}000}}$. Therefore, the maximum half-width of a 95% confidence interval is 0.0098. Therefore, any reported figures for power that differ by more than 0.0196 would be guaranteed to differ significantly at the 5% level. It should be noted that this is the most conservative estimate of the standard errors. For reported power of 0.8, the half-width would shrink to 0.004. Throughout the tables, the results display the superiority of the proposed test statistic over the original Engle-Granger procedure as well as over the Johansen test.

Results for Individual Parameters

This section will review the results in light of each of the varied parameters that were discussed previously. First, we observe that the power increases in all cases with the length of the time series being tested, as expected. Performance for $t = 50$ is unacceptable

in any case, never surpassing 50%. However, results for $t = 250$ are adequate for practical use in many of the cases, depending on other parameters.

In all cases, the power of the tests varies inversely with ρ, again matching expectations. Power suffers significantly in the near cointegrated cases, where $\rho = 0.95$. This is one parameter that demonstrates the superiority of the rank-based test. For example, in Table 1, Panel E the rank statistic has reported power of 92.72% but the OLS test has only 21.97% for $t = 250$ and $\rho = 0.95$. In general, the rank test maintains its power better as this parameter increases, which is one important aspect of its overall performance.

To compare error distribution, we must make comparisons across the panels of each table. Tables A, B, and D all employ symmetric error distributions. Reading across these tables it becomes evident that the LAD and rank tests perform better in the presence of fat-tailed distributions, while the OLS and JO tests have declining power. This is a primary advantage of the rank test over the standard Engle-Granger procedure and stems from the more robust estimation methods employed. As an example, for $\rho = 0.8$ and $t = 100$, the power of the rank test is 60.08% in Panel A of Table 1, but it improves significantly to 87.31% in Panel B and 95.96% in Panel D. For comparison, the OLS test with those same parameters goes from 67.44% to 59.07% and 54.87%.

Panels C and E introduce a skewness parameter of $\beta = -0.25$ in the data generating process. Surprisingly, this change improves performance of the tests in most cases, though this improvement is quite moderate. For instance, the power for the rank test in Panel D, with $\rho = 0.95$, and $t = 250$ is 89.67% but the corresponding power in Panel E is 92.72%. That case is representative of the results across all tests.

To consider the effect of the dispersion parameter ratio, we need to make comparison among the three tables. Again, there is little difference in power due to variation in this parameter; however, the general trend is for power to increase as the ratio increases. This trend can be seen in the slightly higher power in Table 3. For instance, in Panel B, for the rank test with $\rho = 0.95$ and $t = 250$, Table 1 reports power of 70.34%, which increases to 74.91% in Table 2 and 76.38% in Table 3.

Overall Results

In Panel A of all three tables, the most powerful test in all cases is the Engle-Granger test. This is unsurprising, since normally distributed errors are a key assumption of the theoretical validity of that procedure. However, the rank-based procedure sacrifices little power even in this baseline case. The comparison for $\rho = 0.95$ and $t = 250$ is 27.74% for the OLS test and 27.22% for the rank test in Table 1, and the corresponding numbers in Tables 2 and 3 are 29.67% and 28.49%, 31.57% and 29.37%. The loss of power is minimal, so there is little misspecification risk from using the rank test when errors are, in fact, normally distributed.

For non-normally distributed errors, however, the Rank and LAD tests quickly gain power and overtake the Engle-Granger estimation method. In fact, the robust tests perform better as the tails of the error distribution get fatter. For instance, the rank test has power of only 27.22% for the normal distribution, length of 250, and $\rho = 0.95$, but for the same parameters and the skew alpha distribution in Panel E, the power increases to 92.72%.

The particular example cited above is especially powerful evidence in favor of the rank-based test, because it is an example of two series that are near cointegrated. As rho approaches one, previous testing procedures have notoriously low power. For time series

of length 50 and 100, the power of the OLS method barely exceeds the test's size. Clearly, this is a substandard test for drawing any valid conclusions, because the risk of Type II error is so large. The Johansen test, also popular in the finance and economic literature, suffers from the same problem. This is one more argument supporting the proposed rank-based test statistic of this study.

Finally, note that the rank-based test statistic is relatively stable over all simulations. The critical value of the rank-based test range from -3.36 for a time series of length 50 with normally distributed errors to just -2.98 for a time series of length 250 in Panel E of Table 3. This stability is a desirable feature because it demonstrates that slight misspecifications will not dramatically alter the test's conclusions. It also provides a heuristic rule-of-thumb for the test, since all critical values are approximately -3.

Overall, the superiority of the proposed test statistic is evident. It possesses greater power for fat-tailed distributions, while sacrificing little in the baseline case of normal errors. It has stable critical values with a nominal and empirical size of 5%. It performs well for near-cointegrated series, and it is conceptually simple and easy to implement with existing statistical packages.

Table 1. Simulation results, dispersion ratio = 1

Panel A

Error Distribution: Normal (alpha=2, s = 1)						
Method	t	CV	Size	0.8	0.9	0.95
LAD	50	-3.7573	0.0468	0.1278	0.0786	0.0600
	100	-3.5673	0.0529	0.2971	0.1171	0.0738
	250	-3.2925	0.0565	0.8746	0.4409	0.1804
Rank	50	-3.3637	0.0446	0.1710	0.0822	0.0542
	100	-3.3008	0.0549	0.6008	0.1852	0.0890
	250	-3.3049	0.0500	0.9999	0.7917	0.2722
OLS	50	-3.4451	0.0464	0.1970	0.0846	0.0583
	100	-3.3570	0.0525	0.6744	0.2009	0.0926
	250	-3.3595	0.0465	1.0000	0.8424	0.2774
JO	50	16.1066	0.0464	0.0988	0.0611	0.0502
	100	15.4048	0.0506	0.3573	0.1129	0.0614
	250	15.2527	0.0476	0.9956	0.5818	0.1626

Panel B

Error Distribution: alpha = c(1.7, 1.7)						
Method	t	CV	Size	0.8	0.9	0.95
LAD	50	-3.7025	0.0534	0.2423	0.1111	0.0699
	100	-3.4337	0.0471	0.6320	0.3127	0.1304
	250	-3.0901	0.0498	0.9950	0.8765	0.5325
Rank	50	-3.3453	0.0497	0.3364	0.1229	0.0705
	100	-3.2876	0.0464	0.8731	0.4213	0.1557
	250	-3.1685	0.0483	0.9990	0.9844	0.7034
OLS	50	-3.5084	0.0463	0.1648	0.0706	0.0509
	100	-3.4543	0.0482	0.5907	0.1568	0.0719
	250	-3.3896	0.0498	0.9944	0.8009	0.2414
JO	50	17.2826	0.0445	0.0760	0.0508	0.0455
	100	16.6621	0.0460	0.2673	0.0808	0.0535
	250	16.0289	0.0456	0.9920	0.4943	0.1302

Panel C

Error Distribution: Skew alpha=1.7, beta= -0.25						
Method	t	CV	Size	0.8	0.9	0.95
LAD	50	-3.6807	0.0518	0.2503	0.1138	0.0731
	100	-3.3901	0.0531	0.6487	0.3159	0.1417
	250	-3.1056	0.0472	0.9956	0.8756	0.5287
Rank	50	-3.3656	0.0456	0.3334	0.1219	0.0668
	100	-3.2249	0.0553	0.8942	0.4501	0.1688
	250	-3.1552	0.0484	0.9990	0.9861	0.7092
OLS	50	-3.4908	0.0483	0.1707	0.0749	0.0545
	100	-3.3929	0.0553	0.6367	0.1823	0.0757
	250	-3.3895	0.0524	0.9953	0.8187	0.2494
JO	50	16.9852	0.0474	0.0802	0.0532	0.0454
	100	16.1435	0.0496	0.2915	0.0971	0.0592
	250	15.7251	0.0503	0.9926	0.5350	0.1393

Table 1. Simulation results, dispersion ratio = 1

Panel D

Error Distribution: alpha = c(1.5, 1.5)						
Method	t	CV	Size	0.8	0.9	0.95
LAD	50	-3.7795	0.0486	0.3505	0.1483	0.0769
	100	-3.2988	0.0510	0.8554	0.5208	0.2382
	250	-2.9658	0.0468	0.9976	0.9795	0.7997
Rank	50	-3.4042	0.0506	0.4566	0.1653	0.0822
	100	-3.2064	0.0494	0.9596	0.6408	0.2747
	250	-3.0834	0.0475	0.9971	0.9920	0.8967
OLS	50	-3.5393	0.0512	0.1477	0.0591	0.0521
	100	-3.4828	0.0534	0.5487	0.1439	0.0685
	250	-3.4894	0.0503	0.9866	0.7400	0.1828
JO	50	17.7879	0.0517	0.0753	0.0533	0.0469
	100	17.5695	0.0473	0.2067	0.0706	0.0513
	250	16.9348	0.0595	0.9904	0.4661	0.1171

Panel E

Error Distribution: Skew alpha=1.5, beta= -0.25						
Method	t	CV	Size	0.8	0.9	0.95
LAD	50	-3.7221	0.0523	0.3689	0.1578	0.0765
	100	-3.2546	0.0558	0.8665	0.5473	0.2510
	250	-2.8872	0.0498	0.9974	0.9846	0.8294
Rank	50	-3.3592	0.0561	0.4792	0.1865	0.0837
	100	-3.1594	0.0563	0.9663	0.6840	0.2974
	250	-2.9851	0.0511	0.9967	0.9938	0.9272
OLS	50	-3.4979	0.0536	0.1600	0.0744	0.0514
	100	-3.4823	0.0503	0.5510	0.1458	0.0674
	250	-3.4185	0.0528	0.9896	0.8012	0.2197
JO	50	17.7707	0.0515	0.0773	0.0520	0.0448
	100	17.6218	0.0428	0.2035	0.0721	0.0486
	250	16.3744	0.0511	0.9908	0.4578	0.1121

Table 2. Simulation results, dispersion ratio = 2

Panel A

Error Distribution: Normal (alpha=2, s = 1)						
Method	t	CV	Size	0.8	0.9	0.95
LAD	50	-3.7356	0.0508	0.1331	0.0766	0.0616
	100	-3.5960	0.0495	0.2980	0.1246	0.0796
	250	-3.3808	0.0491	0.8659	0.4184	0.1675
Rank	50	-3.3135	0.0516	0.2060	0.0918	0.0610
	100	-3.3197	0.0524	0.6251	0.2007	0.0937
	250	-3.3613	0.0489	1.0000	0.8044	0.2849
OLS	50	-3.4125	0.0520	0.2251	0.0975	0.0610
	100	-3.3722	0.0523	0.6991	0.2170	0.0988
	250	-3.3764	0.0451	1.0000	0.8542	0.2967
JO	50	15.9884	0.0495	0.1017	0.0609	0.0514
	100	15.2867	0.0544	0.3672	0.1178	0.0693
	250	15.4269	0.0469	0.9945	0.5641	0.1574

Panel B

Error Distribution: alpha = c(1.7, 1.7)						
Method	t	CV	Size	0.8	0.9	0.95
LAD	50	-3.7836	0.0477	0.2628	0.1206	0.0705
	100	-3.3833	0.0505	0.6692	0.3492	0.1667
	250	-3.0876	0.0495	0.9964	0.8938	0.5585
Rank	50	-3.3655	0.0457	0.3747	0.1518	0.0733
	100	-3.2332	0.0514	0.9086	0.4913	0.1994
	250	-3.1450	0.0512	0.9998	0.9916	0.7491
OLS	50	-3.4863	0.0466	0.1872	0.0792	0.0549
	100	-3.3956	0.0491	0.6753	0.1918	0.0850
	250	-3.3872	0.0559	0.9977	0.8548	0.2680
JO	50	16.8036	0.0520	0.0846	0.0578	0.0449
	100	16.1418	0.0490	0.2920	0.1009	0.0573
	250	15.9556	0.0517	0.9932	0.5151	0.1292

Panel C

Error Distribution: Skew alpha=1.7, beta= -0.25						
Method	t	CV	Size	0.8	0.9	0.95
LAD	50	-3.6893	0.0543	0.2832	0.1268	0.0741
	100	-3.3937	0.0516	0.6650	0.3515	0.1570
	250	-3.1219	0.0489	0.9964	0.8785	0.5575
Rank	50	-3.3040	0.0559	0.4091	0.1559	0.0800
	100	-3.2002	0.0567	0.9116	0.5174	0.2110
	250	-3.1415	0.0510	0.9998	0.9914	0.7553
OLS	50	-3.4606	0.0505	0.1983	0.0820	0.0542
	100	-3.4496	0.0462	0.6496	0.1801	0.0790
	250	-3.3655	0.0543	0.9977	0.8685	0.2737
JO	50	16.4920	0.0524	0.0936	0.0624	0.0503
	100	16.4871	0.0432	0.2721	0.0843	0.0574
	250	15.7819	0.0492	0.9920	0.5290	0.1308

Table 2. Simulation results, dispersion ratio = 2

Panel D

Error Distribution: alpha = c(1.5, 1.5)						
Method	t	CV	Size	0.8	0.9	0.95
LAD	50	-3.7031	0.0518	0.4242	0.1897	0.0868
	100	-3.2656	0.0532	0.8808	0.5874	0.2961
	250	-2.9564	0.0466	0.9993	0.9867	0.8412
Rank	50	-3.4320	0.0476	0.5273	0.2079	0.0841
	100	-3.2086	0.0472	0.9740	0.7165	0.3377
	250	-3.0381	0.0488	0.9990	0.9965	0.9389
OLS	50	-3.5729	0.0505	0.1621	0.0636	0.0461
	100	-3.4984	0.0487	0.5961	0.1506	0.0682
	250	-3.4643	0.0501	0.9929	0.8189	0.2121
JO	50	17.8688	0.0504	0.0786	0.0522	0.0503
	100	17.2983	0.0479	0.2137	0.0737	0.0529
	250	17.2949	0.0491	0.9817	0.3768	0.0958

Panel E

Error Distribution: Skew alpha=1.5, beta= -0.25						
Method	t	CV	Size	0.8	0.9	0.95
LAD	50	-3.7201	0.0483	0.4127	0.1949	0.0915
	100	-3.3260	0.0486	0.8664	0.5705	0.2769
	250	-2.8859	0.0492	0.9992	0.9878	0.8540
Rank	50	-3.4223	0.0444	0.5353	0.2172	0.0919
	100	-3.1551	0.0542	0.9786	0.7366	0.3588
	250	-3.0096	0.0473	0.9990	0.9979	0.9451
OLS	50	-3.5613	0.0450	0.1544	0.0677	0.0478
	100	-3.4482	0.0526	0.6433	0.1681	0.0773
	250	-3.4730	0.0481	0.9947	0.8262	0.2203
JO	50	17.8953	0.0478	0.0702	0.0525	0.0468
	100	16.7842	0.0544	0.2544	0.0869	0.0619
	250	16.5880	0.0508	0.9883	0.4364	0.1086

Table 3. Simulation results, dispersion ratio = 4

Panel A

Error Distribution: Normal (alpha=2, s = 1)						
Method	t	CV	Size	0.8	0.9	0.95
LAD	50	-3.8014	0.0473	0.1307	0.0753	0.0611
	100	-3.5844	0.0469	0.3063	0.1255	0.0751
	250	-3.3837	0.0466	0.8696	0.4285	0.1762
Rank	50	-3.3361	0.0513	0.2084	0.0929	0.0668
	100	-3.3513	0.0449	0.6134	0.1957	0.0896
	250	-3.3115	0.0506	1.0000	0.8253	0.2937
OLS	50	-3.4352	0.0494	0.2287	0.0955	0.0685
	100	-3.3874	0.0462	0.6993	0.2234	0.0963
	250	-3.3462	0.0503	1.0000	0.8747	0.3157
JO	50	15.9883	0.0478	0.1072	0.0612	0.0529
	100	15.4493	0.0479	0.3533	0.1149	0.0666
	250	15.2203	0.0489	0.9966	0.5958	0.1701

Panel B

Error Distribution: alpha = c(1.7, 1.7)						
Method	t	CV	Size	0.8	0.9	0.95
LAD	50	-3.6671	0.0529	0.2893	0.1385	0.0780
	100	-3.4183	0.0491	0.6668	0.3555	0.1677
	250	-3.0954	0.0490	0.9969	0.8873	0.5729
Rank	50	-3.3428	0.0503	0.4114	0.1724	0.0818
	100	-3.2155	0.0536	0.9154	0.5184	0.2168
	250	-3.1421	0.0541	0.9999	0.9922	0.7638
OLS	50	-3.4652	0.0519	0.1999	0.0874	0.0580
	100	-3.4146	0.0522	0.6849	0.1963	0.0789
	250	-3.3885	0.0532	0.9991	0.8651	0.2740
JO	50	16.7684	0.0550	0.0874	0.0566	0.0491
	100	16.1619	0.0535	0.2945	0.0973	0.0577
	250	16.0871	0.0495	0.9910	0.4958	0.1246

Panel C

Error Distribution: Skew alpha=1.7, beta= -0.25						
Method	t	CV	Size	0.8	0.9	0.95
LAD	50	-3.7416	0.0469	0.2760	0.1350	0.0741
	100	-3.4350	0.0469	0.6558	0.3446	0.1671
	250	-3.1023	0.0465	0.9970	0.8846	0.5736
Rank	50	-3.3232	0.0515	0.4103	0.1719	0.0863
	100	-3.2654	0.0487	0.9031	0.4934	0.2137
	250	-3.1514	0.0489	1.0000	0.9923	0.7595
OLS	50	-3.4661	0.0495	0.2008	0.0849	0.0608
	100	-3.4563	0.0477	0.6604	0.1759	0.0792
	250	-3.4412	0.0475	0.9992	0.8503	0.2536
JO	50	16.7270	0.0488	0.0879	0.0591	0.0508
	100	16.2613	0.0495	0.2841	0.0897	0.0529
	250	15.8871	0.0519	0.9909	0.5164	0.1389

Table 3. Simulation results, dispersion ratio = 4

Panel D

Error Distribution: alpha = c(1.5, 1.5)						
Method	t	CV	Size	0.8	0.9	0.95
LAD	50	-3.7132	0.0521	0.4380	0.2135	0.0991
	100	-3.2826	0.0479	0.8798	0.5891	0.3183
	250	-2.9145	0.0516	0.9996	0.9893	0.8570
Rank	50	-3.4267	0.0508	0.5476	0.2415	0.1016
	100	-3.1765	0.0542	0.9809	0.7413	0.3892
	250	-3.0087	0.0555	0.9994	0.9987	0.9498
OLS	50	-3.5742	0.0496	0.1565	0.0658	0.0447
	100	-3.4334	0.0584	0.6685	0.1812	0.0789
	250	-3.4508	0.0529	0.9974	0.8445	0.2310
JO	50	18.0358	0.0537	0.0678	0.0489	0.0439
	100	17.1151	0.0518	0.2283	0.0805	0.0534
	250	16.8278	0.0542	0.9865	0.4251	0.1041

Panel E

Error Distribution: Skew alpha=1.5, beta= -0.25						
Method	t	CV	Size	0.8	0.9	0.95
LAD	50	-3.7252	0.0520	0.4374	0.2106	0.1007
	100	-3.2833	0.0480	0.8757	0.5917	0.3106
	250	-2.8691	0.0530	1.0000	0.9901	0.8620
Rank	50	-3.3860	0.0510	0.5645	0.2565	0.1091
	100	-3.1752	0.0513	0.9815	0.7448	0.3827
	250	-2.9879	0.0520	0.9999	0.9991	0.9538
OLS	50	-3.5683	0.0474	0.1643	0.0649	0.0484
	100	-3.4858	0.0493	0.6438	0.1646	0.0702
	250	-3.4065	0.0522	0.9975	0.8749	0.2538
JO	50	17.9290	0.0509	0.0773	0.0551	0.0463
	100	17.0147	0.0554	0.2357	0.0814	0.0530
	250	16.8000	0.0458	0.9875	0.4148	0.1043

CHAPTER 5. APPLICATION

This chapter employs the rank based test statistic in a simple stock trading strategy. Statistical arbitrage, also sometimes referred to as pairs trading, is a heavily quantitative, econometric-based trading strategy. Its roots can be traced to the long-short mutual fund investment strategies of A. Winslow Jones in the 1950s. His idea was to create a hedged portfolio of long and short positions to eliminate market risk; the return on the portfolio came from the relative performance of the positions. This investment strategy involved buying undervalued stocks and selling short overvalued companies, a basic contrarian strategy that amounted to betting that the prices would converge, or to put the same thing another way, that the spread would shrink.

The strategy languished in obscurity for years until the mid-1980s when Nunzio Tartaglia assembled a group of traders at Mogan Stanley to create computer models that could efficiently access and analyze more data than previously possible. It was largely this increase in computing power, along with the sanction of several noted academics including Nobel laureates Myron Scholes and Robert Merton, that caused the renewed interest in the strategy. Furthermore, by using both long and short positions, this strategy does not require the calculation of a fundamental value for any securities; they need simply be separated into under- and over-valued subsets. Profits are made on relative performance, and not absolute returns (Vidyamurthy, 2004).

In order to gain a fuller understanding of statistical arbitrage, consider the cash flows involved, as illustrated in Figure C below. The investor is long one portfolio of stocks, which receive dividends and provide a capital gain or loss, while being short another portfolio of stocks on which the trader pays dividends and receives another capital

gain or loss. The investor also earns interest on any necessary collateral cash balance. So long as the prices of the long positions rise more (or drop less) than the short positions, the investor will have a net gain.

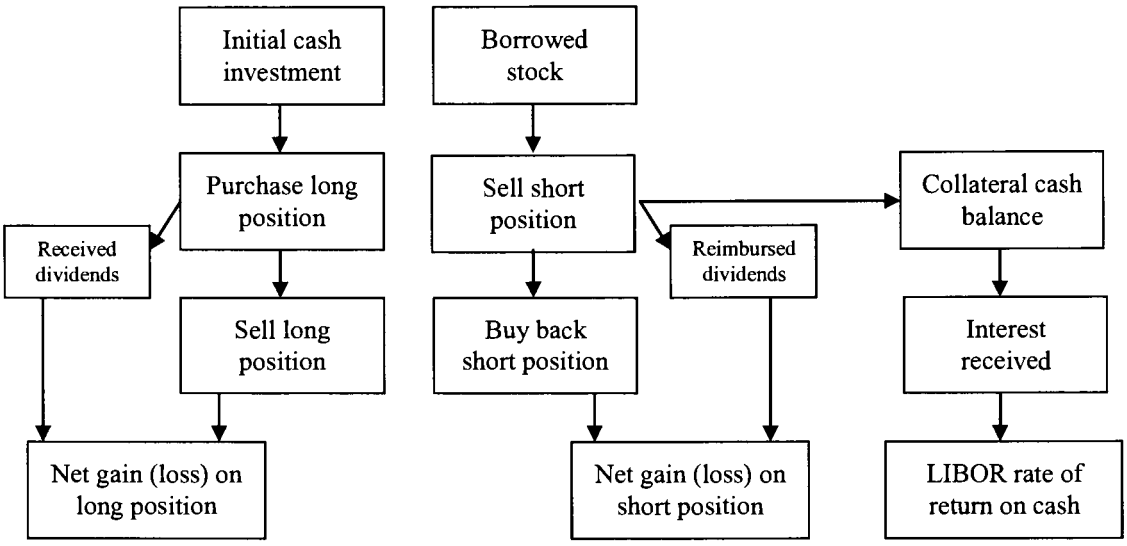

Figure C. Statistical arbitrage diagram

Though the statistical models employed in statistical arbitrage are unique to the individual trader or firm, they all follow the same basic format. First, an initial universe of candidate stocks is identified. This preliminary step allows a trader to focus on a particular industry or other subset that might be presumed to have a high degree of co-movement in stock returns. Second, candidate pairs are tested for cointegration. Third, the spread is calculated, and when the spread widens beyond some predefined threshold, the trader takes the appropriate long and short positions.

Data were available for 28 stocks from the Dow Jones Industrial Average over the entire period. These stocks were chosen in part for their familiarity and easy access to data. Furthermore, because they are all major American companies, there is a reasonable *a priori* expectation that the daily returns on their stocks should be cointegrated over some periods

and that the probability of cointegration is higher for such similar companies than it would be for a randomly selected pair of equities.

Stock return data were gleaned from the Center for Research in Security Prices (CRSP) database. The companies are Alcoa (AA), American Express (AXP), Bank of America (BAC), Boeing (BA), Caterpillar (CAT), Chevron Corporation (CVX), Coca-Cola (KO), DuPont (DD), 3M (MMM), AT&T (T), ExxonMobil (XOM), General Electric (GE), Hewlett-Packard (HPQ), The Home Depot (HD), Intel (INTC), IBM (IBM), Johnson & Johnson (JNJ), JPMorgan Chase (JPM), McDonald's (MCD), Merck (MRK), Microsoft (MSFT), Pfizer (PFE), Procter & Gamble (PG), Travelers (TRV), United Technologies Corporation (UTX), Verizon Communications (VZ), Wal-Mart (WMT), Walt Disney (DIS).

There are 378 potential trading pairs from the 28 stocks. Each of these pairs is tested using the Engle-Granger and rank-based cointegration tests for each trading year. Throughout this application, trading years were defined as 250 contiguous trading days. This definition does not coincide precisely with the calendar year, but it simplifies the calculations to have every period of identical length. Furthermore, the critical values employed were generated by the Monte Carlo simulation using time series of length 250.

The strategy of statistical arbitrage is played out in two periods. First, the identification period is a 250-day period in which the candidate pairs are tested for the existence of cointegration. All pairs that are determined to be cointegrated (at the 5% level) then enter the trading period, which lasts for another 250 days. The length of time in each period can be varied, and the strategy can be made considerably more complex. This simple two-period structure with equal time series length is provided as a case study

illustration to compare the two cointegration tests. This generic strategy is drawn from Whistler (2004).

Table 4 presents results of the rank and OLS (Engle-Granger) cointegration tests for all 16 periods. In every case, the rank-based test identifies more cointegrated pairs. The maximum number of identified pairs is in period 9, in which the rank test identifies 99 pairs (26.19% of all candidates) to enter the trading period; the minimum number occurs for the Engle-Granger test in period 8, in which only 16 pairs are identified (4.23%).

Table 4: Identification of candidate pairs for trading

Period	Days	Rank	OLS	Overlap
1	1 to 250	38	36	34
2	251 to 500	58	54	46
3	501 to 750	44	26	25
4	751 to 1000	57	34	28
5	1001 to 1250	65	28	25
6	1251 to 1500	58	44	39
7	1501 to 1750	50	38	30
8	1751 to 2000	25	16	12
9	2001 to 2250	99	75	70
10	2251 to 2500	49	47	33
11	2501 to 2750	71	51	45
12	2751 to 3000	38	17	14
13	3001 to 3250	37	26	22
14	3251 to 3500	12	4	2
15	3501 to 3750	78	60	40
16	3751 to 4000	6	5	4

One natural question is the extent to which the rank- and OLS-based procedures identify entirely different pairs compared to the degree of overlap between the two procedures. The last column of Table 4 provides the number of such overlaps, in which a candidate pair is identified as cointegrated by both tests. In every period, the number of overlaps is a significant percentage of the identified pairs. In no period is the overlap less than 50% of the OLS pairs, and that occurs in period 14, in which the OLS procedure

identifies only 4 pairs. The next lowest percentages are 67% in period 15 and 70.2% in period 10. As one would hope, there appears to be little loss in trading opportunities by using the rank-based methodology. Furthermore, if a trader were so inclined, he or she could employ both methods to find the greatest possible number of trading opportunities. The primary point, however, is that in this sample, the rank based-method always provides a greater list of stock pairs to be considered for trading and therefore a greater potential for profit.

In total, there are 877 unique trading opportunities to examine in the second, trading step. To calculate trading profits, the first step is to calculate the difference between the two stocks, referred to as the spread. When the spread exceeds its historical average, trades are placed that will profit if the spread reverts to its average. The algorithm employed is based on a 250-day trailing average and standard deviation of the spread. Trades are placed when the spread deviates more than two standard deviations from its 250-day trailing average. Note that this may occur multiple times for the same pair of stocks, as the spread expands, reverts, and expands again.

Positions are closed for a profit when the spread reverts to the mean or closed as a stop-loss condition if the spread widens to exceed three standard deviations. This loss prevention rule is an important aspect of the strategy; it is activated in 5.7% of the trades. All positions are also closed out at the end of the 250 day trading period. Only 1.9% of the candidate pairs reach the end of the period without reverting to the mean, and all of them are profitable at the time of being closed out.

In general, risk and return of statistical arbitrage are affected by a number of factors, including exposure to certain industries and sectors, position concentration or

diversification, use of cash, and leverage. Leverage, in particular, is a vital part of this strategy in order to make significant profits from relatively small spreads. This leverage adds greatly to the risk, however, such that it is sometimes described as picking up nickels in front of bulldozers (Lowenstein, 2001). The scheme employed here is a simple application of the general strategy and is not meant to be realistic, but functions to demonstrate one potential application of the rank-based statistic.

Profits were calculated for all trading opportunities, scaled to a $1 investment for every traded pair. Then compounded annual returns were calculated for the rank-based and OLS identified pairs. Using only the pairs identified by OLS yielded an annual return of 10.81%, in comparison to 12.05% for the pairs from the rank-based test. While both figures represent strong investment returns, the rank test is clearly superior. In terms of total dollars earned in the seventeen year period, the rank test provides 20.8% more earnings. This example provides one application of the test that demonstrates its utility and practicality.

One possible extension of this example would be a high-frequency pairs trading scheme. Such a strategy, which does not allow overnight holding periods, may increase the profits due to increased trading opportunities or diminish the profitability due to increased trading. In general, the optimal sampling frequency is becoming a more pervasive issue now that tick level data is becoming more commonly available. One application has been estimating covariance using high-frequency data (Pooter, Martens, and van Dijk, 2008). Cointegration testing in a high-frequency setting will be an important empirical question as data frequency and availability increase.

CHAPTER 6. CONCLUSION

This paper has developed two new test statistics for cointegration, a rank-based and LAD-based version of the seminal Engle-Granger test. The superiority of the rank-based test, in particular, over the Engle-Granger and Johansen procedures was demonstrated in a wide range of Monte Carlo simulations. The application of statistical arbitrage was presented as one example of a particular case in which the test may be both useful and profitable.

Future work could explore additional applications of this test. A number of heretofore unexplained cointegration results may be due simply to the low power of previous tests. For instance, purchasing power parity is found to be violated in empirical work surprisingly often. Another example would be price discovery and other market microstructure issues, such as the components of the bid-ask spread. Reviewing some of the previous literature, particularly in the field of finance, with this new test might yield interesting results and potentially some changes in conclusions or interpretations.

Natural extensions for this statistic in the field of finance are examinations of high frequency time series and issues of non-synchronicity. Scholes and Williams (1977) and Miller, Muthuswamy, and Whaley (1994) are some of the prominent papers that have explored the problem of non-synchronous financial price series. How will the proposed test statistic function in such a setting? That question is left for future research.

REFERENCES

Agiakloglou, C., & Newbold, P. (1992). Empirical evidence on dickey-fuller-type tests. *Journal of Time Series Analysis, 13,* 471-483.

Aparicio, F., Escribano, A., & Sipols, A. (2006). Range unit-root tests: Robust against nonlinearities, error distributions, structural breaks, and outliers. *Journal of Time Series Analysis, 27,* 545-576.

Banerjee, A., Dolado, J., Galbraith, J., & Hendry, D. (1993). *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data,* Oxford: Oxford University Press.

Banerjee, A., Dolado, J., Hendry, D., & Smith, G. (1986). Exploring equilibrium relationships in econometrics through static models: Some Monte Carlo evidence. *Oxford Bulletin of Economics and Statistics, 48,* 253-278.

Bierens, H. (1997). Nonparametric cointegration analysis. *Journal of Econometrics, 77,* 379-404.

Breitung, J., & Gourieroux, C. (1997). Rank tests for unit roots. *Journal of Econometrics, 81,* 7-27.

Choi, I. (1992). Effects of data aggregation on the power of tests for a unit root: A simulation study. *Economics Letters, 40,* 397-401.

Cochrane, J. (1988). How big is the random walk in GNP? *Journal of Political Economy, 96,* 893-920.

DeJong, D., Nankervis, J., Savin, N., & Whiteman, C. (1992). The power problems of unit root tests for time series with autoregressive errors. *Journal of Econometrics, 53,* 323-343.

Dickey, D., & Fuller, W. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74,* 427-431.

Dickey, D., & Fuller, W. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica, 49,* 1057-1072.

Engle, R., & Granger, C. (1987). Cointegration and error correction: Representations, estimation, and testing. *Econometrica, 55,* 252-276.

Escribano, A., Santos, M., & Sipols, A. (2008). Testing for cointegration using induced-order statistics. *Computational Statistics, 23,* 131-151.

Escribano, A., Sipols, A., & Aparicio, F. (2006). Nonlinear cointegration and nonlinear error correction: Record counting cointegratino tests. *Communications in Statistics: Simulation & Computation, 35,* 939-956.

Fama, E. (1965). The behavior of stock prices. *Journal of Business, 38,* 34-105.

Fotopoulos, S., & Ahn, S. (2003). Rank based Dickey-Fuller test statistics. *Journal of Time Series Analysis, 24,* 647-662.

Granger, C., & Newbold, P. (1974). Spurious regression in econometrics. *Journal of Econometrics, 2,* 111-120.

Granger, C. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics, 16,* 121-130.

Granger, C., & Hallman, J. (1991). Nonlinear transformations of integrated time series. *Journal of Time Series Analysis, 12,* 207-224.

Granger, C., & Orr, D. (1972). "Infinite variance" and research strategy in time series analysis. *Journal of the American Statistical Association, 67,* 275-285.

Gutenbrunner, C., & Jureckova, J. (1992). Regression rank scores and regression quantiles. *Annals of Statistics, 20,* 305-330.

Hall, A. (1989). Testing for a unit root in the presence of moving average errors. *Biometrika, 76,* 49-56.

Hall, A. (1994). Testing for a unit root in time series with pretest data-based model selection. *Journal of Business and Economic Statistics, 12,* 461-470.

Hansen, B., & Phillips, P. (1990). Estimation and inference in models of cointegration: A simulation study. in Fomby, T. & Rhodes, G. (eds.), *Advances in Econometrics: Cointegration, unit roots, and spurious regression, vol. 8,* JAI Press.

Hasan, M. (2001). Rank tests of unit root hypothesis with infinite variance errors. *Journal of Econometrics, 104,* 49-65.

Hasan, M., & Koenker, R. (1997). Robust rank tests of the unit root hypothesis. *Econometrica, 65,* 133-161.

Hendry, D. (1986). Econometric modeling with cointegrated variables: An overview. *Oxford Bulletin of Economics and Statistics, 48,* 201-212.

Herce, M. (1996). Asymptotic theory of LAD estimation in a unit root process with finite variance errors. *Econometric Theory, 12,* 129-153.

Hettmansperger, T. (1984). *Statistical inference based on ranks.* New York: Wiley.

Hettmansperger, T., & McKean, J. (1998). *Robust nonparametric statistical methods.* New York: Wiley.

Hoek, J., Lucas, A., & van Dijk, H. (1995). Classical and Bayesian aspects of robust unit root inference. *Journal of Econometrics, 69,* 27-59.

Hogg, R., McKean, J., & Craig, A. (2005). *Introduction to Mathematical Statistics, 6th edition,* Upper Saddle River, NJ: Pearson Prentice Hall.

Hooker, R. (1901). Correlation of the marriage-rate with trade. *Journal of the Royal Statistical Society.*

Jaeckel, L. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Annals of Mathematical Statistics, 43,* 1449-1458.

Janicki, A., & Weron, A. (1994). Can one see α-stable variables and processes? *Statistical Science, 9,* 109-126.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control, 12,* 231-254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica, 59,* 1551-1580.

Johansen, S., & Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration with applications to the demand for money. *Oxford Bulletin of Economics and Statistics, 52,* 169-210.

Koenker, R. (1997). 8 Rank tests for linear models. *Handbook of statistics, 15,* 175-199.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica, 46,* 33-50.

Kremers, J., Ericsson, N., & Dolado, J. (1992). The power of cointegration tests. *Oxford Bulletin of Economics and Statistics, 54,* 325-348

Kwiatkowski, D., Phillips, P., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics, 54,* 159-178.

Leybourne, S., & McCabe, B. (1994). A consistent test for a unit root. *Journal of Business & Economics Statistics, 12,* 157-166.

Lloyd, C. (2005a). Estimating test power adjusted for size. *Journal of Statistical Computation and Simulation, 75,* 921-934.

Lloyd, C. (2005b). On comparing the accuracy of competing tests of the same hypotheses from simulation data. *Journal of Statistical Planning and Inference, 128,* 497-508.

Lo, A., & MacKinlay, A. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies, 1,* 41-66.

Lowenstein, R. (2001). *When Genius Failed: The Rise and Fall of Long-Term Capital Management,* New York: Random House.

Lucas, A. (1995a). An outlier robust unit root test with an application to the extended Nelson-Plosser data. *Journal of Econometrics, 66,* 153-173.

Lucas, A. (1995b). Unit root tests based on M estimators. *Econometric Theory, 11,* 331-346.

Maddala, G., & Kim, I. (1998). *Unit Roots, cointegration, and structural change.* Cambridge: Cambridge University Press.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business, 36,* 394-419.

Miller, M., Muthuswamy, J., & Whaley, R. (1994). Mean reversion in of Standard & Poor's 500 Index basis changes: Arbitrage-induced or statistical illusion. *Journal of Finance, 49,* 479-513.

Nelson, C., & Plosser, C. (1982). Trends and random walks in macroeconomic time series. *Journal of Monetary Economics, 10,* 139-162.

Ng, S., & Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association, 90,* 268-281.

Ng, S., & Perron, P. (1997). Estimation and inference in nearly unbalanced nearly cointegrated systems. *Journal of Econometrics, 79,* 53-81.

Nolan, J. (2003). Modeling financial data with stable distributions. *Handbook of Heavy Tailed Distributions in Finance,* ed. Rachev, S., Amsterdam: Elsevier Science B.V.

Osterwald-Lenum, M. (1992). A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics. *Oxford Bulletin of Economics and Statistics, 54,* 461-472.

Park, J. (1992). Canonical cointegrating regressions. *Econometrica, 60,* 119-143.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica, 57,* 1361-1401.

Pfaff, B. (2006). *Analysis of integrated and cointegrated time series with R, First Edition.* New York City: Springer.

Phillips, P. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika, 74,* 535-547.

Phillips, P. (1995). Fully modified least squares and vector autoregression. *Econometrica, 63,* 1023-1078.

Phillips, P., & Hansen, B. (1990). Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies, 57,* 99-125.

Phillips, P., & Loretan, M. (1991). Estimating long-run economic equilibria. *Review of Economic Studies, 58,* 407-436.

Phillips, P., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika, 75,* 335-346.

Pooter, M., Martens, M., & van Dijk, D. (2008). Predicting the daily covariance matrix for S&P 100 stocks using intraday data – but which frequency to use? *Econometric Reviews, 27,* 199-229.

Rachev, S., & Mittnik, S. (2000). *Stable Paretian Models in Finance.* New York City: Wiley.

Rothenberg, T., & Stock, J. (1997). Inference in a nearly integrated autoregressive model with nonnormal innovations. *Journal of Econometrics, 80,* 269-286.

Said, S., & Dickey, D. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrica, 71,* 599-607.

Sargan, J., & Bhargava, A. (1983). Testing residuals from least squares regression for being generated by the Gaussian random walk. *Econometrica, 51,* 153-174.

Scholes, M., & Williams, J. (1977). Estimating betas from nonsynchronous data. *Journal of Financial Economics, 5,* 309-327.

Schwert, G. (1989). Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics, 7,* 147-159.

Stigler, S. (1981). Gauss and the invention of least squares. *Annals of Statistics, 9,* 465-474.

Stock, J. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica, 55,* 1035-1056.

Stock, J. (1994). Unit roots, structural breaks, and trends. *Handbook of Econometrics, vol 4,* Engle, R. & McFadden, D. (eds.), Amsterdam: Elsevier.

Terpstra, J., & McKean, J. (2005). Rank-based analyses of linear models using R. *Journal of Statistical Software, 14,* 1-26.

Terpstra, J., McKean, J., & Naranjo, J. (2001). Weighted Wilcoxon estimates for autoregression. *Australian and New Zealand Journal of Statistics, 43,* 399-419.

Vidyamurthy, G. (2004). *Pairs Trading: Quantitative Methods and Analysis,* Hoboken, New Jersey: Wiley and Sons, Inc.

Whistler, M. (2004). *Trading Pairs,* New York: Wiley and Sons, Inc.

Wilcox, R. (1997). *Introduction to Robust Estimation and Hypothesis Testing.* San Diego: Academic Press.

Yule, G. (1926). Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time series. *Journal of the Royal Statistical Society, 89,* 1-63.

APPENDIX

The R code below was employed in the Monte Carlo simulations to create critical values and test for size and power. The two functions are COINT.SIM and MY.DF.TEST. The code requires R packages fUnitRoots, urca, tseries, tframe, quantreg, MASS, SparseM, as well as the code from Terpstra and McKean (2005) that can be found at http://www.stat.wmich.edu/mckean/HMC/Rcode/AppendixB/ww.r

```
coint.sim=function(t=25,alpha=c(2,2),beta=c(0,0),gamma=c(sqrt(1/2),sqrt(1/2)),
                delta=c(0,0),rho=0.9,B=2,N=1,lags=1,pvalue=.05,coint=T,
                        cv.lad=qnorm(pvalue),cv.rank=qnorm(pvalue),cv.ls=qnorm(pvalue),
                cv.jo=0) {
 res.rank=rep(NA,N)
 rej.rank=rep(NA,N)
 res.ls=rep(NA,N)
 rej.ls=rep(NA,N)
 res.jo=rep(NA,N)
 rej.jo=rep(NA,N)
 res.lad=rep(NA,N)
 rej.lad=rep(NA,N)
 t=t+20 #Used to remove "start-up" effects
 for (i in 1:N){
  #Generate the two time series
  if (coint==T){
   u1=cumsum(rstable(n=t,alpha=alpha[1],beta=beta[1],delta=delta[1],gamma=gamma[1]))
```

```
  u2=arima.sim(n=t,list(ar=rho),innov=rstable(n=t,alpha=alpha[2],beta=beta[2],
        delta=delta[2],gamma=gamma[2])@.Data)
  u1=u1[21:t]
  u2=u2[21:t]
  y1=(B/(B-1))*u1-(1/(B-1))*u2
  y2=-(1/(B-1))*u1+(1/(B-1))*u2
 }
 if (coint==F){
  y1=cumsum(rstable(n=t,alpha=alpha[1],beta=beta[1],delta=delta[1],gamma=gamma[1]))
  y2=cumsum(rstable(n=t,alpha=alpha[2],beta=beta[2],delta=delta[2],gamma=gamma[2]))
  y1=y1[21:t]
  y2=y2[21:t]
 }
###LAD test
  ladreg=rq(y2~y1)    #Fits an LAD regression
  err.lad=ladreg$residuals
  stat.lad=my.df.test(err.lad,p=lags,mu=T,type="LAD")$tau #ADF test with intercept and
no time trend
  names(stat.lad)=NULL
  res.lad[i]=stat.lad
  if (stat.lad<cv.lad) rej.lad[i]=1
   else rej.lad[i]=0
###Rank based test
```

```
  wilreg=wwfit(y1,y2)  #Fits a simple linear regression *with* intercept
  err.rank=wilreg$residuals
  stat.rank=my.df.test(err.rank,p=lags,mu=TRUE,type="WIL")$tau #ADF test with
intercept and no time trend
  names(stat.rank)=NULL
  res.rank[i]=stat.rank
  if (stat.rank<cv.rank) rej.rank[i]=1
   else rej.rank[i]=0
###OLS test
  reg=lm(y2~y1)   #Fits a simple linear regression *with* intercept
  err.ls=resid(reg)
  stat.ls=my.df.test(err.ls,p=lags,mu=TRUE,type="LS")$tau #ADF test with intercept and
no time trend
  names(stat.ls)=NULL
  res.ls[i]=stat.ls
  if (stat.ls<cv.ls) rej.ls[i]=1
   else rej.ls[i]=0
###JO test
  system=cbind(y1,y2)
  stat.jo=ca.jo(system)
  res.jo[i]=stat.jo@teststat[2]
  if (stat.jo@teststat[2]>cv.jo) rej.jo[i]=1
   else rej.jo[i]=0  }
```

```
ans.lad=mean(rej.lad)
 ans.rank=mean(rej.rank)
 ans.ls=mean(rej.ls)
 ans.jo=mean(rej.jo)
 names(ans.lad)='LAD test'
 names(ans.rank)='Rank test'
 names(ans.ls)='OLS test'
 names(ans.jo)='JO test'
 ans=list(ans.lad,ans.rank,ans.ls,ans.jo)
 return(list(TS.lad=res.lad,TS.rank=res.rank, TS.ls=res.ls, TS.jo=res.jo,
        Q05.lad=quantile(res.lad,.05),Q05.rank=quantile(res.rank,.05),Q05.ls=quantile(res.l
s,.05),Q05.jo=quantile(res.jo,.95),
        ERR.lad=ans.lad,ERR.rank=ans.rank,ERR.ls=ans.ls, ERR.jo=ans.jo))
}
my.df.test=function(x, p=1, mu=TRUE, type="LS") {
  n <- length(x)
  d <- diff(x)
  Xtm1 <- x[1:(n-1)]
  if(p > 1)
    z <- cbind(Xtm1,input(d,p-1))
  else
    z <- cbind(Xtm1,d)
  z <- z[!apply(z,1,function(x){any(is.na(x))}),]
```

```
if (type=="LS") {
  lsest <- lsfit(z[,1:(dim(z)[2]-1)],z[,dim(z)[2]],intercept=mu)
  res <- lsest$resid
  lsest <- ls.print(lsest,print.it=F)
  tau <- ifelse(mu,lsest[[2]][[1]][2,3],lsest[[2]][[1]][1,3])
 }
 #WIL rank-based version of DF test
 if (type=="WIL") {
  wilest <- wwest(z[,1:(dim(z)[2]-1)],z[,dim(z)[2]],print.tbl=F)
  res <- wilest$tmp1$residuals
  tau <- wilest$ans[2,3]
 }
 #LAD version of DF test
 if (type=="LAD") {
  ladest <- rq(z[,dim(z)[2]]~z[,1:(dim(z)[2]-1)])
  res <- ladest$residuals
  temp <- summary(ladest,se='iid')
  tau <- temp$coefficients[2,3]
 }
 list(tau=tau,xy=z,r=res)
}
```