

STATISTICAL ARBITRAGE:
A STUDY IN TURKISH EQUITY MARKET

ÖZKAN ÖZKAYNAK
103622017

ISTANBUL BILGI UNIVERSITY
INSTITUTE OF SOCIAL SCIENCES
MASTER OF SCIENCE IN ECONOMICS

UNDER SUPERVISION OF
DOÇ. DR. EGE YAZGAN
2007

STATISTICAL ARBITRAGE:
A STUDY IN TURKISH EQUITY MARKET

İstatistiksel Arbitraj:
Türk Hisse Senedi Piyasasında Bir Çalışma

ÖZKAN ÖZKAYNAK
103622017

Doç. Dr. Ege Yazgan :
Prof. Dr. Burak Saltoğlu :
Yrd. Doç. Dr. Koray Akay :

Tezin Onaylandığı Tarih :

Toplam Sayfa Sayısı : 55

Anahtar Kelimeler:

- 1) İstatistiksel Arbitraj
- 2) İkili Alım-Satım
- 3) Kantitatif Portföy
- 4) Nötr Piyasa Portföyü
- 5) Kointegrasyon

Keywords:

Statistical Arbitrage
Pairs Trading
Quantitative Portfolio
Market Neutral Portfolio
Cointegration

ABSTRACT

Statistical Arbitrage is an attempt to profit from pricing inefficiencies that are identified through the use of mathematical models. One technique is Pairs Trading, which is a non-directional strategy that identifies two stocks with similar characteristics whose price relationship is outside of its historical range. The strategy simply buys one instrument and sells the other in hopes that relationship moves back toward normal. The idea is the price relationship between two related instruments tends to fluctuate around its average in the short term, while remaining stable over the long term. From the academic view of weak market efficiency theory, pairs trading shouldn't work since the actual price of a stock reflects its past trading data, including historical prices. This leaves us the question: Does a statistical arbitrage strategy, pairs trading, work for the Turkish stock market? The main objective of this research is to verify the performance and risks of pairs trading in the Turkish equity market. The main conclusion is that pairs trading may be a profitable strategy in the Turkish Market. Such profitability was found consistent over different time frames. Another result of the research is that integrating each stock's fundamentals (P/E, price-to-book, and market capitalization, dividend yield, cash position etc.) to the pure quantitative trading strategy improves the backtesting results.

ÖZETÇE

İstatistiksel Arbitraj, matematiksel modellerin kullanılmasıyla fiyatlardaki etkinsizliğin belirlenerek bu fırsattan getiri elde edilmesidir. İstatistiksel Arbitraj yöntemlerinden bir tanesi İkili Alım-Satımdır. Bu strateji, aralarında istatistiksel ilişki olan iki hissenin belirlenmesi ve bu hisselerin birbirlerine göre fiyat hareketlerindeki sapmalardan faydalanarak aynı anda kısa ve uzun pozisyonlarla kar edilmesini amaçlar. Bu çalışmada, Türkiye hisse senedi piyasasında ikili alım-satım yönteminin sınanması amaçlanmıştır. Çalışmadan çıkarılabilecek temel sonuç, bu stratejinin Türkiye piyasasında karlı bir metodoloji olabileceği yönündedir. Diğer bir sonuç ise, şirketlerin temel rasyolarının da kantitatif yöntemlere entegre edilmesiyle daha iyi getiri rakamlarına ulaşabileceğidir.

ACKNOWLEDGEMENTS

I would like to express my gratefulness to my brilliant advisor Ege Yazgan, for his ideas and guidance throughout the work at Istanbul Bilgi University. Especially I would like to thank him for helping me on understanding all the financial theories, which give me a lot of inspirations and thoughts. I would also like to acknowledge the support of my professors and friends at Istanbul Bilgi University MSc in Economics Department. Thank you all.

Table of Contents

1. INTRODUCTION.....	6
1.1 Statistical Arbitrage and Market Efficiency	
1.2 Statistical Arbitrage and Pairs Trading	
1.2.1 Pairs Trading History	
1.2.2 Pairs Trading Background	
1.2.3 Cointegration Based Strategies	
2. METHODOLOGY OF THE RESEARCH.....	20
2.1 Cointegration Methodology	
2.2 Data	
2.3 Pairs Selection	
2.4 Trading Rules	
2.5 Performance of the Strategy	
3. CONCLUSION AND FURTHER RESEARCH.....	47
4. REFERENCES.....	52

1. INTRODUCTION

1.1 Statistical Arbitrage and Market Efficiency

In 1900, Louis Bachelier described the variation of stock price using Brownian Motion in his dissertation. [Bachelier, (1900)] He is the first to anticipate much of what later became standard fare in the financial theory: Random Walk of financial market prices, Brownian Motion and Martingales, before both Einstein and Weiner. [Einstein, (1956)]

This first paper in the history of Mathematical Finance was widely recognized only in the 1950s. The modernization of finance would date to the year 1952 with the publication in Journal of Finance of Harry Markowitz's article, "Portfolio Selection". In this paper, Markowitz gave a precise definition of risk and return, as the mean and variation of the outcome of an investment. [Markowitz, (1952)]

This made the powerful methods of mathematical statistics available for the study of strategy of portfolio selection. An issue that is the subject of intense debate among academics and financial professionals is the Efficient Market Hypothesis (EMH), which says that security prices fully reflect all available information at any time.

This implies that there is no arbitrage opportunity in a perfectly "efficient" market. One can neither buy securities which are worth more than the selling price, nor sell securities worth less than the selling price.

A significant development of EMH in 1960's by Eugene Fama and his later work in 1998 asserts that price movements in the market are unpredictable. The Random Walk theory can be connected to Bachelier's work in 1900. [Fama, (1965)]

Over more than half a century, much empirical research was done on testing the market efficiency, which was traced to 1930's by Alfred Cowles. Many studies have found that stock prices are at least partially predictable. The contradiction to market efficiency gives the

possibilities to search for a statistical arbitrage opportunity. S.Hogan, R.Jarrow and M.Warachka demonstrate a method to test the existence of statistical arbitrage, and proved that it is incompatible with market efficiency.

Statistical Arbitrage is a heavily quantitative and computational approach to equity trading. It describes a variety of automated trading systems which commonly make use of data mining, statistical methods and artificial intelligence techniques. A popular strategy is “pairs trading”, in which stocks are put into pairs by fundamental or market based similarities.

One stock in the pair is bought long, the other is sold short. This strategy hedges risk from whole market movements. First, we find two securities in the same industry/sector which have historically traded in a certain range but which are now at an extreme.

Based on analysis of the relative prospects for each company, we may determine that the relative price relationship appears to be wrong. We will short the shares of the more expensive stock and buy an equal amount of the other. As the pair moves toward their norm, gains will be made from both the long and the short. This strategy is a low volatility one that is little affected by market direction.

This study implements the statistical method from and experiments on real market data. Before we come to the description of the test, some background discussion about arbitrage and statistical arbitrage is necessary.

An arbitrage is a transaction or portfolio that makes a profit without risk. A portfolio is said to be an arbitrage if it costs nothing to implement, has a positive probability of a positive payoff and a zero probability of a negative payoff. Loosely speaking, “buy low” and “sell high” trade.

To relax the condition of arbitrage, we can define a statistical arbitrage; in other words, an arbitrage is a special case of statistical arbitrage. One major distinction is that a statistical arbitrage is not riskless. Like an arbitrage, a statistical arbitrage costs nothing to implement.

However, it has a positive expected payoff and a zero probability of a negative payoff only as time approaches infinity, and its variance vanishes at time infinity. The test for statistical arbitrage opportunity is applied to one trading strategy; specifically, the profits generated from the trading strategy every business day.

This evaluation of profits of the trading strategy is based on a period of market data from the current trading day. It is believed that a statistical arbitrage opportunity appears as an abnormal behavior in the market. It then becomes essential to distinguish whether an abnormal market figure is a “potential opportunity” or simply erroneous data.

The market efficiency theory has been tested by different type of research. Such concept provides, on its weak form, that the past trading information of a stock is fully reflected on its value, meaning that past data has no potential for predicting future behavior of asset’s prices.

The main theoretical consequence of this concept is that no logical rules of trading based on historical data should have a significant and positive excessive return over a benchmark portfolio.

In opposition to the market efficiency theory, several papers have showed that past information is able to explain future stock market returns. Such predictability can appear in different ways, including time anomalies (day of the week effect) [French (1980)] and correlation between asset’s returns and other variables. [Fama and French, (1992)].

A respectable amount of papers have tried to use quantitative tools in order to model market and building trading rules based on that. The basic idea of this type of research is to look for some kind of pattern in the historical price behavior and using only historical information, take such pattern into account for the creation of long and short trading positions.

One of the most popular approaches to model the market and infer logical rules is technical analysis. [Murhpy, (1999)] Such technique is based on quantitative indicators and also visual patterns in order to identify entry and exit points on the short term behavior of stock prices. The popularization of technical analysis leads to a number of tests that had as objective to verify if such tools were profitable or not.

It is worth to say that even though the majority of papers have showed that technical analysis is profitable, several problems can be addressed with such studies, including data snooping problems, transaction costs and liquidity. All this incompleteness of the research still makes technical analysis a subject to be studied. With the advent of computer power, more sophisticated mathematical methods could be employed in case of trading rules.

A popular strategy that has made its reputation in the early 80's is the so called pairs trading. Such methodology as designed by a team of scientists from different areas, which were brought together. The main objective of such team was to use statistical methods to develop computer based trading platforms, where the human subjectivity had no influence whatsoever in the process of making the decision to buy and sell a particular stock.

Basically, the idea of pairs trading is to take advantage of market inefficiencies. The first step is to identify two stocks that move together and trade them every time the absolute distance between the price paths is above a particular threshold value. If the stocks, after the divergence, return to the historical behavior, then is expected that the one with the highest price is going to have a decrease in value and the one with the lowest price has an increase. All long and short position are taken according with this logic.

The main objective of this research is to investigate the profitability and risk of the pairs trading strategy in the Turkish equity market.

The paper is organized as follows: The first part introduces general concept of statistical arbitrage, and its special form, pairs trading. The second part provides the main guidelines of the methodology, including the way the pairs are going to be formed, the logical rules of trading and performance assessment. The results and its discussion are made and after that the paper finishes with some concluding remarks and further research suggestions.

1.2 Statistical Arbitrage and Pairs Trading

1.2.1 Pairs Trading History

The Wall Street Quant Nunzio Tartaglia gathered a team of physicist, mathematicians and computer scientists to research arbitrage opportunities in the equities markets in the in the mid-1980's. Tartaglia's group used sophisticated statistical methods to develop high technology trading programs, executable through automated trading systems. Tartaglia's programs also identified pairs of securities whose prices tended to move together.

They traded these pairs with great success in 1987 and made a \$50million for the firm. Although the Morgan Stanley group disbanded in 1989 after a couple of bad years of performance, pairs trading has since become an increasingly popular market-neutral investment strategy used by institutional traders as well as hedge fund managers. The increased popularity of quantitative-based statistical arbitrage strategies has also apparently affected profits. [Saul, (1989)]

In this paper, we examine the risk and return characteristics of pairs trading with daily data over the period 2003-2007. Using a simple algorithm for choosing pairs, we test the profitability of several straightforward, self financing trading rules. We find average annualized return of about 30.87% percent for pairs trading portfolio in Turkey.

Pairs trading has recently been the subject of academic interest. Gatev, Goetzmann and Rouwenhorst (1999) present evidence that this simple trading strategy produced statistically significant excess returns for the period 1963-1997. Zebedee (2001) analyzes the impact of pairs-trading at the microstructure level within the airline industry.

Pairs trading was also used in different fields, beyond equities. For instance, Kato, Linn and Schallheim (1991) and Wahab, Lashgari, and Cohn (1992) studied arbitrage opportunities in the ADR market, and found very little evidence for profitable opportunities in the ADR market. In particular, Wahab et al. (1992) followed an implicit pairs trading strategy with two portfolios: an ADR portfolio and an underlying shares portfolio. They sell the portfolio with the highest returns over a period of two weeks and buy the portfolio with lowest returns over the same two-week

period. They found limited profits for their pairs trading strategy, and they attributed their small profits, around 4%, to transaction costs and data limitations. The conclusion was that pairs trading using ADRs do not seem to be profitable.

In the academic literature on the U.S. Treasury securities market, Krishnamurthy (2000) examines the classic trade involving a short position in a newly issued 30-year Treasury bond and a long position in the old 30-year Treasury bond. He estimates that the profits from this strategy are greatly reduced once the cost of financing in the repo markets is taken into account.

Gatev et al (1999) examine pairs trading in the U.S. equity market. They confirm that this popular Wall Street investment strategy is profitable after an allowance for trading costs, and that these profits are inherently different from a pure mean-reversion strategy. But, they do not address the difficulties or costs of shorting in that market.

Most referenced works in pairs trading include Gatev, Goetzmann and Rouwenhorst (1999), Vidyamurthy (2004), and Elliott, van der Hoek and Malcolm (2005).

The first paper is an empirical piece of research that, using a simple standard deviation strategy, shows pairs trading after costs can be profitable. The second of these papers details an implementation strategy based on a cointegration based framework, without empirical results. The last paper applies a Kalman filter to estimating a parametric model of the spread. These methods can be shown to be applicable for special cases of the underlying equilibrium relationship between two stocks.

To define the boundary of this research, it is necessary to identify pairs trading relative to other seemingly related hedge fund strategies. There are as many classification themes in the industry as the number of strategies. After compiling both academic sources and informal, internet-based sources, pairs trading falls under the big umbrella of the long/short investing approach that is based on simultaneous exploitation of overpricing and under-pricing, by going long on perceived under-priced assets and short on perceived overpriced ones. Under the long/short investing umbrella, as opposed to say, event driven strategies, there are market neutral strategies and pairs

trading strategies. Originally suggested by Jacobs and Levy (1993), Jacobs, Levy and Starer (1998, 1999), and debated in Michaud (1993), market neutral investing is a portfolio optimization exercise that aims to achieve negligible exposure to systematic risks, whilst “harvesting” two alphas, or active returns, one from the long position on the winners and one from the short position in the losers. There are also market neutral strategies that earn both beta return and two alphas, via the use of derivatives, such as the equitized strategy and hedge strategy (Jacobs and Levy, 1993).

There are basically three main methods to implement pairs trading: The distance method, the cointegration method and the stochastic spread method.

The distance method is used in Gatev et al (1999) and Nath (2003) for empirical testing whereas the cointegration method is detailed in Vidyamurthy (2004). Both of these are known to be widely adopted by practitioners. The stochastic spread approach is recently proposed in Elliot et al (2004).

a. The Distance Method

The co-movement in a pair is measured by what is known as the distance, or the sum of squared differences between the two normalized price series under the distance method. Trading is triggered when the distance reaches a certain threshold, as determined during a formation period. In Gatev et al (1999), the pairs are selected by choosing, for each stock, a matching partner that minimizes the distance.

The trading trigger is two historical standard deviations as estimated during the formation period. Nath (2003) keeps a record of distances for each pair in the universe, in an empirical distribution format so that each time an observed distance crosses a trigger of 15 percentile; a trade is entered for that pair. Risk control is instigated by limiting a trading period at the end of which positions have to be closed out regardless of outcomes. Nath (2003) also adopts a stop-loss trigger to close the position whenever the distance widens further to hit the 5 percentile. This distance approach is purely statistical and economic model-free, hence it has the advantage of not being exposed to

model misspecification and misestimation. On the other hand, being non-parametric means that the strategy does not have the flexibility of incorporating prior knowledge by the trader in representing the relationship between the two time series. For example, in the US market, GM and Ford are expected to exhibit some form of co-movement, due to their industry similarity. A trader wishing to trade the pair would naturally like to incorporate her prior knowledge on the pair in designing the trading strategy, instead of just blindly relying on a statistic which may not be stable. Another disadvantage of this approach, inherent in any model-free approach, is its lack of forecasting ability regarding the convergence time or expected holding period.

b. The Cointegration Method

The cointegration approach outlined in Vidyamurthy (2004) is an attempt to parameterize pairs trading, by exploring the possibility of cointegration (Engle and Granger, 1987). Cointegration is the phenomenon that two time series that are both integrated of order d , can be linearly combined to produce a single time series that is integrated. Cointegrated time series can also be represented in an Error Correction Model (ECM) in which the dynamics of one time series at the current time is a correction of last period's deviation from the equilibrium (called the error correction component) and possibly some lag dynamics. In this research, we are using cointegration to select stocks and the methodology is explained in detail in the next sections. We didn't use the error correction mechanism, but gave the description of whole methodology and suggested it for the further research.

One major issue with this cointegration approach is the difficulty in associating it with theories on asset pricing. Although pairs trading has been originally premised on pure statistical results, economic theory considerations are necessary in verifying the strategy as the trader should not lose sight of fundamentals driving the values of the assets. Vidyamurthy attempts to relate the cointegration model to the Arbitrage Pricing Theory (APT) (Ross, 1976)

c. The Stochastic Spread Method

Elliott et al (2005) explicitly model the mean reversion behavior of the spread between the paired stocks in a continuous time setting, where the spread is defined as the difference between the two prices. The spread is driven by a latent state variable, assumed to follow an Ornstein - Uhlenbeck process.

$$dx_t = \kappa(\theta - x_t)dt + \sigma dB_t \quad (1)$$

Where dB_t is a standard Brownian motion in some defined probability space. The state variable is known to revert to its mean θ at the speed κ . By making the spread equal to the state variable plus a Gaussian noise, or:

$$y_t = x_t + H\omega_t \quad (2)$$

the trader asserts that the observed spread is driven mainly by a mean reverting process, plus some measurement error where $\omega_t \sim N(0, 1)$.

The above OU model offers three major advantages from the empirical perspective. First, it captures mean reversion which underlies pairs trading. The fact that x can be negative is not a problem because the spread so defined can take on negative values. Generally, the long term mean of the level difference in two stocks should not be constant, but widens as they go up and narrows as they go down. The exception is when the stocks trade at similar price points. By using the spread as log differences, this is no longer a problem. Second, being a continuous time model, it is convenient for forecasting purposes. The trader can compute the expected time that the spread converges back to its long term mean, so that questions critical to pairs trading such as the expected holding period and expected return can be answered explicitly. A third advantage is that the model is completely tractable, with its parameters easily estimated by the Kalman filter in a state space setting. The estimator is a maximum likelihood estimator and optimal in the sense of minimum mean square error (MMSE).

Despite the several advantages, this approach does have a number of shortcomings. The first one arises from the use of the OU process. Albeit fully tractable, this process fails to capture the property of heteroskedasticity in financial time series, or the dependence of volatility on the level of the variable being modeled. The diffusion coefficient in the OU process is constant which implies that the volatility of the driving force is not adaptive to the level of the force in its effort to adjust back to its long run mean.

Pairs trading strategies are generally applied in developed markets, but especially in recent years, we can see researches for emerging markets, as these markets' liquidity have been increasing. Perlin (2007) applied the pairs trading methodology in Brazilian market. Minimum squared distance rule was used, meaning that each stock was searched a corresponding pair that offers the minimum squared distance between normalized price series. To evaluate the performance of the strategy two methods were used. The first one is the computation of the excessive return of the strategy over a properly weighted portfolio and the second was the use of bootstrap methods for evaluating the performance of the trading rule against the use of random pairs for each stock. The main conclusion of the research was that the long positions of pairs trading were profitable and the long positions obtained excessive returns against the benchmark. On the other hand, the performance of short positions was not good because of the uptrend of the market in the period the research was executed.

1.2.2 Pairs Trading Background

Statistical arbitrage is an attempt to profit from pricing discrepancies in a group of assets. The detection of mispricings is based on the identification of a linear combination of assets, else a synthetic asset, whose time series is mean reverting with finite variance.

The standard approach to identify statistical mispricings is to run a regression of the values of one asset, say X_{1t} , against to the others X_{2t}, \dots, X_{nt} and test the residuals for mean-reversion.

Several tests have been developed for this purpose in the econometric literature, the most famous of which are the Dickey-Fuller and Phillip-Perron tests. Note that the residuals of the regression model represent the mispricing at each time t of X_{1t} relatively to $\{X_{2t}, \dots, X_{nt}\}$

The next step is to create a model that describes the dynamics of mispricings – how errors of different magnitude and sign are corrected over time.

To take advantage of predictability, price forecasts need to be incorporated in a dynamic trading strategy. An arbitrage trading system identifies the “turning points” of the mispricings time-series and takes proper positions on the constituent assets when mispricings become exceptionally high.

An arbitrage strategy as described above is not without risk, although profitable in the long run, its instant profit depends heavily on the ability of market prices to return to the historical or predicted norm within a short period of time.

Generally, the weaker mean-reversion, the higher the probability of observing adverse movements of the synthetic. Several authors have suggested approaches that attempt to take advantage of price discrepancies by taking proper transformations of financial time series.

Contrary to other intelligent approaches, we do not base arbitrage trading strategies on point forecasts, but on the conditional probability density for the future value of mispricing. We obtain more realistic confidence bounds on the value of synthetic that take into account short-term changes in volatility of mispricings movement. Our approach shows also a satisfactory degree of adaptivity and robustness. It adjusts the combination of the asset so as to control the mean-reversion of the synthetic time series and also detects shifts in equilibrium levels of the time series and adapts to the new stationary combinations. However, the profitability of these trades in a real market environment is still questionable given the various trading costs and market frictions.

Pairs trading relies on the principle of equilibrium pricing for near-equivalent shares. In efficient markets, capital asset pricing model-based valuation theory and the law of one price require price equality for equivalent financial assets over time. [Reilly and Brown (2000)]

The price spreads of near-equivalent assets should also conform to a long-term stable equilibrium over time. Hendry and Juselius use this principle to show that short-term deviations from these equivalent pricing conditions may create opportunities for arbitrage profits depending upon the size and duration of the price shock. [Hendry and Juselius, (2001)]

When a sufficiently large deviation of price spread from the long-run norm is identified, a trade is opened by simultaneously buying the undervalued share and selling the overvalued share. The trade is closed out when prices return to their equilibrium price spread levels by selling the long position and offsetting the short position.

Net trading profit sums the profit from long and short positions, calculated as the difference between the opening and closing prices. (net of trading costs less interest on short sale receipts) [Gillespie and Ulph, (2001)]

The risk-free characteristic of pairs trading arises from the simultaneous long-short opening market positions. The opposing positions ideally immunize trading outcomes against systematic market-wide movements in prices that may work against uncovered positions. [Jacobs and Levy, (1993)]

But arbitrage trading of the convergence trade type is rarely riskless. Market events, persistent pricing inefficiencies or structural price changes may invalidate statistical pricing models, confound future price expectations or require parameter re-estimation.

Price spreads after position opening may escalate rather than revert, or the equilibrium position may shift. The inherent nature of losses was dramatically demonstrated by the unraveling of the Long-Term Capital Management's highly leveraged long-short sovereign bond positions in the late 90s. [Lowenstein, (2000)]

Pairs trading is also exposed to risk from the inherent limitations in the statistical techniques used to identify and extract profit potential. Traditional techniques may appeal in their simplicity, but suffer severe limitations as a foundation for trading decision of choices that determine arbitrage profit potential and extraction.

The profit reduction consequences of these risks may be offset by loss limitations strategies including stop loss and time limit orders and derivatives hedging. But these strategies are costly and only limit rather than prevent loss. With regard to statistical inefficiency, a preferable situation is integrating loss protection into the statistical modeling itself.

Pairs trading is a comparative-value form of statistical arbitrage designed to exploit temporary random departures from equilibrium pricing between two shares. However, the strategy is not riskless. Market events as well as poor statistical modeling and parameter estimation may all erode potential profits. Since the conventional loss limiting trading strategies are costly, a preferable situation is to integrate loss limitation within statistical modeling itself.

Pairs trading is a strategy with a long history of modest, but persistent profits. [Peskin and Boudreau, (2000)] The strategy identifies pairs of shares whose prices are driven by the same economic forces, and then trades on any temporary deviations of those two-share prices from their long-run average relationship. [Gillespie and Ulph, (2001)]

The arbitrage or risk-free nature of the strategy arises from the opening of opposing positions for each trade – shorting the over-valued share and buying the under-valued share. [Burgess, (1999)]

The simple statistical techniques used for share pairs selection and trading decisions makes pairs trading an appealing arbitrage strategy. But simplicity comes at a cost. Correlation, covariance and regression analysis of share price associations provide an imprecise, simplistic statistical definition of a long-run equilibrium relationship between share prices. Moreover, they do not necessarily imply mean reversion to a long-run equilibrium price spread.

This research assumes that such deficiencies of the statistical techniques are best dealt with by systematic improvement within the underlying statistical modeling itself, rather than left to costly hedging and conditional order. We use Cointegration theory to provide a statistically precise foundation for the decisions involved pairs trading. In this paper, we use the principle of

cointegrated series to derive a precise, dynamic definition of long-run equilibrium price spread that inherently implies mean reversion in component series.

1.2.3 Cointegration Based Strategies

The applicability of the cointegration technique to asset allocation was pioneered by Lucas and Alexander. Its key characteristics, mean reverting tracking error, enhanced weights stability and better use of information comprised in the stock prices allow a flexible design of various funded and self-financing trading strategies, from index and enhanced index tracking to long-short market neutral strategies. [Alexander, Giblin, and Weddington, (2001)]

Alexander demonstrate that the arbitrage profit potential between two shares depends critically on the presence of a long-term equilibrium spread between share prices, the existence of short-run departures (price shocks) from the equilibrium and re-convergence to equilibrium. [Alexander, Giblin, and Weddington, (2001)]

In this situation, the statistical technique used for pairs trading must be able to provide an effective model of share price time behavior; detect equilibrium value relationships, and provide a measure of the extent and size of short-term variations from that equilibrium relationship. Gatev, Gillespie and Ulph and Alexander and Dimitriu all suggest that Cointegration theory offers a more integrative framework for statistical arbitrage strategies than current techniques. [Gatev, Goetzmann and Rouwenhorst, (1999)]

Cointegrated price series possess a stationary long-run stable equilibrium relationship with the associated property of mean-reversion. By the definition, the linear combination of cointegrated price series is stationary and will always revert back to the mean of the stationary series. This is an important fact, which will ensure that the pairs trading technique developed in this paper becomes predictable. [Gillespie and Ulph, (2001)]

Further details on Cointegration analysis can be found in Harris. [Harris, (1995)]

2. METHODOLOGY OF THE RESEARCH

2.1 Cointegration Methodology

Granger identified a link between non-stationary processes and the concept of a long-run equilibrium. If an economic time series $[y_t]$ follows a random walk, its first difference forms a stationary series. As we have already encountered, y_t is integrated of order one, and has to be differenced once in order to achieve stationary. This is usually expressed as $y_t \sim I(1)$. [Alexander and Dimitriu, (2002)]

The early work of Granger and Newbold highlighted the dangers of generating a spurious regression by regressing one $I(1)$ time series on another. Granger, however, later identified a situation when such a regression did not yield a spurious relationship. This was the case when two $I(1)$ series were cointegrated. Cointegration is shown to be an exception to a general rule. The general rule is that if two series, y_t and x_t are both $I(1)$, then any linear combination of the two series will yield a series which is also $I(1)$.

The exception to this general rule is when a linear combination of two (or more) series is integrated of a lower order. In this case the common stochastic trends have cancelled out yielding a series that is stationary. Thus, in the case of the regression of two $I(1)$ series, we do not obtain something that is spurious but something that may be relatively sensible in economic terms.

The following exposition attempts to explain this further. Assume we have two variables y_t and x_t . If we ignore cyclical and seasonal terms, we can decompose each variable into a random walk and an irregular component. Thus, we can write:

$$y_t = m_{yt} + u_{yt} \quad (3)$$

$$x_t = m_{xt} + u_{xt}$$

where m_{it} is a random walk process representing the trend in variable i at time t and u_{it} is the stationary (irregular) component of variable i at time t .

If y_t and x_t are cointegrated, there must be non-zero values of β_1 and β_2 such that the linear combination $b_1y_t + b_2x_t$ is stationary. In other words,

$$\begin{aligned} b_1y_t + b_2x_t &= b_1[m_{yt} + u_{yt}] + b_2[m_{xt} + u_{xt}] \quad (4) \\ &= [b_1m_{yt} + b_2m_{xt}] + [b_1u_{yt} + b_2u_{xt}] \end{aligned}$$

For $b_1y_t + b_2x_t$ to be stationary, the term $[b_1m_{yt} + b_2m_{xt}]$ must vanish. If either of these two trends appear, the linear combination $b_1y_t + b_2x_t$ will also contain a trend. Since the second term $[b_1u_{yt} + b_2u_{xt}]$ is by assumption stationary, the necessary and sufficient condition for y_t and x_t to be cointegrated is:

$$[b_1m_{yt} + b_2m_{xt}] = 0 \quad (5)$$

This holds for all t , if and only if:

$$m_{yt} = -b_2m_{xt} / b_1 \quad (6)$$

For non-zero values of b_1 and b_2 , the only way to ensure the equality is for the stochastic trends to be identical up to a scalar. The scalar is given by $-b_2/b_1$. Thus, up to a scalar, two $I(1)$ variables must have the same stochastic trend if they are cointegrated.

Empirical Tests for Cointegration

Once pre-testing has demonstrated that both component series are integrated of order one, variants of the Dickey-Fuller and Augmented Dickey-Fuller tests can be used to test for cointegration in the simple linear case. In addition, a modified Durbin-Watson statistic known as the Cointegrating Regression Durbin-Watson (CRDW) can also be used to undertake tests for

cointegration. All three tests are applied to the residuals obtained from a linear relationship estimated using the OLS procedure.

As noted earlier, cointegration provides an exception to a general rule. The general rule is that if two series, y_t and x_t , are both $I(1)$ then any linear combination of the two series will yield a series which is also $I(1)$. Thus if we posit the following long-run relationship:

$$y_t = a + bx_t + u_t \quad (7)$$

the linear combination is given by:

$$u_t = y_t - a - bx_t \quad (8)$$

and thus $u_t \sim I(1)$

The exception to this general rule is when a linear combination of two (or more) series are integrated of a lower order. In this case the common stochastic trends have cancelled out yielding a series that is stationary. The linear combination in this case is the errors which are assumed $u_t \sim I(0)$, and is thus a stationary series. In this case the regression of two $I(1)$ series does not lead to something that is spurious but something that may be relatively sensible in economic terms.

We apply OLS to the following equation:

$$y_t = a + bx_t + u_t \quad (9)$$

This is a regression of one $I(1)$ variable on another. This is called the cointegrating regression and represents a long-run relationship between the levels variables y_t and x_t . It has been shown that the application of OLS to $I(1)$ series yields what are called ‘super-consistent’ estimates, that is estimates that converge on their true values at a faster rate that would be the case if $I(0)$ (or

stationary variables) are used in estimation. However, the large sample distributions are not standard and the use of t-tests for inferential purposes in such cases is invalid. However, the parameter values are used to compute the residuals for the equation and these provide the empirical basis for the cointegration tests. The residuals are a linear combination of the explanatory variables and may be expressed as:

$$u_t = y_t - a - bx_t \quad (10)$$

The test for cointegration is similar in form to the Dickey-Fuller and Augmented Dickey-Fuller tests for the univariate case described before. [Dickey and Fuller (1979)] However, there are some subtle differences.

The following Dickey-Fuller regression is performed:

$$u_t - u_{t-1} = (\rho - 1)u_{t-1} + v_t \quad (11)$$

There is no constant term in this equation. This follows from the fact that a constant was included in the long-run equation above and the mean residual is therefore zero. In other words, there is no drift term. There is also no trend included in the equation.

$H_0 : \rho - 1 = 0$ [u_t is non-stationary \Rightarrow non-cointegration]

$H_a : \rho - 1 < 0$ [u_t is stationary \Rightarrow cointegration]

In other words, if there is a unit root present in the residuals, then the series in the regression cannot be cointegrated. This is because the estimated residuals are non-stationary. On the other hand, if the null hypothesis of a unit root is rejected, the estimated residuals are stationary and this is consistent with cointegration.

As with the application of this test in the last lecture, the distribution is non-standard and critical values need to be computed. The McKinnon tables can be again used to compute the appropriate critical values.

The augmented Dickey-Fuller test can also be used and is obtained through estimation of this regression:

$$u_t - u_{t-1} = (\rho - 1)u_{t-1} + \sum_{j=1}^k \delta_j u_{t-j} + v_t \quad (12)$$

The null and alternatives are given by the following:

$$H_0 : \rho - 1 = 0 \quad [u_t \text{ is non-stationary} \Rightarrow \text{non-cointegration}]$$

$$H_a : \rho - 1 < 0 \quad [u_t \text{ is stationary} \Rightarrow \text{cointegration}]$$

and the computed critical values will alter given the reduced sample size that occurs through the introduction of lagged variables.

One final test that has some intuitive appeal is based on the Durbin-Watson statistic. AR(1) process is expressed as:

$$u_t = \rho u_{t-1} + v_t \quad (13)$$

This could be re-expressed as:

$$u_t - u_{t-1} = (\rho - 1)u_{t-1} + v_t \quad (14)$$

Thus if $\rho = 1$, we have a unit root and the error series are not stationary. Recall that the Durbin-Watson is expressed as:

$$DW = 2(1 - \rho) \quad (15)$$

If $\rho = 1$, this implies that the Durbin-Watson is zero. Thus, under the null hypothesis of non-cointegration, the DW is zero. Values for the DW that are statistically different from zero imply a stationary error process and hence cointegration. This provides a new application for the DW and in this form it's referred to as the cointegrating regression Durbin-Watson (CRDW). Thus, the null and alternative hypotheses can be expressed as:

$H_0: \rho - 1 = 0$ [u_t is non-stationary \Rightarrow non-cointegration]

$H_a: \rho - 1 < 0$ [u_t is stationary \Rightarrow cointegration]

Specially tabulated critical values for the CRDW are provided and are distinct from the conventional tests used to test for the presence of serial correlation. In this case we are determining critical values at which the CRDW is statistically different from zero.

The following table provides a the critical values for the Engle-Granger Cointegration Test:

5% Significance Level for CRDW Statistic for T=50

Table 1 - Critical Values for the Engle-Granger Cointegration Test

T	1%	5%	10%
50	-4.123	-3.461	-3.130
100	4.008	-3.398	-3.087
200	-3.954	-3.368	-3.067
500	-3.921	-3.350	-3.054

The critical values are for cointegrating relations estimated using the Engle-Granger methodology. Source: MacKinnon (1991)

The Engle-Granger Two-Stage Procedure

The earliest procedure developed to explore the relationship between ECM models and cointegration was due to Engle and Granger in the mid-1980s.

1. Establish the Order of Integration of the Variables

Pre-test the variables for their order of integration using, for example, F-tests, DF and ADF tests. It is argued by some that establishing the order of integration in a pre-testing framework may be somewhat misleading. What is ultimately important is whether a combination of variables are cointegrated or not, and this could be achieved through combinations of subsets being I(1) rather than individual series being I(1). It should be noted that some econometricians feel this first step is redundant. However, if all the variables are integrated of order zero, and are thus stationary, it

is not necessary to proceed any further as standard estimation techniques can be used. The inclusion of variables with different orders of integration may lead to an ‘unbalanced’ equation. It might be useful to think of data transformations that transform I(2) variables into I(1) variables. For instance, nominal wages and prices are sometimes found to be I(2) but real wages are generally found to be I(1). One has to be cautious here, however, and, given the weak power of the tests, it may be useful to proceed with caution if the evidence is marginal one way or the other.

2. Estimate the Long-Run Relationship between the Levels Variables

Estimate the following equation using OLS

$$y_t = \alpha + \beta x_t + u_t \quad (16)$$

The following estimates are obtained:

$$y_t = \alpha + \beta x_t + u_t \quad (17)$$

$$u_t = y_t - \alpha - \beta x_t$$

Retrieve the estimated residuals, which we will define as u_t . Establish whether the residuals are stationary using, for example, DF, ADF or the CRDW tests. In the former two cases, the McKinnon tables could be used to compute the relevant critical values. If the tests indicate that the residuals are stationary, proceed to the next step. This suggests that y_t and x_t are cointegrated. Again, if the results are marginal (one way or the other), it is advisable to proceed to the next stage, as this stage offers an additional framework for testing for cointegration.

Error Correction Mechanism

Cointegration refers to the study of the possible dynamic relationship between n series where one or more have unit roots (are not stationary). The purpose of cointegration analysis is to test whether a linear combination of variables having unit roots is in fact stationary. If this position is fulfilled then it can be concluded that there exists an equilibrium relationship among a set of non-stationary variables, which would imply that their stochastic trends must be linked. Since the trends of cointegrated variables are linked, the dynamic paths of such variables must be linked to the current deviation from the equilibrium relationship. Error correction model looks at this important relationship between the change in the variable and the deviation from the equilibrium.

The formal analysis of cointegration, as introduced by Engle and Granger (1987), begins by variables in long-run equilibrium when:

$$\beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_n x_{nt} = 0 \quad (18)$$

If we define $\beta = (\beta_1, \dots, \beta_n)$ and $x_t = (x_{1t}, \dots, x_{nt})'$ then the long-run equilibrium would imply that $\beta x_t = 0$. For equilibrium to be meaningful, it must be the case that any deviation from long-run equilibrium (the equilibrium error $e_t = \beta x_t$) must be stationary. Engle and Granger (1987) provide the following definition of cointegration:

The components of the vector $x_t = (x_{1t}, \dots, x_{nt})'$ are said to be cointegrated of order d, b , denoted by $x_t \sim CI(d, b)$

1. All components of x_t are integrated of order d .
2. There exist a vector $\beta = (\beta_1, \dots, \beta_n)$ such that linear combination βx_t is integrated of order $(d - b)$, where $b > 0$. β is cointegrating vector.

An important point necessary to note about the definition is that all variables in the model must be integrated of the same order. If all series are not integrated of the same order, they cannot be cointegrated and therefore, there cannot be a long run relationship between these series.

Error Correction Model is a step forward to determine how variables are linked together. The causal flows that must exist in any cointegrated system are revealed during the second stage of cointegration modeling, the building of an error correction model (ECM). This is a dynamic model based on correlations of returns but with the constraint that short run deviations from the long-run equilibrium will eventually be corrected.

In the simplest case that there are two cointegrated log price series x and y the ECM takes the form:

$$\Delta y_t = \alpha_1 + \alpha_y (y_{t-1} - \beta x_{t-1}) + \sum_{i=1}^p a^{(i)}_{11} \Delta y_{t-i} + \sum_{i=1}^q a^{(i)}_{12} \Delta x_{t-i} + e_{yt} \quad (19)$$

$$\Delta x_t = \alpha_1 + \alpha_x (y_{t-1} - \beta x_{t-1}) + \sum_{i=1}^p a^{(i)}_{21} \Delta y_{t-i} + \sum_{i=1}^q a^{(i)}_{22} \Delta x_{t-i} + e_{xt}$$

Where the dynamic structure is captured by the difference terms, while the error correction term captures the levels (long-run) information.

If the variables are cointegrated, the residuals from the equilibrium regression can be used to estimate the error-correction model and analyze the long-run and short-run effects of the variables as well as to see the adjustment coefficients, which the coefficient of the lagged residual terms of the long-run relationship identified in cointegration process.

The intuition behind the error-correction model (ECM) is that long run errors have to be corrected in the short-run dynamics such that the process can move closer to its long run target.

The ECM is important and popular for many reasons. Firstly, it is a convenient model measuring the correction from disequilibrium of the previous period which has a very good economic implication. Secondly, since ECMs are formulated in terms of the first differences, which typically eliminate the trends from the variables involved, they can play an important role dealing with potential problems leading to spurious regressions. A third very important advantage of the ECM's is the ease with which they can fit into the general-to-specific approach to econometric modeling, which is in fact a search for the most parsimonious ECM model that best fits the given data sets. Finally, the fourth and the most important feature of the ECM is when the variables under examination are found to be cointegrated. In this case, the

disequilibrium error term is a stationary variable. The fact that the two variables are cointegrated implies that there is some adjustment process which prevents the errors in the long-run relationship becoming larger and larger. Engle and Granger have shown that any cointegrated series have an ECM representation. This is very useful when it is wished to test and incorporate both the economic theory relating to the long-run relationship between variables, and short-run disequilibrium behaviors.

2.2 Data

The database for this research is based on the Istanbul Stock Exchange's (ISE) most liquid 50 stocks between the periods 2003 and 2007 end-June.

We divided the data series into two parts: One part for modeling and the other for back testing. The first part of the data includes closes between January 2003 and December 2005. The back testing period is till end-June of 2007 (January 2006-June 2007).

The data summary is given in the Table below:

Table 2 – Data for the Research

Number of Sectors	Number of Selected Stocks	Number of Observations	Total Number of Observations
17	50	1,125	56,250

* Out of Stocks from ISE50

* The data for the research is based on the 50 most liquid stocks from Istanbul Stock Exchange between the periods of January 2003 and June 2007

All calculations in this research were updated on a daily basis from the data provider, Reuters. A general list of the stocks, their company names, tickers and floating market capitalizations are given on a Table 3 as of June 29, 2007. All of the model system was designed with Visual Basic Application in Microsoft Excel software and all steps are automated.

Table 3 – Sectors & Companies

SECTOR & TICKER	COMPANY NAME	MCAP(M\$)	SECTOR & TICKER	COMPANY NAME	MCAP(M\$)		
AUTOMOTIVE			PUBLISHING				
1	DOAS	Dogus Otomotiv	514.33	22	HURGZ	Hurriyet Gazetecilik	1162.76
2	FROTO	Ford Otosan	3308.44	RETAILING			
3	KARSN	Karsan Otomotiv	245.29	23	MIGRS	Migros	2972.46
4	TOASO	Tofas Otomobil	2433.70	TELECOM			
CEMENT			24	TCELL	Turkcell	14924.11	
5	AKCNS	Akcansa	1181.32	TRANSPORT			
GLASS			25	TAVHL	TAV	2134.87	
6	TRKCM	Trakya Cam	1065.38	26	THYAO	Turk Hava Yollari	1294.46
CONSUMER DURABLES			FINANCIALS				
7	ARCLK	Arcelik	3464.32	27	AKBNK	Akbank	16786.75
8	VESTL	Vestel	382.93	28	HALKB	Halkbank	8096.35
FOOD (other)			29	GARAN	Garanti Bankasi	11831.21	
9	ULKER	Ulker	1053.35	30	ISCTR	Is Bankasi (C)	12994.64
MALT & BREWERY			31	VAKBN	Vakifbank	6477.08	
10	AEFES	Anadolu Efes	4542.41		Yapi Ve Kredi		
METAL PROCESSING			32	YKBNK	Bankasi	7049.21	
11	EREGL	Eregli Demir Celik	5175.65	33	FORTS	Fortis Bank	2008.28
12	KRDMD	Kardemir (D)	262.57	34	FINBN	Finansbank	5700.98
PAPER & CARDBOARD			35	SKBNK	Sekerbank	900.66	
13	KOZAD	Koza Davetiye	420.24	36	TSKB	T.S.K.B.	622.41
PETROCHEMICALS			37	ASYAB	Asya Bank	1851.14	
14	AKSA	Aksa	325.46	38	AKGRT	Aksigorta	1852.98
15	PETKM	Petkim	1404.65	39	ANSGR	Anadolu Sigorta	417.37
OIL & GAS			40	GLYHO	Global Yatirim Holding	219.89	
16	AYGAZ	Aygaz	1056.89	HOLDINGS			
17	PTOFS	Petrol Ofisi	2055.34	41	ALARK	Alarko Holding	450.13
18	TRCAS	Turcas Petrol	611.68	42	DOHOL	Dogan Holding	3035.41
19	TUPRS	Tupras	6046.45	43	DYHOL	Dogan Yayin Holding	2472.06
PHARMACEUTICALS			44	GSDHO	Gsd Holding	257.55	
20	ECILC	Eczacibasi Ilac	847.43	45	IHLAS	Ihlas Holding	221.14
21	SELEC	Selcuk Ecza	1078.95	46	KCHOL	Koc Holding	7025.08
			47	NTHOL	Net Holding	210.13	
			48	SAHOL	Sabancı Holding	9037.25	
			49	SISE	Sise Cam	1752.95	
			REITS				
			50	ISGYO	Is REIT	734.71	

Although, the Float Market Capitalization amounts were not directly used in the modeling or trading section in this research for simplicity, the numbers for all pairs are provided at the last columns.

2.3 Pairs Selection

The investment strategy we aim at implementing is a market neutral long/short strategy. The strategy implies that we will try to find shares with similar betas, where we believe one stock will outperform the other one in the short term. By simultaneously taking both long and short positions, the beta of the pair approaches zero and the performance generated equals alpha.

The challenge in this strategy is identifying stocks that tend to move together and therefore make potential pairs. Our aim is to identify pairs of stocks with mean-reverting relative prices. To find out if two stocks are mean-reverting, the cointegration test is conducted to the log ratio of the pair. Engle-Granger Cointegration Test procedure was applied to determining the stationary in the log-ratio:

$$y_t = \log S_{1t} - \log S_{2t} \quad (20)$$

$$\Delta y_t = \mu + \gamma y_{t-1} + \varepsilon_t$$

In other words, we are regressing Δy_t on lagged values of y_t . The null hypothesis is that $\gamma=0$, which means that the process is not mean reverting.

If the null hypothesis can be rejected on the 90% confidence level, the price is following a weak stationary process and is thereby mean-reverting.

2.4 Trading Rules

In order to execute the strategy, we need a couple of trading rules to follow when to open and when to close a trade. Our basic rule will be to open a position when the ratio of two share prices

hits the 2 rolling standard deviations and close when the ratio returns to the mean. However, we do not want to open a position a pair with a spread that is wide and getting wider.

This can be partly avoided by the following procedure: We actually want to open a position when the price ratio deviates with more than 2 standard deviations from the 132 days rolling mean. The position is not opened when the ratio breaks the 2 standard deviations limit for the first time, but rather when it crosses it to revert the mean again.

We have an open position when the pair is on its way back again. In summary, we:

- Open position when the ratio hits 2 standard deviations bands for two consecutive times.
- Close position when the ratio hits the mean.

Furthermore, there will be some additional rules to prevent us from losing too much money on one single trade. If the ratio develops in an unfavorable way, we will use a stop-loss and close the position as we have lost 10% of the initial size of the position.

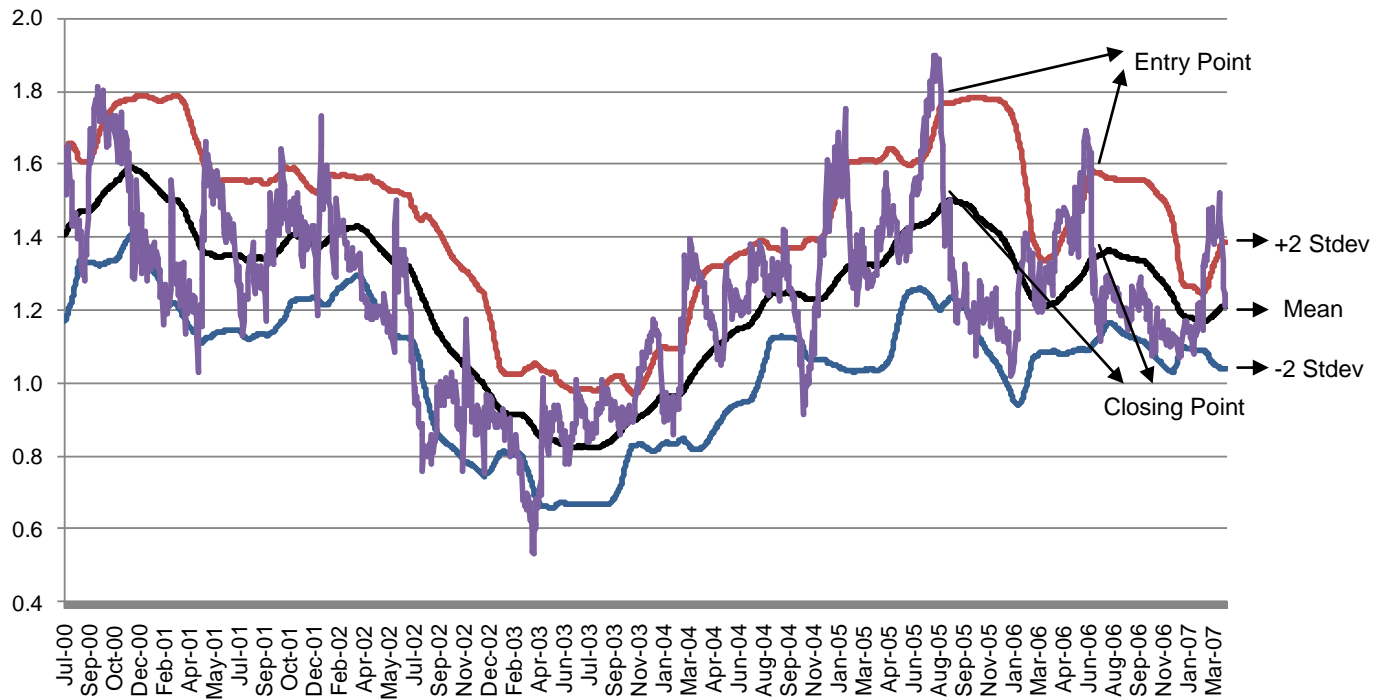
We will never keep a position for more than 132 days. There is no reason to wait for a pair to revert fully. The maximum holding period of a position is therefore set to 6 months (132 trading days). This should be enough time for the pairs to revert, but also a short enough time not to lose time value.

A pair trading using spread bets requires simply two transactions where one share is bought and another share, usually in the same sector, is sold short. The spread bet trader who utilizes a pairs trading strategy is simply speculating that one share will out-perform the other, in both an up and down stock market. A pairs trading is therefore not a bet on the overall stock market direction.

An example of pairs trading would be buying Garanti Bank (GARAN) and selling short Anadolu Sigorta (ANSGR), buying Fort Otosan (FROTO) and selling short Tofas Otomotiv (TOASO) or Koc Holding (KCHOL) and selling short Sabanci Holding (SAHOL).

The chart below shows Tekstilbank's price (TEKTL) minus Sekerbank's (SKBNK) as an example, and also takes into account the ratio of the two stock prices as they are not the same.

Chart 1 – Pairs Trading, Basic Trading Rule



The stocks' values (SKBNK and TEKST) at the beginning of 2000 were taken as an index level 100 and the graph shows how their prices changed until mid-2007. Both stocks have risen during this period. As a first impression from the spread between stocks, it may be possible to imply a long/short strategy to gain money from these stocks.

But firstly, in order to prove the applicability of long/short strategy, cointegration test was performed and the null hypothesis that asserts the process is not mean reverting was rejected at the 90% confidence level, meaning that the price is following a weak stationary process and is thereby mean-reverting.

The Engle-Granger Cointegration Test was used to determine the stationarity in the log-ratio:

$$Y_t = \log (SKBNK) - \log(TEKST) \quad (21)$$

$$\Delta y_t = \mu + \gamma y_{t-1} + \varepsilon_t$$

We are regressing Δy_t on lagged values of y_t . The null hypothesis is that $\gamma = 0$, which means that the process is not mean reverting.

The Engle-Granger Cointegration Test procedure was applied to all of ISE50 stocks within the same sector and the following results are obtained. The optimal lag, Coefficient and T-Stats are also given in the table below. Critical Values for the Engle-Granger Cointegration Test were used to identify the pairs that have cointegration relation and the optimal lags are identified through AIC criteria.

Table 4 – Cointegration Results

Sector & Pairs	Optimal Lag	Coefficient	T-Stats	Cointegration
Automotive				
DOAS FROTO	1	-0.03	-2.43	Not Cointegrated
DOAS KARSN	1	-0.03	-2.76	Not Cointegrated
DOAS TOASO	1	-0.02	-1.77	Not Cointegrated
FROTO KARSN	1	-0.01	-2.27	Not Cointegrated
FROTO TOASO	2	-0.12	-3.40	Cointegrated
KARSN TOASO	1	-0.01	-2.15	Not Cointegrated
Consumer Durables				
ARCLK VESTL	1	-0.01	-1.36	Not Cointegrated
Metal Processing				
EREGL KRDMD	1	-0.01	-2.00	Not Cointegrated
Petrochemicals				
AKSA PETKM	2	-0.12	-3.35	Cointegrated
Oil & Gas				
AYGAZ PFOTS	1	-0.01	-1.61	Not Cointegrated
AYGAZ TRCAS	1	-0.01	-2.78	Not Cointegrated
AYGAZ TUPRS	2	-0.15	-3.99	Cointegrated
PTOFS TRCAS	1	0.00	-1.63	Not Cointegrated
PTOFS TUPRS	1	-0.01	-2.43	Not Cointegrated
TRCAS TUPRS	1	-0.01	-1.70	Not Cointegrated
Financials				
AKBNK GARAN	2	0.12	3.18	Not Cointegrated
AKBNK ISCTR	1	0.00	0.29	Not Cointegrated
AKBNK YKBNK	3	0.12	3.32	Not Cointegrated
AKBNK FORTS	1	-0.01	-2.10	Not Cointegrated
AKBNK FINBN	1	0.00	-1.00	Not Cointegrated
AKBNK SKBNK	1	0.00	-0.03	Not Cointegrated
AKBNK TSKB	1	0.00	-1.37	Not Cointegrated
AKBNK AKGRT	1	0.00	-1.05	Not Cointegrated
AKBNK ANSGR	1	0.00	-0.79	Not Cointegrated
AKBNK GLYHO	1	0.00	0.04	Not Cointegrated
GARAN ISCTR	1	0.00	-1.18	Not Cointegrated
GARAN YKBNK	2	0.14	3.91	Not Cointegrated
GARAN FORTS	1	-0.02	-2.60	Not Cointegrated
GARAN SKBNK	1	-0.01	-2.03	Not Cointegrated
GARAN SKBNK	1	0.00	-0.82	Not Cointegrated

Table 4 – Cointegration Results (continued)

Sector & Pairs		Optimal Lag	Coefficient	T-Stats	Cointegration
GARAN	TSKB	1	-0.02	-2.83	Not Cointegrated
GARAN	AKGRT	1	-0.01	-2.16	Not Cointegrated
GARAN	ANSGR	2	0.11	3.09	Cointegrated
GARAN	GLYHO	1	0.00	-0.90	Not Cointegrated
ISCTR	YKBNK	1	-0.02	-2.43	Not Cointegrated
ISCTR	FINBN	1	0.00	-0.61	Not Cointegrated
ISCTR	FINBN	1	-0.01	-1.68	Not Cointegrated
ISCTR	SKBNK	1	-0.01	-2.01	Not Cointegrated
ISCTR	TSKB	1	-0.01	-1.51	Not Cointegrated
ISCTR	AKGRT	1	-0.01	-1.05	Not Cointegrated
ISCTR	ANSGR	1	-0.01	-1.59	Not Cointegrated
ISCTR	GLYHO	1	-0.01	-1.55	Not Cointegrated
YKBNK	FORTS	1	0.00	-0.31	Not Cointegrated
YKBNK	FINBN	1	0.00	-0.87	Not Cointegrated
YKBNK	SKBNK	1	-0.02	-2.44	Not Cointegrated
YKBNK	TSKB	1	0.00	-0.97	Not Cointegrated
YKBNK	AKGRT	1	0.00	-0.60	Not Cointegrated
YKBNK	ANSGR	1	0.00	-0.89	Not Cointegrated
YKBNK	GLYHO	1	-0.01	-1.94	Not Cointegrated
FORTS	SKBNK	1	-0.01	-1.41	Not Cointegrated
FORTS	SKBNK	1	0.00	-0.33	Not Cointegrated
FORTS	TSKB	2	-0.12	-3.40	Cointegrated
FORTS	AKGRT	1	-0.01	-1.95	Not Cointegrated
FORTS	ANSGR	1	-0.01	-1.48	Not Cointegrated
FORTS	GLYHO	1	0.00	-0.45	Not Cointegrated
FINBN	SKBNK	1	0.00	-0.73	Not Cointegrated
FINBN	TSKB	1	-0.01	-1.77	Not Cointegrated
FINBN	AKGRT	1	-0.01	-2.15	Not Cointegrated
FINBN	ANSGR	1	-0.01	-1.75	Not Cointegrated
FINBN	GLYHO	1	0.00	-0.87	Not Cointegrated
SKBNK	TSKB	1	0.00	-0.76	Not Cointegrated
SKBNK	AKGRT	1	0.00	-0.89	Not Cointegrated
SKBNK	ANSGR	1	0.00	-1.03	Not Cointegrated
SKBNK	GLYHO	1	-0.03	-3.54	Cointegrated
TSKB	AKGRT	1	-0.01	-2.48	Not Cointegrated
TSKB	ANSGR	1	-0.02	-2.70	Not Cointegrated
TSKB	GLYHO	1	-0.01	-1.56	Not Cointegrated
AKGRT	ANSGR	1	-0.02	-3.17	Cointegrated
AKGRT	GLYHO	1	0.00	-0.72	Not Cointegrated
ANSGR	GLYHO	1	-0.01	-1.16	Not Cointegrated
Holdings & REITS					
ALARK	DOHOL	1	0.00	-1.43	Not Cointegrated
ALARK	DYHOL	1	0.00	-1.22	Not Cointegrated
ALARK	GSDHO	2	0.10	2.75	Not Cointegrated
ALARK	IHLAS	1	0.00	-1.78	Not Cointegrated
ALARK	KCHOL	2	-0.16	-4.29	Cointegrated
ALARK	NTHOL	1	0.00	-1.33	Not Cointegrated
ALARK	SAHOL	2	-0.15	-4.00	Cointegrated
ALARK	SISE	2	-0.15	-4.12	Cointegrated
DOHOL	DYHOL	1	-0.01	-2.09	Not Cointegrated
DOHOL	GSDHO	1	-0.01	-1.86	Not Cointegrated
DOHOL	IHLAS	1	-0.01	-2.31	Not Cointegrated
DOHOL	KCHOL	1	0.00	-1.03	Not Cointegrated
DOHOL	NTHOL	1	-0.02	-2.41	Not Cointegrated
DOHOL	SAHOL	1	0.00	-1.07	Not Cointegrated

Table 4 – Cointegration Results (continued)

Sector & Pairs	Optimal Lag	Coefficient	T-Stats	Cointegration
DOHOL SISE	1	-0.01	-1.45	Not Cointegrated
DYHOL GSDHO	2	0.12	3.40	Not Cointegrated
DYHOL IHLAS	1	-0.01	-2.37	Not Cointegrated
DYHOL KCHOL	1	0.00	-1.14	Not Cointegrated
DYHOL NTHOL	1	-0.01	-1.39	Not Cointegrated
DYHOL SAHOL	1	0.00	-1.12	Not Cointegrated
DYHOL SISE	1	-0.01	-1.79	Not Cointegrated
GSDHO IHLAS	1	-0.02	-2.57	Not Cointegrated
GSDHO KCHOL	1	0.00	-0.30	Not Cointegrated
GSDHO NTHOL	1	-0.02	-2.72	Not Cointegrated
GSDHO SAHOL	2	0.11	2.96	Not Cointegrated
GSDHO SISE	1	0.00	-0.54	Not Cointegrated
IHLAS KCHOL	1	0.00	-1.65	Not Cointegrated
IHLAS NTHOL	1	-0.01	-1.66	Not Cointegrated
IHLAS SAHOL	1	0.00	-1.73	Not Cointegrated
IHLAS SISE	1	-0.01	-1.64	Not Cointegrated
KCHOL NTHOL	1	0.00	-0.73	Not Cointegrated
KCHOL SAHOL	2	-0.15	-4.04	Cointegrated
KCHOL SISE	2	-0.12	-3.31	Cointegrated
NTHOL SAHOL	1	0.00	-0.85	Not Cointegrated
NTHOL SISE	1	0.00	-0.81	Not Cointegrated
SAHOL SISE	1	-0.01	-2.14	Not Cointegrated
ISGYO ALARK	2	-0.12	-3.39	Cointegrated
ISGYO DOHOL	1	-0.01	-1.64	Not Cointegrated
ISGYO DYHOL	1	0.00	-1.22	Not Cointegrated
ISGYO GSDHO	1	0.00	-1.05	Not Cointegrated
ISGYO IHLAS	1	-0.01	-1.68	Not Cointegrated
ISGYO KCHOL	1	0.00	-0.92	Not Cointegrated
ISGYO NTHOL	1	-0.01	-1.72	Not Cointegrated
ISGYO SAHOL	1	0.00	-0.97	Not Cointegrated
ISGYO SISE	2	-0.12	-3.31	Cointegrated

14 pairs have cointegration relation at 90% significance level for the testing period, January 2003 - December 2005.

Table 5 – Pairs

Pairs		
1	Garan	Ansg
2	Akgrt	Ansg
3	Forts	Tskb
4	Skbnk	Glyho
5	Aksa	Petkm
6	Alark	Kchol
7	Alark	Sahol
8	Alark	Isgyo
9	Kchol	Sahol
10	Kchol	Sise
11	Sise	Isgyo
12	Froto	Toaso
13	Alark	Sise
14	Aygaz	Tuprs

Pairs trading is better suited to when we have a view over a week or two, perhaps even longer. Everyday pairs trading is often not the way to go because relationships between stocks in the short term can be somewhat fickle. One of the best times to get involved with pairs is when the markets get volatile and there are some clear imbalances between stocks within the same sector, or even against the index itself. Another way to pairs trading is to trade a stock against the underlying index. For example, buy GarantiBank stock (GARAN) and sell short the underlying ISE100 index, on assumption that whatever way the ISE100 moves, GarantiBank share price is likely to outperform.

2.5 Performance of the Strategy

In this research, Istanbul Stock Exchange's (ISE) most liquid 50 stocks were used between the periods 2003 and 2007 end-June. Only the liquid top 50 stocks included in the research because it is not possible to short all of stocks listed in the Istanbul Stock Exchange. In fact, short selling costs are much higher in Turkey compared with the developed markets. A short sales level fee of 5% per annum was used in this paper. (The rebate rate is between 4% and 5% per annum for large liquid Turkish stocks on the basis of a sample survey of prime brokers.)

The data series were used in two distinct parts: One part for modeling and the other for back testing. The first part of the data includes closes between January 2003 and December 2005. The back testing period is till end-June of 2007 (January 2006-June 2007).

There are 14 pairs selected that have cointegration relationship in the modeling period. (Cointegration details and results were given in the previous section.) Pure quantitative strategy gives the following outputs. 35 of the 42 pairs give positive returns, while the market neutrality is mostly satisfied.

The outputs of the backtests are given on the Table 6 below:

First column of the table lists the sector of the companies. There are 17 different sectors listed in Istanbul Stock Exchange and we found cointegration relationship within 5 sectors. Note that we looked for cointegration relation between stocks within the same sector.

The first columns of the tables show which one of the stocks is long and short. Both the open and close dates of the trades and the stocks' prices at position open and close dates are given on the tables. The long names of the companies are given in the appendix.

Total return column provides cumulative returns since mid-2004 based on the trades according to pure quantitative pairs trading methodology. The position opening and closing dates of the stocks were automatically identified by the trading system and total returns were calculated.

All the trading system was automated with Visual Basic Application (VBA) in Microsoft Excel software and data were downloaded from data provider, Reuters.

The data series used in this study is till end-June, so the trading situation (open or close) at the last day of June 2007 was given at the row of the table. For position-closed pairs, the last close date; for open ones, the last position open dates are given on the Last Close/Open Date column.

According to last trading signals, 9 of the 14 pairs' positions are open at the end-June.

Table 6 – Pairs Trade Open & Close Dates, Returns

GARAN - ANSGR

Long / Short	Open Date	GARAN	ANSGR	Close Date	GARAN	ANSGR	Return
GARAN / ANSGR	1/26/2006	6.19	3.50	5/23/2006	4.51	2.47	2.19%
ANSGR / GARAN	6/16/2006	3.81	1.91	8/1/2006	4.17	2.34	13.54%
ANSGR / GARAN	1/29/2007	5.46	2.43	3/16/2007	5.56	2.80	13.28%
ANSGR / GARAN	4/16/2007	7.25	3.10	Long ANSGR, Short GARAN (Open)			

AKGRT - ANSGR

Long / Short	Open Date	AKGRT	ANSGR	Close Date	AKGRT	ANSGR	Return
ANSGR / AKGRT	3/9/2006	6.33	3.12	4/7/2006	5.91	3.13	7.17%
ANSGR / AKGRT	5/24/2006	5.28	2.38	8/23/2006	5.23	2.49	5.50%
ANSGR / AKGRT	2/12/2007	6.01	2.50	2/26/2007	6.05	2.70	7.13%
AKGRT / ANSGR	3/20/2007	5.91	2.80	3/28/2007	6.44	2.82	8.31%
ANSGR / AKGRT	4/17/2007	7.35	3.06	4/24/2007	6.90	3.08	6.78%
ANSGR / AKGRT	5/7/2007	6.90	2.82	Long ANSGR, Short AKGRT (Open)			

FORTS - TSKB

Long / Short	Open Date	FORTS	TSKB	Close Date	FORTS	TSKB	Return
FORTS / TSKB	4/14/2006	2.62	2.63	6/21/2006	1.74	1.58	6.55%
FORTS / TSKB	8/22/2006	1.90	1.95	11/28/2006	2.01	2.01	3.11%
TSKB / FORTS	1/9/2007	1.98	1.81	1/18/2007	1.98	2.07	14.52%
TSKB / FORTS	5/16/2007	2.28	2.09	Long TSKB, Short FORTS (Open)			

SKBNK - GLYHO

Long / Short	Open Date	SKBNK	GLYHO	Close Date	SKBNK	GLYHO	Return
GLYHO / SKBNK	6/22/2006	7.45	1.12	6/29/2006	5.45	1.12	26.04%
SKBNK / GLYHO	1/3/2006	5.30	1.67	1/27/2006	7.30	1.54	45.34%
GLYHO / SKBNK	2/1/2007	5.15	1.10	Long GLYHO, Short SKBNK (Open)			

AKSA - PETKM

Long / Short	Open Date	AKSA	PETKM	Close Date	AKSA	PETKM	Return
PETKM / AKSA	3/31/2006	2.76	6.25	7/19/2006	2.15	4.88	0.06%
PETKM / AKSA	9/22/2006	3.21	5.00	2/8/2007	3.58	6.00	8.38%
AKSA / PETKM	3/5/2007	3.12	5.80	Long AKSA, Short PETKM (Open)			

ALARK - KCHOL

Long / Short	Open Date	ALARK	KCHOL	Close Date	ALARK	KCHOL	Return
KCHOL / ALARK	6/14/2006	4.34	4.76	6/28/2006	3.60	4.50	11.61%
ALARK / KCHOL	8/11/2006	3.86	5.35	9/18/2006	4.18	5.30	9.22%
ALARK / KCHOL	1/26/2007	3.62	6.10	5/31/2007	3.38	5.51	3.07%

Table 6 – Pairs Trade Open & Close Dates, Returns (Continued)

ALARK - SAHOL

Long / Short	Open Date	ALARK	SAHOL	Close Date	ALARK	SAHOL	Return
ALARK / SAHOL	2/2/2006	5.36	6.28	5/10/2006	5.08	5.73	3.68%
SAHOL / ALARK	6/14/2006	4.34	4.02	7/4/2006	3.88	4.43	20.91%
ALARK / SAHOL	1/22/2007	3.46	5.86	Long ALARK, Short SAHOL (Open)			

ALARK - ISGYO

Long / Short	Open Date	ALARK	ISGYO	Close Date	ALARK	ISGYO	Return
ALARK / ISGYO	1/25/2006	5.49	3.31	5/12/2006	4.99	3.05	-1.40%
ISGYO / ALARK	5/25/2006	4.58	2.34	6/20/2006	4.00	2.53	20.87%
ALARK / ISGYO	12/15/2006	3.76	3.00	5/31/2007	3.38	2.27	14.12%

KCHOL - SAHOL

Long / Short	Open Date	KCHOL	SAHOL	Close Date	KCHOL	SAHOL	Return
KCHOL / SISE	1/24/2006	6.23	6.22	5/1/2006	6.68	6.12	8.86%
KCHOL / SISE	10/2/2006	4.88	5.22	12/28/2006	5.60	5.61	7.21%
SISE / KCHOL	3/29/2007	6.35	5.56	5/31/2007	5.51	6.15	23.83%

KCHOL - SISE

Long / Short	Open Date	KCHOL	SISE	Close Date	KCHOL	SISE	Return
KCHOL / SISE	2/21/2006	7.23	5.97	2/24/2006	7.27	5.68	5.51%
KCHOL / SISE	3/16/2006	6.86	5.87	5/24/2006	5.55	4.50	4.10%
KCHOL / SISE	10/5/2006	5.00	4.75	12/22/2006	5.40	4.85	5.96%
SISE / KCHOL	1/15/2007	6.00	4.77	2/15/2007	6.35	5.87	17.13%
SISE / KCHOL	4/11/2007	6.70	5.34	5/31/2007	5.51	5.35	18.05%
KCHOL / SISE	6/29/2007	5.25	5.40	Long KCHOL, Short SISE			

SISE - ISGYO

Long/Short	Open Date	SISE	ISGYO	Close Date	SISE	ISGYO	Return
ISGYO / SISE	10/20/2006	5.53	2.86	12/6/2006	4.85	2.92	14.38%
ISGYO / SISE	5/29/2006	4.75	2.78	6/15/2006	3.98	2.42	3.38%
SISE / ISGYO	4/5/2007	5.38	3.24	5/31/2007	5.35	2.27	29.20%

FROTO - TOASO

Long / Short	Open Date	FROTO	TOASO	Close Date	FROTO	TOASO	Return
FROTO / TOASO	3/27/2006	9.85	4.13	4/21/2006	12.12	4.19	21.69%
FROTO / TOASO	7/26/2006	8.61	3.80	9/18/2006	10.70	4.19	13.99%
FROTO / TOASO	10/20/2006	9.94	4.76	12/20/2006	11.18	4.76	12.50%
TOASO / FROTO	2/1/2007	12.90	5.34	2/2/2007	12.80	5.73	8.08%
TOASO / FROTO	4/11/2007	12.00	5.24	4/16/2007	11.90	5.93	13.92%
FROTO / TOASO	4/17/2007	12.00	5.73	Long FROTO, Short TOASO (Open)			

Table 6 – Pairs Trade Open & Close Dates, Returns (Continued)

ALARK - SISE

Long / Short	Open Date	ALARK	SISE	Close Date	ALARK	SISE	Return
ALARK / SISE	2/22/2006	5.22	5.82	5/25/2006	4.58	4.56	9.49%
ALARK / SISE	10/12/2006	3.98	5.24	4/25/2007	3.66	5.48	-12.67%
ALARK / SISE	6/5/2007	3.26	5.40	Long ALARK, Short SISE (Open)			

AYGAZ - TUPRS

Long / Short	Open Date	AYGAZ	TUPRS	Close Date	AYGAZ	TUPRS	Return
TUPRS / AYGAZ	10/26/2005	3.62	19.64	3/24/2006	4.77	22.29	-18.30%
AYGAZ / TUPRS	5/8/2006	4.91	25.08	10/5/2006	3.43	21.19	-14.66%
TUPRS / AYGAZ	1/8/2007	3.78	22.39	2/20/2007	4.06	25.34	5.84%
AYGAZ / TUPRS	2/28/2007	3.70	25.11	4/13/2007	4.62	28.33	11.75%

The main conclusion is that a quantitative statistical arbitrage strategy, pairs trading in this case, is providing positive absolute returns in most of the trades. The main reason for the negative trades was the structural change of the companies after the modeling period. To overcome this problem, we can improve the trading system by designing more dynamic model that can realize this structural change and warn about this critical change.

Note that our strategy used in this research includes only basic methodologies and can be improved in many ways. Instead of cointegration methodology, many other strategies, such as correlation, fundamental analysis etc., can be used in order to catch the co-movement between stocks. In addition, multiple trading rules can be applied to obtain better backtesting results.

Also, asymmetric upper and lower rolling standard deviations may be used. For example, the profit objective can be designed as two standard deviations above the entry (near the mean), and the stop should be 1.5 standard deviations below the entry (3.5 total standard deviations of divergence). But, for simplicity, symmetric standard deviation rates used in this research.

The pairs trading strategy can also be applied to stocks within the same group companies, like KOC, Sabanci, Dogan Yayin Holding etc. Using long/short strategy between stocks within the same group may be a profitable strategy in Turkey. This is because holding companies tend to move together and investors, especially in emerging countries, like Turkey, closely monitor

group companies' stocks. We are currently conducting research for only group companies and the first results indicate that it is generating much better results than the classical same sector long/short strategy. For instance, the trade between, Koc Holding companies, Froto and Toaso generated 139.5% total return between mid-2004 and June 2007 and all of the trades provided positive returns.

The aim of pairs trading strategy is to satisfy market neutrality, and get consistent returns through the use of quantitative approaches, independent from the market movement. Portfolio selection techniques are essential for the performance of portfolios and risks taken for the desired returns. In this research, we focused on long term, stable, very liquid stocks that have high market capitalizations (ISE50). In order to explain strategy better, let assume an investor wants to manage his portfolio with pairs trading strategy in Turkey.

Table 7 – Stock Pairs in the Pairs Trading Portfolio

Pairs		
1	Garan	Ansg
2	Akgrt	Ansg
3	Forts	Tskb
4	Skbnk	Glyho
5	Aksa	Petkm
6	Alark	Kchol
7	Alark	Sahol
8	Alark	Isgyo
9	Kchol	Sahol
10	Kchol	Sise
11	Sise	Isgyo
12	Froto	Toaso
13	Alark	Sise
14	Aygaz	Tuprs

The stocks portfolio which includes stocks through which we identified with cointegration methodology from ISE50 stocks within the same sector is given on the Table 6.

In the strategy, models are backtested utilizing actual market-based costs to make the trades as realistic as possible. Specifically, the following costs are included:

- a. A short sales level fee of 5% per annum (The rebate rate is between 4% and 5% per annum for large liquid Turkish stocks on the basis of a sample survey of prime brokers)
- b. Equity transactions cost: 20 basis points on the face amount for each long and short position.
- c. Bid-ask spread at each position initiation or closing: 80 basis points
- d. One time FX transaction fee of 50 basis points at the initiation of the trade and also at the close of the trade.

In summary,

1. Data set for the research was chosen: ISE50 (Section 2.2)

The data for the research is based on the 50 most liquid stocks from Istanbul Stock Exchange between the periods of January 2003 and June 2007. Data were divided into two groups: One part for the modeling (January 2003 – December 2005) and the other for backtesting (January 2006 – June 2007). The cointegrated pairs were identified with the use of the modeling period data.

2. The methodology for pairs selection was identified: Cointegration. The cointegrated stock pairs were selected from ISE50. (Section 2.4)

The challenge in this strategy is identifying stocks that tend to move together and therefore make potential pairs. Our aim is to identify pairs of stocks with mean-reverting relative prices. To find out if two stocks are mean-reverting, the cointegration test is conducted to the log ratio of the pair. Engle-Granger Cointegration Test procedure was applied to determining the stationary in the log-ratio:

$$y_t = \log S_{1t} - \log S_{2t} \quad (22)$$

$$\Delta y_t = \mu + \gamma y_{t-1} + \varepsilon_t$$

In other words, we are regressing Δy_t on lagged values of y_t . The null hypothesis is that $\gamma = 0$, which means that the process is not mean reverting.

The cointegration results based on this methodology is given on the Table 4.

3. Position open & close dates were identified based on the trading rule. The long /short open and close points of the pairs were identified through the trading strategy. (Section 2.5)

Our basic rule is to open a position when the ratio of two share prices hits the 2 rolling standard deviations and close when the ratio returns to the mean. However, we do not want to open a position a pair with a spread that is wide and getting wider. We basically open position when the ratio hits 2 standard deviations bands for two consecutive times and close position when the ratio hits the mean. In the previous section, all of the trade outputs that depend on this basic rule are shown.

4. An equally weighted portfolio, consisting of all of the pairs selected from ISE50 through cointegration methodology, was formed to test the performance of the strategy. With a stock pairs and their mark-to-market values at the end of each month were calculated. If the position of any pair is closed, the value is invested to fixed income, benchmark T-Bill, until a new position open. Each pairs' total return during this period is given on the table below.

Table 8 – Cumulative Returns of the Pairs

Pairs	Total Return
GARAN - ANSGR	1.52%
AKGRT - ANSGR	13.69%
FORTS - TSKB	18.46%
SKBNK - GLYHO	54.15%
AKSA - PETKM	-6.94%
ALARK - KCHOL	48.56%
ALARK - SAHOL	35.30%
ALARK - ISGYO	60.85%
KCHOL - SAHOL	85.00%
KCHOL - SISE	89.59%
SISE - ISGYO	113.31%
FROTO - TOASO	105.98%
ALARK - SISE	19.02%
AYGAZ - TUPRS	21.87%
Portfolio	49.71%
ISE100	18.39%

As seen from the table, all of the pairs had positive return at the backtesting period, except AKSA – PETKM trade. All of the costs listed at the beginning of this section were included to the calculations (a short sales level fee of 5% per annum; equity transactions cost of 20 basis points on the face amount for each long and short position; bid-ask spread cost, 80 basis points, at each position initiation or closing; one time FX transaction fee of 50 basis points at the initiation and close of the trades).

If an equal weighted portfolio were formed with these 14 pairs, the cumulative return would be 49.71% higher than the initial investment. A portfolio that has a value of TRY100 at the end of December 2005 would reach to TRY149.71 at the end of June 2007. If invested in ISE100 Index, TRY100 would only reach TRY118.39. Both investments' values at the end of each month are given on the chart below.

Chart 2 – Pairs Trading Portfolio & ISE100 Cumulative Returns

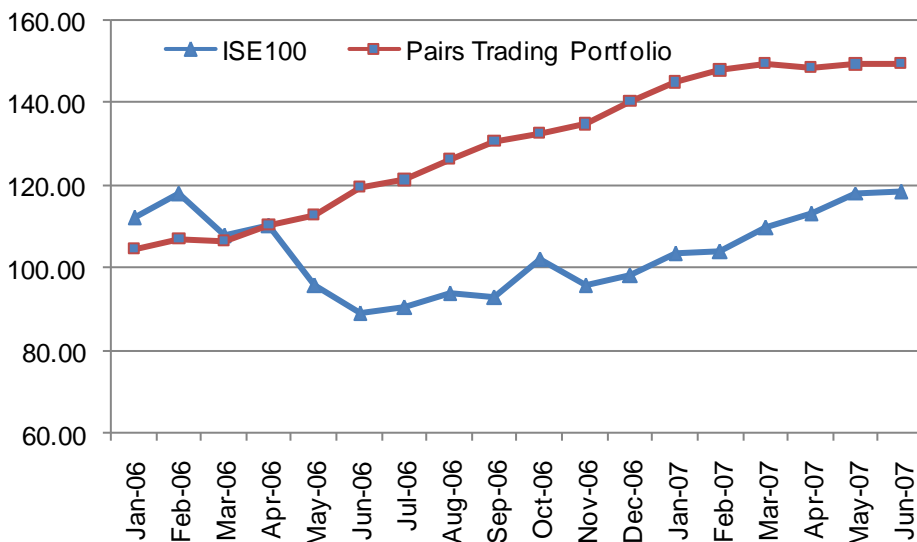


Table 9 – Performance of the Pairs Trading Portfolio

	Portfolio	ISE100
Annualized Return (Geometric)	30.87%	11.91%
Annualized Standard Deviation	6.20%	21.57%
Sharpe Ratio	2.00	-0.30
Information Ratio	2.93	-

The portfolio overperformed the Istanbul Stock Exchange, ISE100 Index. It has annualized return of 30.87%, while the ISE100 index’s return was 11.91% for the same period. Additionally, the standard deviation is significantly smaller for the portfolio. For the backtesting period, the standard deviations of the portfolio and ISE100 are 6.20% and 21.57%, respectively. Additionally, the Sharpe Ratio of the portfolio is 2.00, while its Information Ratio is 2.93. (The risk-free rate was taken as 18.45% for this period. It is the average of benchmark T-Bill rates for this period.) On the other hand, the ISE100 Index which had negative Sharpe Ratio, -0.30, performed poorly at this period. The aim in the pairs trading is to satisfy market neutrality that is whatever the market condition is, to get positive returns. As seen from the results, although the ISE100 Index had periods that had tough declines, the model portfolio performed well and got positive and more stable returns in these periods.

The pairs trading portfolio above had showed that a risk-averse investor may choose to use this strategy in Turkey. On the other hand, if the investor sees the sentiment positive and believes the market will be bullish, then he can use directional - more aggressive - strategies. Many hedge funds today use directional strategies in emerging markets, especially in the last few years. They believe, if the country has good macro fundamentals and the global environment supports this picture, and they invest in the emerging market with directional strategies. In these circumstances, directional strategies provide higher gains than market neutral statistical arbitrage strategies, like pairs trading. But, this strategy comes with a cost: they are more risky than neutral strategies.

3. CONCLUSION AND FURTHER RESEARCH

Statistical Arbitrage is an attempt to profit from pricing inefficiencies that are identified through the use of mathematical models. One technique, pairs trading in this research, is a non-directional strategy that identifies two stocks with similar characteristics whose price relationship is outside of its historical range. The strategy simply buys one instrument and sells the other in hopes that relationship moves back toward normal. The idea is the price relationship between two related instruments tends to fluctuate around its average in the short term, while remaining stable over the long term. The main objective of this research is to verify the performance and risks of pairs trading in the Turkish equity market. The main conclusion of this paper is that pairs trading had a good performance when applied to the Turkish financial market, especially for the daily frequency. In the end, we feel that pairs trading is an attractive strategy for the Turkish equity market. However, as a standalone strategy, it is faced with some risks which can be diversified away. Thus, we believe that a pairs trading strategy makes sense as part of a larger portfolio.

An equally weighted portfolio, consisting of all of the pairs selected from ISE50 through cointegration methodology, was formed to test the performance of the strategy. With a stock pairs and their mark-to-market values at the end of each month were calculated. If the position of

any pair is closed, the value is invested to fixed income, benchmark T-Bill, until a new position open.

All of the pairs had positive return at the backtesting period, except one trade (AKSA – PETKM). All relevant costs were included to the calculations (a short sales level fee of 5% per annum; equity transactions cost of 20 basis points on the face amount for each long and short position; bid-ask spread cost, 80 basis points, at each position initiation or closing; one time FX transaction fee of 50 basis points at the initiation and close of the trades). If an equal weighted portfolio were formed with these 14 pairs, the cumulative return would be 49.71% higher than the initial investment. A portfolio that has a value of TRY100 at the end of December 2005 would reach to TRY149.71 at the end of June 2007. If invested in ISE100 Index, TRY100 would only reach TRY118.39.

The portfolio overperformed the Istanbul Stock Exchange, ISE100 Index. It has annualized return of 30.87%, while the ISE100 index's return was 11.91% for the same period. Additionally, the standard deviation is significantly smaller for the portfolio. For the backtesting period, the standard deviations of the portfolio and ISE100 are 6.20% and 21.57%, respectively. Additionally, the Sharpe Ratio of the portfolio is 2.00, while its Information Ratio is 2.93. (The risk-free rate was taken as 18.45% for this period. It is the average of benchmark T-Bill rates for this period.) On the other hand, the ISE100 Index which had negative Sharpe Ratio, -0.30, performed poorly at this period. The aim in the pairs trading is to satisfy market neutrality that is whatever the market condition is, to get positive returns. As seen from the results, although the ISE100 Index had periods that had tough declines, the model portfolio performed well and got positive and more stable returns in these periods.

Expecting that the mispricing will be eliminated in the future and playing on the convergence of the asset price to its true may be risky because the convergence of the price is not immediate and its exact date is uncertain. There are two intuitively appealing strategies: The first one is to invest only if the mispricing exceeds a threshold and to keep the position unchanged until the mispricing falls below another threshold. For this strategy the relevant questions are what are the optimal thresholds and the properties of the investment portfolio corresponding to this strategy.

The second type of strategy is to continuously change positions according to the level of mispricing. In this case, we are interested in the optimal functional form of the dependence of the position on the mispricing.

In this research, there are some additional rules to prevent us from losing too much money on one single trade. If the ratio of the two stocks develops in an unfavorable way, we will use a stop-loss and close the position as we have lost 10% of the initial size of the position. Additionally, we will never keep a position for more than 132 trading days in order to prevent losses. In general, shorter period is used for closing period, but this strategy may change with the preference of investors. We think in Turkey, investors are generally long-term investors, so identify maximum holding period of 6 months (132 trading days approximately). This should be enough time for the pairs to revert to their long term mean, but also a short time not to lose time value. Moreover, the pairs are chosen among stocks that have beta spread smaller than 0.2. In order to reach sector neutrality, all pairs are selected from the same sector. Through this strategy, we want to avoid the systematic market risk. The risk there is some market risk exposure is that a minor beta spread is allowed for. In order to find sufficient number of pairs, we have to accept this beta spread, but the spread is so small that in practice the market risk we are exposed to is ignorable. Also the industry risk is eliminated, since we are only investing in pairs belonging to the same industry.

The main risk the strategy being exposed to is then the risk of stock specific events that is the risk of fundamental changes implying that the prices may never again, or at least 132 days. In order to control this risk, we use the rules of stop-loss and maximum holding period. This risk is further reduced through diversification, which is obtained by simultaneously investing in several pairs. In addition, there is a possibility for us to make our own decisions. If we, for example, are aware of fundamentals that are not taken into account in the calculations and that indicates that there will be no mean reversion for a specific pairs, we can of course avoid investing in such pairs.

Note that statistical arbitrage is subject to model weakness as well as stock-specific risk. The statistical relationship on which the model is based may be spurious, or may break down due to

changes in the distribution of returns on the underlying assets. Factors which the model may not be aware of having exposure to, could become the significant drivers of price action in the markets, and the inverse applies also.

On a stock-specific level, there is risk of M&A activity or even default for an individual corporate. Such an event would immediately end any historical relationship assumed from empirical statistical analysis.

The strategy of pairs trading is attractive since it is market neutral. This means that we are not bear any market risk and can expect to make profits irrespective of rising or falling prices. This is because we bet on the spread between the two stocks instead of betting on the absolute price levels. One of the main risks involved with pairs trading based only on quantitative strategy is that a fundamental change in the relationship between the two stocks can get masked and we can enter positions when the prices are not expected to revert to historical means. This can happen when for example, there is a fundamental change in the strategy of one of the companies as a result of which price level changes permanently.

While fundamental analysis based pairs trading normally looks at stocks within the same industry, pure quantitative strategies may not. This means that we are somewhat protected against adverse movements affecting any particular industry since he is diversified. However, the diversification may not be large enough given only two stocks in the pair.

Like any other contrarian, mean reversion strategy, our strategy was subject to noise trader risk. This is the risk that the deviations which signal a trading opportunity may not converge in the short-run and therefore, lead to higher losses, also resulting in larger margin requirements.

Diversification is another common way of decreasing the risks involved. From the point of view of pairs trading, it could mean two things. Firstly, it might mean investing only a portion of total capital in the pairs trading strategy so that there is adequate diversification through other investments. Secondly, it could mean holding positions in several pairs and limiting the amount

invested in any pair like we did. However, this may not be easy to implement since it is reasonable to assume that trading opportunities would not arise in many pairs at the same time.

Another step for this research may be examining each stock fundamental: At present, we are conducting further research to combine fundamental ratios with pure quantitative techniques in order to improve the strategy. First results are encouraging. The goal is to check the fundamentals do not contradict the trade's premise. The strategy might be buying the stock with a lower P/E, price-to-book, and market capitalization as well as a higher dividend yield, cash position, and short-interest ratio.

Another idea can be checking the recent and anticipated news for each stock to make sure a news-driven price move is not responsible for the pair diverging from its average value. Red flags include announcements of earnings, litigation, regulation, or major management changes. These events could change the pair's relationship in a fundamental, permanent way, and the pair's current value may not revert back to its mean.

Finally, we should analyze each stock's statistical and sentimental outlook for warning signs that suggest the pair will continue to diverge from its mean instead of revert towards it. Also, setting a profit objective and stop-loss upon entry might be used. For instance, it may be a good idea to exit the pairs trade if it achieves 50% of its profit target in any one day or total losses reaches 10% in one trade.

4. REFERENCES

Alexander, C. and Dimitriu, A. (2002), The Cointegration Alpha: Enhanced Index Tracking and Long-Short Equity Market Neutral Strategies, *Discussion Papers in Finance ISMA Center*, University of Reading

Alexander, C., Giblin, I. and Weddington, W. (2001), Cointegration and Asset Allocation: A New Active Hedge Fund Strategy, *Discussion Papers in Finance ISMA Center*, University of Reading –

Bachelier, B. (1900), The Theory of Speculation, PhD Thesis

Bondarenko, O. (2003), Statistical Arbitrage and Security Prices, *The Review of Financial Studies*, Vol.16, No.3, p.875

Burgess, A.N. (1999), Statistical Arbitrage Models of FTSE 100, *Computational Finance*

Conrad, J. and Kaul, G. (1998), An Anatomy of Trading Strategies, *The Review of Financial Studies*, Vol.11, No.3, p.489

Dickey, D.A. and W.A. Fuller (1979), Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, 74, p. 427–431

Dimson, E. and Mussavian, M. (1998), A Brief History of Market Efficiency, *European Financial Management*, p.91-193

Einstein, A. (1956), Investigations on the Theory of Brownian Movement, New York: Dover

- Fama, E. and French, K. (1992), The Cross-section of Expected Stock Returns, *Journal of Finance*, p.427-465 – 8
- Fama, E. (1965), Random Walks in Stock Market Prices, *Financial Analysts Journal*, September/October
- Fama, E. (1991), Efficient Capital Markets II., *Journal of Finance*, p.1575
- French, K. (1980), Stock Returns and Weekend Effect, *Journal of Financial Economics*, p.55-69
- Gatev, E., Goetzmann, W. and Rouwenhorst, K. (1999), Pairs Trading: Performance of a Relative Value Arbitrage Rule, Working Paper, *Yale University School of Management*
- Gillespie, T. and Ulph, C. (2001), Pairs Trades Methodology: A Question of Mean Reversion, *Proceedings of International Conference on Statistics*, Unpublished Paper
- Hall, G.E. (1998), Market Neutral Trading Strategies, *Advances in the Valuation and Management of Mortgage-Backed Securities*, p.27-49
- Hargreaves, C.P. (1994), Nonstationary Time Series Analysis and Cointegration, *Oxford University Press*
- Harris, R. (1995), Using Cointegration Analysis in Econometric Modeling, *Prentice Hall*, London
- Hendry, D. and Juselius, K.(2001), Explaining Cointegration Analysis: Part II, *Energy Journal*, 75-120
- Jacobs, B. and Levy, K. (1993), Long-Short Equity Investing, *Journal of Portfolio Management*, 52-63

Jeradeesh, N. and Titman, S. (2001), Profitability of Momentum Strategies: An evaluation of Alternative Explanations, *Journal of Finance*

Kargin, V. (2004), Optimal Convergence Trading

Larson, E. ad Aberg, J. (2003), A Market Neutral Statistical Arbitrage Trading Model, *Stockholm School of Economics*

Lowenstein, R. (2000), When Genuis Failed: The Rise and Fall of Long-Term Capital Management, *Random House*, New York

Malkiel, G. (2003), The Efficient Market Hypothesis and Its Critics, Working Paper

Markowitz, Harry M. (1952), Portfolio selection, *Journal of Finance*, 7 (1), 77-9

Markowitz, Harry M. (1999), The early history of portfolio theory: 1600-1960, *Financial Analysts Journal*, 55 (4), 5-16.

Murhpy, J. (1999), Technical Analysis of the Financial Markets, *New York Institute of Finance*

Park, C. and Irwin, (2004), The Profitability of Technical Analysis: A Review, *Agmas Project Research Report*, N. 04

Perlin, M.S. (2007), Evaluation of Pairs Trading Strategy at the Brazilian Financial Market, *ICMA/Reading*

Peskin, M. and Boudreau, B. (2000), Why Hedge Funds Make Sense,

Phillips, P.C.B. (1987), Time series Regression with a Unit Root, *Econometrica*, 55, p.277-302

Reilly, F. and Brown, K. (2000), *Investment Analysis and Portfolio Management*, 6th edition, *Harcourt College*, New York –

Ross, S.A. (1976), The Arbitrage Theory of Capital Asset Pricing, *Journal of Economic Theory* 13, 341-360 -4

Saul, H. (1989), Inside Morgan Stanley's Black Box, *Institutional Investor*, May, p.204

Shleifer, A. and Vishny, R.W. (1997), The Limits of Arbitrage, *The Journal of Finance*, Vol.52, No.1, p.35-55

<http://thefla.org/articles/1.pdf>

Tobias, A. (2001), Inference and Arbitrage: The Impact of Statistical Arbitrage on Stock Prices, *MIT*

Vidyamurthy, G. (2004), Pairs Trading: Quantitative Methods and Analysis, *John Wiley & Sons*, p.23-84