STOCKHOLM SCHOOL OF ECONOMICS

MASTER THESIS

Statistical Arbitrage Using Cross-Market Pairs Trading

Author:
Erik BOLTENSTÅL
22790@student.hhs.se

Supervisor:
Michael HALLING
michael.halling@hhs.se

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Finance

in the

Department of Finance
Stockholm School of Economics

May 14, 2018

Abstract

Pairs trading is a statistical arbitrage strategy that offers appealing properties for the sophisticated investor. The concept relies on the creation of a mean-reverting spread between two assets, where there is assumed to exist a long-term equilibrium relationship. This paper applies cointegration testing to model such equilibrium relationships between different pairs of 27 European equity indices. A selection algorithm based on the Engle-Granger two-step procedure picks the five most mean-reverting pairs in a formation period that are consequently traded in a trading period. The process is shifted in time in a rolling window fashion to obtain out-of-sample results for the period January 2006 to December 2017. Performance measures after transaction costs are encouraging with annualized excess returns between 3.9% and 13.3%, as well as information ratios between 0.52 and 1.29 for different parameter sets. The returns do not load on a conventional systematic risk-factor, but are almost completely beta neutral during the entire sample period.

KEYWORDS: Statistical Arbitrage, Cointegration, Kalman Filter, Pairs Trading

Acknowledgements

I extend my gratitude to friends and family for having patience with me during this stressful but rewarding learning experience. I also thank my advisor Michael Halling for providing useful comments and ideas during the course of the thesis.

Contents

1. Introduction

Hedge funds, proprietary trading firms and Commodity Trading Advisors (CTAs) have long been trying to "beat-the-market" on a risk-adjusted basis. One strategy to potentially achieve this is particularly appealing, namely pairs trading. There is now substantial evidence on the profitability of relative value strategies on a daily time-frame, and that these strategies do not load on well-known risk-factors. Pairs trading is a trading or investment strategy used to exploit financial markets that are out of equilibrium as described by Elliott et al., (2005) and Litterman, (2004). Markets may not be in equilibrium all the time, but over time they move to a rational equilibrium, and the arbitrageur will exploit these dislocations. The equilibria we refer to here are not directly observable, but can be seen as hidden stochastic relationships between certain assets. More formally, these relationships are often referred to as common stochastic trends.

The pairs trading strategy was first introduced by Nunzio Tartaglia, working at Morgan Stanley in 1987 (Vidyamurthy, 2004). Tartaglia and his group utilized a systematic investment approach and developed what is believed to be some of the very first automated trading systems for pairs trading. The process of a pairs trade is conceptually very simple, where one first identifies a pair of assets (traditionally stocks) that seem to move together. When there is a relative mispricing in the relationship, one would bet on the mean-reversion of this mispricing given the assumption of a common long-term trend. These relative mispricings are described as the *spread* between two assets in the literature. The practical implementation would require a long position in the stock that has relatively underperformed and a short position of equal nominal value in the stock that has relatively overperformed. We construct a mean-reverting synthetic asset (the spread) that is effectively market-neutral. The prospect of using a zero-cost portfolio to harvest excess returns has been particularly appealing even to academia since it does not rely on directional trading. During the last two decades, there has emerged a substantial academic literature documenting the profitability of pairs trading which has fallen under the category of what is often called statistical arbitrage. This wording is somewhat misleading since we are in fact not dealing with a real arbitrage opportunity, but a risky bet given a statistical "edge".

Vidyamurthy, (2004) is the most cited work for cointegration based pairs-trading utilizing a Engle-Granger two-step approach to formally test for cointegration between pairs. Somewhat simplified, the spread in this case can be thought of as the difference in log-price levels between two assets, where one asset is scaled by a cointegration coefficient. The spread can then be normalized and used as a trading signal. This technique has been extensively studied in a defined universe of stocks (Dunis et al., 2010; Rad et al., 2016). However, there has not been much research on applying these concepts across markets. The work that has been done is mostly related to "Siamese Twin" companies as described by Froot and Dabora, (1999), and dual listed companies (DLCs) as described by Hong and Hong and Susmel, (2012).[1] Only Burgess, (2000) presents a short case study of pairs trading between the French CAC index and the German DAX index.

This paper applies the cointegration approach of pairs trading to 27 European equity indices using a robust backtesting procedure on daily spot prices for the period 2006-01-01 to 2017-12-31. We also make use of a rolling OLS model and the Kalman filter to estimate cointegrating relationships, similarly to Dunis et al., (2010). The empirical results suggest that our strategy outperforms its benchmark across all parameterisations. We observe high information ratios between 0.52 and 1.29, even after transaction costs. We also find that the strategy is inherently less risky than a buy-and-hold strategy, with a standard deviation less than half of the benchmark portfolio. Our results also suggest that the performance of the strategy is positively related to volatility.

We now proceed to a brief review of the relevant literature in section 2, describe our methodology in section 3, and our empirical results in section 4. Section 5 concludes the paper by presenting the most important findings.

[1] Dual listings comprise of a primary listings and a second listing as an ADR. This is commonplace with Asian stocks, that often have a secondary listing in the U.S.

2. Literature Review

2.1 Cointegration of International Stock Markets

2.1.1 Common Stochastic Trends

Common stochastic trends in international stock markets can have important implications for the international investor. If systematic risk is priced similarly in different markets, the result is a common stochastic trend. The trend is then driven by some unobservable risk factor that is common across markets (Burgess, 2000). Evidence of a single common stochastic trend that drives the long-run comovement of the equity markets in the U.S., Japan, England, Germany and Canada is presented by Kasa, (1992). Interestingly, the presence of a single common stochastic trend would imply that these markets are highly correlated over long time horizons. Consequently, a investor holding only a small portfolio of international equity indices will not reap the full benefits of diversification. There also seems to be an increasing integration of European markets in particular (Mylonidis and Kollias, 2010). This is not surprising given the growing rate of economic integration in the European Union, with many countries adopting a common currency in the euro and monetary policy largely driven by the European Central Bank.

2.1.2 Volatility Spillover Effects

Volatility spillovers effects are prominent across markets both internally within Europe, but also between Europe and the United States. These effects vary in strength over time and can be triggered by both scheduled and unscheduled news events in the short run (Baele, 2005; Jiang et al., 2012). A study on *ex-ante* measures of integration of higher order moments using options data find that there is strong interdependence in variance across different markets, although the interdependencies are not as strong when it comes to skewness and kurtosis (Gagnon et al., 2016). The results also show that these interdependencies increased in persistence and in the speed of adjustment during the financial crisis of 2007-2009 for all four moments. The implication being that we will see more spillover effects across markets in the event of market turmoil.

2.1.3 The Case for Relative Value Strategies

Given that there is substantial evidence that many international stock markets are driven by the same fundamental risk factors, forming long-short pairs of cointegrated equity indices should cancel out the exposure to such common risk factors (Burgess, 2000). The relative magnitude of the assets' idiosyncratic components should be enhanced in such a portfolio. Assuming that a common stochastic trend between two equity indices follows a random walk process, and that the asset specific components exhibit mean reverting characteristics (i.e. predictable behavior), we could trade this mean reverting relative price series using a pairs trading strategy. The idiosyncratic and mean-reverting component here could be the cyclical earnings of a industry in a certain country, exchange rate fluctuations, or even an underwhelming earnings report by a heavily weighted company in an equity index. Regardless of what drives dislocations in relative value between two cointegrated markets, such dislocations tend to eventually revert back to some long term mean. The variance ratio test is proposed by Burgess, (2000) as a worthwhile exercise to truly understand how a relative price series might be inherently mean-reverting. This is important since it highlights the exact behavior which we are trying to harvest by building a pairs trading strategy. The test is presented in Appendix A.1 for the interested reader.

2.2 Pairs Trading

We have now presented the theoretical motivation for the underlying dynamic of a meanrevering spread. This section briefly describes the different approaches to pairs trading currently present in the academic literature, without claiming completeness or exhaustiveness. All methods build on similar fundamental principles and have mainly evolved from the logic of the more rudimentary but surprisingly effective distance approach.

We identify four main categories of pairs trading literature directly or indirectly relevant for this study, the *distance approach*, the *cointegration approach*, the *stochastic spread* *approach* and the *copula approach*, listed below in conjunction with relevant papers (Krauss, 2017). Novel and less well studied approaches as well as approaches not directly relevant to our study are only covered briefly under a separate section.

- Distance Approach (Do and Faff, 2010, 2012; Gatev et al., 2006)
- Cointegration Approach (Rad et al., 2016; Vidyamurthy, 2004)
- Stochastic Spread Approach (Cummins and Bucca, 2012; Elliott et al., 2005)
- Copula Approach (Krauss and Stübinger, 2017; Rad et al., 2016)
- Other Approaches: Stochastic Control, Machine Learning, PCA (Avellaneda and Lee, 2010; Huck, 2009, 2010; Jurek and Yang, 2007)

2.2.1 Distance Approach

The most influential and cited paper in the pairs trading domain of academic research is Gatev et al., (2006). The authors use a simple distance approach where they identify pairs that have prices that move together historically in a formation period. This is measured as the distance between normalized prices, forming a mean reverting series of distances (the spread). The top 20 pairs with minimum historic sum of squared distances (SSD) are picked for trading. Their trading rules automatically takes bets in the spread when it diverges by more than two standard deviations from its mean and closes the position when the spread reverts to the mean, at the end of the trading period, or when a delisting occurs. This simple trading strategy achieves an average annualized excess return of 11%. The profit persists even when accounting for conservative transaction cost estimates. The authors also show that their results are robust to alternative explanations such as reversals (Jegadeesh, 1990) and momentum (Jegadeesh and Titman, 1993) using a bootstrapping approach. An obvious advantage with this approach is that it is model-free, making it robust to model misspecifications.

The findings of Gatev et al., (2006) is confirmed by Do and Faff, (2010, 2012), although they note that profitability has decreased due to an increasing share of non-converging pairs and that the strategy is largely unprofitable after transaction costs. They also refine the selection criteria for pairs selection, only allowing for matching pairs within the same industry, potentially reducing the amount of spurious correlations. After this refinement, the trading strategy is again slightly profitable even after transaction costs.

2.2.2 Cointegration Approach

The most cited work for cointegration based pairs-trading is Vidyamurthy, (2004), mainly targeted for practitioners rather than for academics. The author proposes the Engle-Granger two-step approach to formally test for cointegration between pairs, which has not been the case in studies using the distance approach. An economic (regression) model is now introduced and defined as:

$$ln(P_t^A) - \gamma ln(P_t^B) = \mu + \epsilon_t \tag{2.1}$$

where $ln(P_t^A)$ and $ln(P_t^B)$ represents the $I(1)$-nonstationary log-price series of stocks A and B. In this case, γ is the cointegration coefficient, μ is the long-run equilibrium, and ϵ_t is the spread (the residuals of the regression model). The residuals are tested for stationarity using a unit-root test and the price series are formally tested for cointegration in a second stage. Regular t-statistics of the cointegration coefficients are sufficient to test for cointegration in the second stage. The spread is then normalized by subtracting the mean and dividing by the standard deviation, and subsequently used as a trading signal.

The most comprehensive empirical study is performed by Rad et al., (2016), testing the strategy on the entire U.S. equity market from 1962 to 2014 reporting a mean monthly excess return of 0.85% before transaction costs and 0.33% after transaction costs. The authors also test other strategies such as the distance approach and the copula approach (covered later) and find that all methods produce significant alphas after accounting for various risk factors. The cointegration method is found to perform best during turbulent market conditions.

Huck, (2015) and Huck and Afawubo, (2015) show that the cointegration approach is superior in selecting effective pairs for trading, compared to other methods such as the distance approach. This holds true both for the S&P 500 and Nikkei 225 stock universe. The papers also make the same observation as Rad et al., (2016), that pairs trading strategies seem to have produced impressive performance during the 2007-2009 financial crisis.

An improved selection algorithm for pairs is proposed by Caldeira and Moura, (2013) based on in-sample performance measures such as the Sharpe Ratio. The authors report annualized excess returns of 16.38% and a Sharpe ratio of 1.34 using their selection algorithm in conjunction with the cointegration approach on the Brazilian stock market.

2.2.3 Stochastic Spread Approach

The Stochastic Spread approach, first proposed by Elliott et al., (2005) sets out to model the spread as a stochastic process. The spread in this case is defined as the difference in log-price levels between two assets. The spread is modeled using a mean reverting Gaussian Markov chain model that is observed in Gaussian noise. Specifically, the spread is defined by the *state equation* where x_k is the state variable following a mean reverting process:

$$x_{k+1} - x_k = (a - bx_k)\tau + \sigma\sqrt{\tau}\epsilon_{k+1} \tag{2.2}$$

where a, b, and σ are constant values and ϵ_{k+1} is a standardised Gaussian noise term. The mean can be defined as $\mu = \frac{a}{b}$ where b is the mean reversion strength. Equation 2.2 can be rewritten as:

$$x_{k+1} = A + Bx_k + C\epsilon_{k+1} \tag{2.3}$$

where $A = a\tau$, $B = 1 - b\tau$ and $C = \sigma\sqrt{\tau}$. In a continuous time setting, we can describe the state process as a Ornstein-Uhlenbeck (OU) procees that satisfies the following stochastic differential equation:

$$dX_t = \rho(\mu - X_t)dt + \sigma dW_t \tag{2.4}$$

where W_t is a standard Wiener process, $\mu = \frac{a}{b}$ denoting the mean and $\rho = b$ describing the mean reversion strength. We now define the *observation process* (y_k) as the sum of the state variable and some standardised Gaussian noise:

$$y_k = x_k + D\omega_k, \qquad D > 0 \tag{2.5}$$

The parameters A, B and C in equation 2.3 can now be estimated using the recursive Kalman filter algorithm. The Kalman filter will be discussed in more detail in section 3.4.2, although with a slightly different application. The algorithm essentially takes a series of noise measurements and returns a optimal estimate for the unobservable true value. The estimate is a weighted average of the prediction by the state equation (given the observation in the previous period) and the observation in the current period. When we have estimated the model, we can enter trades at certain thresholds defined by the fixed parameter c, e.g. when $y_k \geq \mu + c(\frac{\sigma}{\sqrt{2\rho}})$ or when $y_k \leq \mu - c(\frac{\sigma}{\sqrt{2\rho}})$. The paper does not define any fixed exit thresholds but assumes that the trade will be unwound at a certain time T later. No empirical results are provided in the paper, only a remark that experiments with real data has been performed with a hedge fund.

Do et al., (2006) criticize the method (as they do with the distance approach) with regards to the fact that it assumes return parity - meaning that in the long-run the two stocks must provide the same return, to the extent that any departure from it is expected to be corrected in the future. In practice it is rare to find two such stocks, the only possible candidates being companies that adopt a dual listed company structure. This critique is largely only valid for stocks. In other asset classes such as commodity derivatives we can find pairs that have a strong fundamental relationship and where short term deviations will almost certainly correct given enough time. Such pairs could be formed for instance between crude oil futures and refined products such as gasoline and heating oil futures (Cummins and Bucca, 2012; Girma and Paulson, 1999).

Bertram, (2010) use a mean-reverting Ornstein-Uhlenbeck process to model the spread, and finds analytical solutions to optimal entry and exit levels by maximising the expected return per unit of time. Cummins and Bucca, (2012) apply this approach empirically on Energy derivatives and find Sharpe Ratios consistently above two for each year between 2003 and 2010.

The stochastic spread approach does not offer any method for pairs selection, but assumes that one has already identified suitable pairs for trading. This could be done by one of the previously mentioned methods for pairs selection. Another disadvantage with the method is the Gaussian nature of a OU-process, given that financial data tends to be leptokurtic.

2.2.4 Copula Approach

The copula method deals with some of the problems inherent in dealing with Gaussian assumptions in the cointegration and stochastic spread approach. Several new papers has been added during the last few years as this is a rather new approach to the literature (Krauss and Stübinger, 2017; Liew and Wu, 2013; Rad et al., 2016; Stander et al., 2013; Xie and Wu, 2013). First, pairs are selected according to a correlation or cointegration criteria. Log-returns are calculated and the marginal distributions F_{R_A} and F_{R_B} for the two assets A and B are obtained using either a parametric or non-parametric approach as described in Stander et al., (2013). The returns are plugged in to their own distribution functions to obtain uniform variables $U_A = F_{R_A}(R_A)$ and $U_B = F_{R_B}(R_B)$, after which an appropriate copula function $C(u_A, u_B)$ can be selected. The copula is then used to calculate the conditional marginal distribution functions using first derivatives of the copula function as seen below.

$$h(u_A \mid u_B) = P(U_A \leq u_A \mid U_B = u_B) = \frac{\partial C(u_A, u_B)}{\partial u_B} \tag{2.6}$$

$$h(u_B \mid u_A) = P(U_B \leq u_B \mid U_A = u_A) = \frac{\partial C(u_A, u_B)}{\partial u_A} \tag{2.7}$$

When the conditional probability is greater (less) than 0.5, an asset would be considered relatively overvalued (undervalued). However, trades will be triggered well in the tail regions of the conditional probabilities, e.g. below the 5% and above the 95% confidence level. Formally speaking when $P(U_A \leq u_A \mid U_B = u_B) = 0.05$ and $P(U_A \leq u_A \mid U_B = u_B) = 0.95$.

2.2.5 Other Approaches

The following methods are only covered conceptually without going in to much technical detail.

Stochastic Control Approach

Jurek and Yang, (2007) model their spread using a Ornstein-Uhlenbeck process, and make use of stochastic control theory to develop a new methodology for pairs trading. The authors apply their approach empirically to two pairs of Siamese twin shares, the Royal Dutch - Shell pair as well as the Unilever PLC - Unilever N.V. pair. They compare their results to the distance approach proposed by Gatev et al., (2006) and find that the stochastic control method delivers a significant improvement in realized Sharpe ratio as well as terminal wealth when applied to the same pairs. One important conclusion from their paper is that there is a critical level of mispricing where arbitrageurs will *decrease* their position due to negative wealth effects. This finding could be seen as a formal argument for the limits to arbitrage argument found in the behavioral finance literature.

Liu and Timmermann, (2013) build on the method proposed by Jurek and Yang, (2007) but allow for non-delta-neutral positions. The significant finding of the paper is that it can be optimal to hold both assets long or short at the same time under some circumstances. It can also be optimal to only hold one of the two assets. This is an interesting finding since it is hard to reconcile with the standard definition of a pair trade.

Machine Learning

Huck, (2009, 2010) uses Elman neural networks to generate return forecast for the upcoming week conditional on past return data. These forecasts are then compared pairwise in a ranking step and relative performance evaluated, the relative performance being the difference in the return forecasts. The most undervalued stocks are bought and the most overvalued stocks are sold in a trading step. This approach is different from other pair trading models in that there is no long run equilibrium model. The trades are instead determined by the final ranking of the assets.

PCA

Avellaneda and Lee, (2010) decompose stock returns into systematic and idiosyncratic components by regressing the return of each stock on their respective sector ETF. This concept is then extended to a multifactor model with m factors. They then use principal component analysis (PCA) to create m eigenportfolios, in line with the multifactor model, that can be traded relative to the sector ETFs. The idiosyncratic components can thus be traded separately from the systematic components.

3. Methodology

3.1 Data

3.1.1 European Equity Indices

The data set is comprised by daily closing prices of 27 European equity price indices as listed in table C.1 in Appendix C. The programming language used for data handling and analysis in this paper is R. The data is fetched using the `https://stooq.com/` API for the period 2004-01-01 to 2017-12-31 and complemented with manually downloaded data from `https://finance.yahoo.com/` since the Yahoo finance API is no longer supported. The out of sample period spans from 2006-01-01 to 2017-12-31. This period is selected considering the fact that some indices do not have data dating further back in time. Furthermore, we have to consider a possibly important structural break in the data during the decimalization of stock markets in the early 2000s.

Spot data is used for the analysis to simplify the process of forming a dataset. Implementing a trading strategy on equity indices would generally be done using index futures or ETFs rather than buying/selling the underlying basket of stocks. This holds true considering transaction costs and capital efficiency (margin requirements vs owning outright). Futures data generally are available on a per-contract basis and would thus have to be merged into a longer time-series, a non-trivial task prone to some error. ETFs do not have data available for all the markets in our sample for the chosen period. Spot data has the advantage that one does not have to account for roll-over periods and data is more readily available for longer periods and more indices. Using spot data as a proxy for the price a trader could expect to receive is not an unreasonable assumption given that we use daily data. If one wishes to use data of a higher frequency, using spot data is certainly not optimal.

The dataset is comprised of *price* indices, meaning that dividends are not incorporated in the index values. When the stocks go *ex-dividend* however, the index will be affected and fall. Since futures prices factor in this drop, there will be a slight difference between futures prices and spot prices ahead of the ex-dividend date. This problem is however not a great concern, and the price index can be considered a good proxy for the return we would expect to receive from any given trade.

One problem that is eliminated by using indices as opposed to stocks is survivorship bias. Since stocks are continuously replaced when their market cap becomes too small, holding the index will weed out companies that are not performing naturally.

3.1.2 Time-zone Differences and Holidays

Time-zone differences and different opening hours across exchanges presents an apparent problem with cross-market trading given only daily observations. There is however some homogeneity in trading hours across Europe as can be seen in table C.2 (Appendix C), listing the exchanges included in the sample. All the countries in the sample are captured by the time zones UTC, UTC+1, UTC+2 and UTC+3. The opening hours mostly conform to the London trading session (08:00-16:30 UTC), although there are some local variations, especially in closing hours. The biggest discrepancy is with regards to the Baltic countries, all closing at 14:00 UTC, 2.5 hours before the end of the London session. Moreover, the DAX is formally open until 19:00 UTC, although trading after the London session is generally very quiet. The morning opening hours are very homogeneous, where all exchanges except the Bratislava Stock exchange and Iceland Stock Exchange are open at 08:00 UTC when the London session starts. The varying closing hours are a problem that is hard to eliminate completely given the frequency of the data, as closing prices will inevitably represent different times of the day in some indices. Given this, one could not be certain to receive the daily closing prices for both assets when entering the long and short leg simultaneously (opening a pair). A practical solution could be to enter all trades at 14:00 UTC when the Baltic exchanges close, and treat that time as our closing price across all exchanges. Running a backtest with this specification is however not possible using only daily closing prices (intra-day data is needed in this case). The reader should keep this shortcoming in mind when interpreting the results. Lacking access to intra-day data, this problem could be greatly reduced by implementing the strategy on a weekly time-frame. A few hours in between data points will almost certainly not affect returns measured on a weekly time-frame substantially. While the primary backtest is conducted using daily observations, the strategy is also tested on a weekly time-frame for this reason as a robustness test.

National holidays have not been removed from the sample but simply appear as a missing value in the raw dataset. Common holidays such as Christmas and Easter will be missing across most of the countries, meaning that no new trades can be initiated since there is no movement in price. In both cases the closing price from the last day will simply be carried forward to the missing value, i.e. the return will appear as zero. This is common practice in the literature and could only cause problems if there is a national holiday in one country but not in the other, in which case a cross-market spread trade could potentially be hindered. This is however a small source of bias, and not likely to affect the results significantly.

3.1.3 Foreign Exchange Risk

Foreign exchange risk is another important issue to consider in cross-market trading. Indices are denoted in local currency, meaning that a cross-market spread will be affected by currency fluctuations if the long and short leg are denoted in different currencies. One solution to this would be to recalculate all index prices to the same currency using time series of daily exchange rates. We opted against this method since a cointegrating relationship is primarily a statistical relationship generally performed without any type of currency conversion. This relationship will be altered by introducing new variables that are allowed to interact with the data. The pairs trading strategy presented in this paper is consequently exposed to foreign exchange risk, and the possibility that results are driven by currency fluctuations has to be addressed. A spread could converge only due to an appreciation or depreciation of a certain currency against another. To test this hypothesis, we run the strategy only on a subset of the 15 indices denoted in EUR in table C.1 (Appendix C), instead of the 27 in our original sample. If results persist even after this restriction, it is reasonable to assume that the results are not primarily driven by currency fluctuations.

3.2 Overview of the Trading Strategy

The trading strategy can be described by three main processes that run in tandem but with different purposes: (1) *pairs selection*, (2) *continuous estimation* of the cointegrating relationship using time-varying models and (3) *continuous tracking* of the normalized spread to trigger trades.

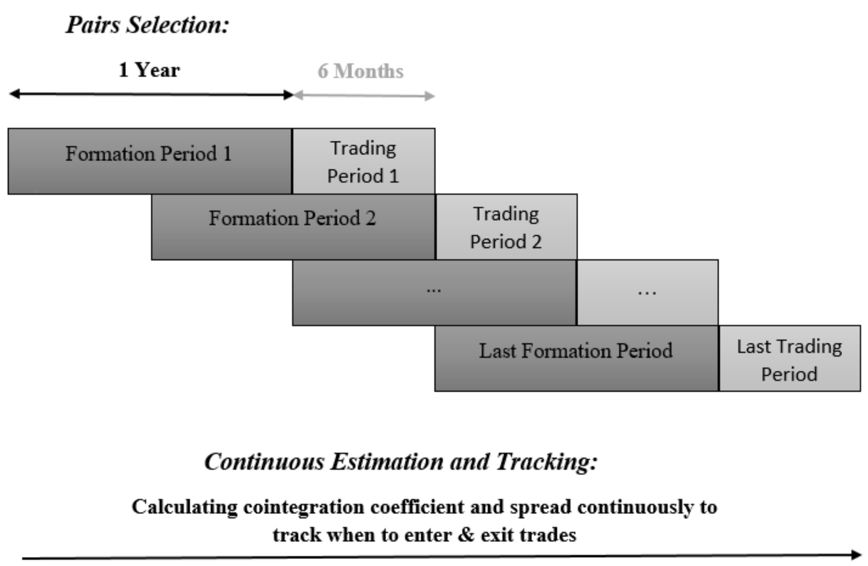
The first process, pairs selection, is described in detail in section 3.3. It is performed semiannually using data from the past year (the *formation period*) and then moved forward through the sample by six months at the time using a rolling window approach. This process decides which pairs that will be traded during the following six months (the *trading period*). The choice of a formation period of one year and a trading period of six months follows the standard practice in the literature of a formation period twice as long as the trading period and is the proposed window sizes by Gatev et al., (2006). A one year formation period is also shown to have better performance than a two year formation period sometimes used (Huck and Afawubo, 2015).

The second process is the continuous estimation of the cointegration coefficient and calculation of the spread, as described in detail in section 3.4. We utilize a time-varying model that updates the cointegration coefficient each day similarly to Dunis et al., (2010). This is done using two different methods that will be compared in the results section, a rolling OLS regression and a Kalman filter approach.

The third process is tightly linked to the second. We now normalize the spread using a Z-score as described in section 3.5. The Z-score defines deviations from the average spread in terms of standard deviations.

The entire backtesting procedure can be seen in figure 3.1 and is designed to eliminate lookahead bias, so that the rules could be applied in real-time as a trading strategy without modification. Step (2) and (3) as described above are illustrated together in the bottom of the figure as a continuous process occurring every day.

FIGURE 3.1: Backtesting procedure with periodic pairs-selection, continuous estimation of the cointegrating relationship and tracking of the normalized spread



The procedure of periodically shifting a formation and trading period forward in time as seen in 3.1 is commonly known as walk-forward analysis among practitioners. This is a cross-validation technique that has gained traction among practitioners as-well as academics since one can obtain robust out-of-sample results even when the amount of data is limited. The end result is a long vector of returns comprising of all the trading periods combined together. Performed correctly, these returns will be truly out-of-sample. In our case, we end up with a return vector built by combining 24 trading periods between January 2006 and December 2017 to obtain a out-of-sample time series of 12 years. This is the return series that is presented in the results section.

3.3 Pairs Selection

The process described below is the pairs selection procedure and is performed semiannually using a rolling window approach.

3.3.1 Engle-Granger Two-Step Method

The Engle and Granger, (1987) two-step methodology is preferred in financial applications as noted by Alexander, (2001) and Dunis et al., (2010) due to its simplicity and and lower variance compared to the Johansen, (1988) method. Since we are only concerned with pairs of assets in this paper, the former method is chosen. The Johansen method has some interesting applications when there is more than one cointegrating vector and the synthetic asset is comprised of more than two assets (Dunis and Ho, 2005). This application is however outside the scope of this paper.

Step 1

The first step of the Engle-Granger method is to test that the log-price series of all the assets in the sample are indeed integrated of order one and thus contain a unit root. That the price series are nonstationary might seem obvious, but nevertheless this has to be formally tested. This is traditionally done using a unit root test such as the Augmented Dickey-Fuller (ADF) test or Phillips-Perron (PP) test. The ADF test and PP test is known to perform poorly when there are structural breaks in the data and the sample is small, both conditions that will be present given our chosen method. Structural breaks can certainly occur when a cointegrating relationship "breaks" and our sample size for each individual test will only be one year as defined by the formation period. Considering these shortcomings, we decide to use another unit root test based upon the weighted symmetric estimator of Pantula, Gonzales-Farias and Fuller (PGF) defined in equation 3.1 (Pantula et al., 1994):

$$\hat{\rho}_{WS} = \frac{\sum_{t=2}^{n} Y_{t-1} Y_t}{\sum_{t=2}^{n-1} Y_t^2 + n^{-1} \sum_{t=1}^{n} Y_t^2} \tag{3.1}$$

where Y is the time-series to be tested for a unit root and n is the length of the vector. The distribution of $\hat{\rho}_{WS}$ are then determined through simulation and table lookup is used to determine the p-value associated with the test statistic. This test has been shown to perform better than both the ADF test and PP test in simulation studies (Pantula et al., 1994). Specifically, it rejects the null hypothesis of a unit root more accurately in small samples than the ADF test and PP test.

After confirming that the price series are indeed nonstationary, all assets are combined into pairs. A universe of 27 assets means that there are 351 potential pairs available to trade. The log-price levels of one asset A will be regressed on the log-price levels of the other asset B in the pair according to the specification in equation 3.2 using a OLS regression. This is sometimes called the *cointegrating regression* or the *static regression* (Do et al., 2006).

$$ln(P_t^A) = \mu + \gamma ln(P_t^B) + \epsilon_t \tag{3.2}$$

The relationship is often rewritten as in equation 3.3 to highlight the fact that the cointegrating relationship contains a long-run mean μ and a residual term ϵ_t sometimes called the error correction term. Equation 3.3 implies that if ϵ_t is stationary, then the combination $ln(P_t^A) - \gamma ln(P_t^B)$ must also be stationary.

$$ln(P_t^A) - \gamma ln(P_t^B) = \mu + \epsilon_t \tag{3.3}$$

The constant μ can also be thought of as a "premium" in one asset versus another as described by Vidyamurthy, (2004) and Do et al., (2006). Vidyamurthy, (2004) discusses possible reasons for such a premium between two stocks, such as a a relative liquidity premium, takeover potential or pure charisma. These explanations are not directly transferable to stock indices, although there might exist other reasons for a premium in equity indices such as perceived political risk or an unstable currency.

The residuals ϵ_t are tested for stationarity using the PGF test described above, although the condition now is that the regression residuals are $I(0)$-stationary. If the null hypothesis of a unit root is rejected, we can proceed to step two with the pair. It is important to note that it is not possible to perform any hypothesis testing about the actual cointegrating relationship at this stage or make any inferences.

Step 2

An Error Correction Model (ECM) can now be specified as in equation 3.4 containing the differenced price series as well as the error correction term, being the lagged residuals from the previous step. The ECM should now account for both a short-term relationship in the differenced series and a long-term relationship in the error correction term.

$$ln(\frac{P_t^A}{P_{t-1}^A}) = c + \alpha ln(\frac{P_t^B}{P_{t-1}^B}) + \beta\epsilon_{t-1} + v_t \tag{3.4}$$

The t-statistics of the coefficient β can now be used to test for cointegration. The parameter β will now represent the mean-reverting behavior of the cointegrating relationship, where a negative value will predict convergence towards the long-run mean in in the next period when the spread is high or low in the current period. The higher the magnitude of β, i.e. the absolute value $\mid \beta \mid$ for a negative coefficient, the more mean reverting behavior is expected.

AR(1) Model for the Error Correction Term

While an ECM is the traditional way to represent the cointegrating relationship, we elect to use another specification. We model the spread as a $AR(1)$-process as in equation 3.5 similarly to the stochastic spread approach described above, but in discrete time. As Moura et al., (2016) notes, ARMA dynamics can always be considered as valid attempts for modeling the spread, given their mean reverting behavior. The $AR(1)$-process could be thought of as the discrete time counterpart to the Ornstein-Uhlenbeck process used in the stochastic spread approach (Moura et al., 2016; Neumaier and Schneider, 1998).

$$\epsilon_t = \rho \epsilon_{t-1} + \eta_t \tag{3.5}$$

The process we use can thus be described by estimating parameters μ, γ and ρ using equations 3.2 and 3.5. If $\mid \rho \mid < 1$ and significant, then $ln(P_t^A)$ and $ln(P_t^B)$ are cointegrated. The parameter ρ will now give us a measure of the mean reversion if the spread can adequately be modeled by a $AR(1)$-process, which is tested using a Ljung-Box test of the residual series (Ljung and Box, 1978). The Ljung-Box test is a portmanteau test for the "overall" randomness of a time series based on not one specific lag but many lags. The null hypothesis is clearly defined as: $H_0 : r_1 = r_2 = ... = r_k = 0$, whilst the alternative hypothesis is defined less clearly as: H_1 : not all $r_j = 0$. A rejection of the null hypothesis thus only tells us that there is some non-random component in the time series. The specification in 3.5 has the advantage of fewer parameters to estimate and the interpretation of ρ is somewhat more intuitive than β in the ECM.

3.3.2 Selection Algorithm

The selection algorithm for which pairs to trade is based on the Engle-Granger procedure, and performed for all 351 combinations of possible pairs. The process is described in a stepwise fashion below:

1. Check that the two individual price series are integrated of order one using a unit-root test such as the PGF test.

2. Run the cointegrating regression and test the residuals for stationarity (again using a unit root test), proceed to next step if stationary.

3. Form a $AR(1)$-model of the spread, if the residuals can be described by a $AR(1)$-process and $\mid \rho \mid < 1$ in equation 3.5, pair is cointegrated.

4. Sort cointegrated pairs on ρ (lower values are better, since more mean reverting), and trade the five best pairs.

3.4 Time-Varying Equilibrium Models

We now proceed to describe the second process occurring continuously and in tandem with the pairs selection, namely estimating the cointegrating relationship using a time-varying model.

To see the benefit of an adaptive model for the long run equilibrium, we present two examples, starting by describing a static model where the cointegrating relationship is stable through time. Figure 3.2 uses a static regression between the log-prices of BEL 20 and FTSE 250 estimated over two years in 2005-2006, and then applied for the period 2005-2010. We can clearly see how there seems to be a common trend throughout the sample, although the level defined by γ and μ in the relationship has changed after 2007, "breaking" the static relationship. As noted by Burgess, (2000), there is every reason to believe that the markets opinion as to what constitutes a "fair price" of one asset relative to another will change over time. This effect can also be observed as a downward drift in the spread (figure 3.2), i.e. the relationship has not disappeared but persists in a modified form. The mean-reversion dynamics now exists around a drift of the spread, creating a type of "channel" effect (Burgess, 2000). The persistence of a mean reverting behavior between the two assets can also be confirmed by the variance ratio function (VRF) described in A.1 (Appendix A) and plotted in figure B.1 (Appendix B).

Although the traditional static relationship can be sufficient in some cases such as an in-sample cointegration test as described in the pairs selection, it is not optimal when we need the best possible estimate in real-time trading out-of-sample. The cointegration analysis is done *ex post*, meaning that it would contain lookahead bias if we used the information to backtest a trading strategy within the same time-window we have estimated the model on. We thus need to form a estimate for γ and μ that only is estimated on past data, and preferably able to adapt to a changing long-run equilibrium. This can be achieved using a time-varying model such as a rolling OLS regression or a dynamic linear model estimated using the recursive Kalman filter algorithm. Figure 3.3 depicts a rolling OLS model applied on the same pair as before, and we observe that the relationship is more well behaved. Figure 3.4 plots the rolling estimates of γ and μ adjusting the long-run equilibrium level of the relationship as time goes by.

FIGURE 3.2: Static cointegrating regression estimated over 2005-2006, BEL 20 and FTSE 250

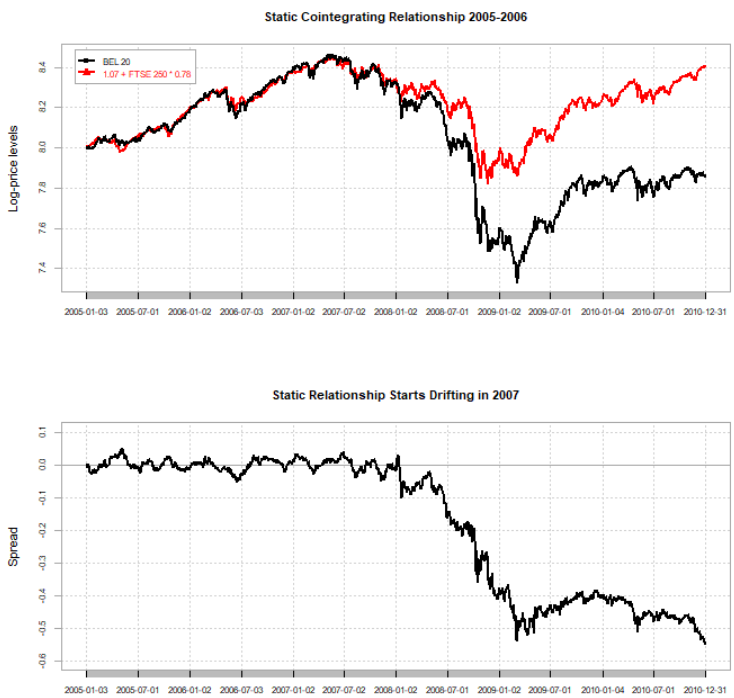


FIGURE 3.3: Dynamic cointegrating regression with a lookback period of 250 Days, BEL 20 and FTSE 250

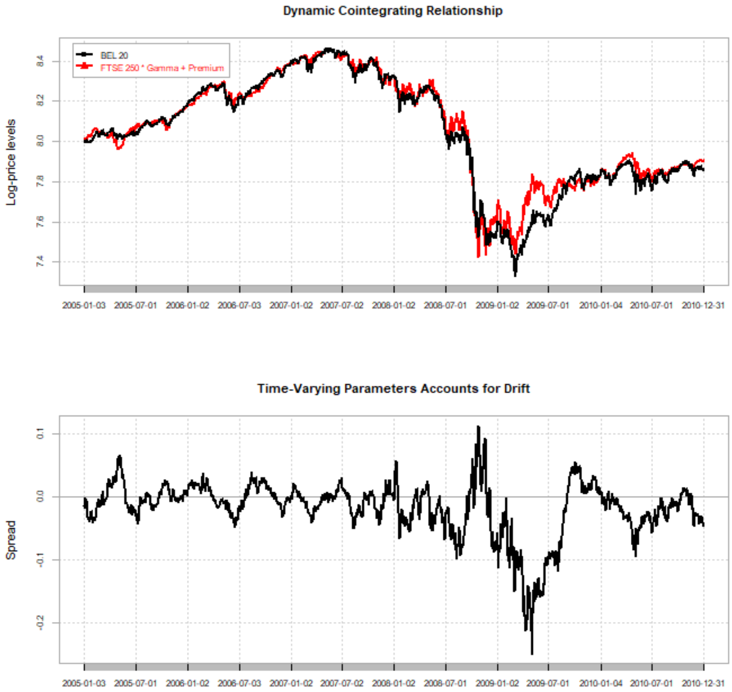
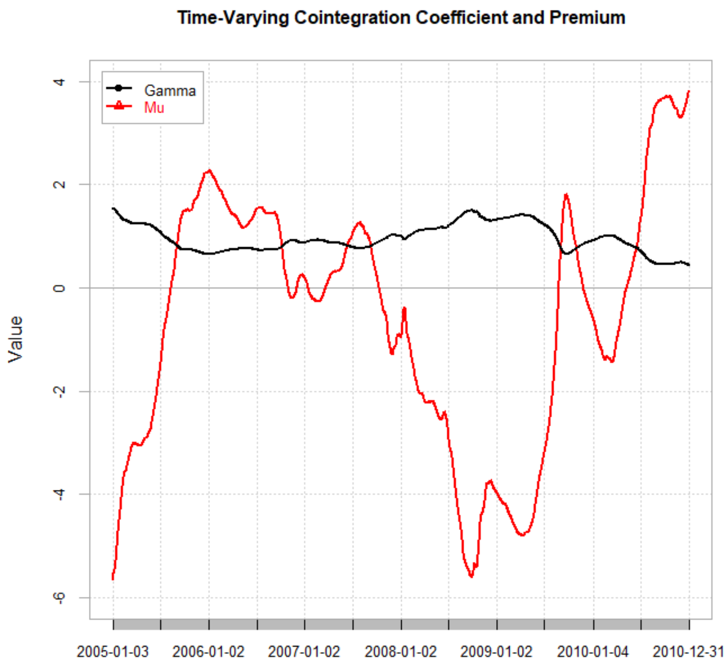


FIGURE 3.4: Time-varying cointegration coefficient and premium estimated using rolling OLS regression with 250-days lookback period, BEL 20 and FTSE 250



3.4.1 Rolling OLS

Given the benefits of adaptive models, we utilize a rolling OLS regression that is estimated in each period (daily) using data from the last year (250 trading days). The lookback period for the regression is consistent with the length of our static regression in the pairs selection. For each time-step, we estimate the regression in equation 3.2 and can obtain the de-trended residuals ϵ_t by rewriting 3.2 as 3.6. We add time subscripts to the cointegration coefficient γ and the premium μ since they are continuously estimated as previously illustrated in figure 3.4.

$$\epsilon_t = ln(P_t^A) - \gamma_t ln(P_t^B) - \mu_t \tag{3.6}$$

Estimating the model completely without an intercept μ is sometimes done when implementing short term strategies (Dunis et al., 2010; Krauss, 2017). This simplifies the interpretation of the cointegration coefficient if one uses non-log-prices in the model. The coefficient is then also the *hedge-ratio* - i.e. the ratio one must sell (buy) asset B to offset the long (short) position in asset A. In this case, the spread will include the premium component μ which is not significant in short-term trading but can be important in longer-term strategies.

3.4.2 Kalman Filter

We define a state space regression model, also known as a dynamic linear model in the equation system 3.7:

$$\begin{aligned}
ln(P_t^A) &= \mu_t + \gamma_t ln(P_t^B) + v_t, \qquad & v_t &\sim N(0, \sigma_v^2) \\
\mu_t &= \mu_{t-1} + w_{\mu,t}, & w_{\mu,t} &\sim N(0, \sigma_\mu^2) \\
\gamma_t &= \gamma_{t-1} + w_{\gamma,t}, & w_{\gamma,t} &\sim N(0, \sigma_\gamma^2)
\end{aligned} \tag{3.7}$$

where $ln(P_t^A)$ is the dependent variable and log-price level of asset A and $ln(P_t^B)$ is the independent variable and log-price level of asset B. As previously, μ_t is the premium and γ_t is the cointegration coefficient. The error terms v_t, $w_{\mu,t}$ and $w_{\gamma,t}$ are assumed to be uncorrelated with variances σ_v^2, σ_μ^2 and σ_γ^2 respectively. The first equation in system 3.7 is known as the *measurement equation*, and the two latter equations are the *state equations*. The states evolve according to a random walk model which is called the *prediction step*, and the estimates are then revised in the *updating step* where the predictions are compared to the noisy measurements (realizations of $ln(P_t^A)$). The updated values of μ_t and γ_t (and some noise) will then be our prediction in the next time-step when the process is iterated. This recursive process is known as the Kalman filter and is used to obtain optimal estimates for the parameters of the model. For details on the Kalman filter algorithm, see Appendix A.2.

The specification in 3.7 is in many ways very similar to our previous definition in 3.2 and 3.6, although the premium μ_t and cointegration coefficient γ_t are now estimated using the Kalman filter. The spread is defined exactly as before but v_t now denotes the spread:

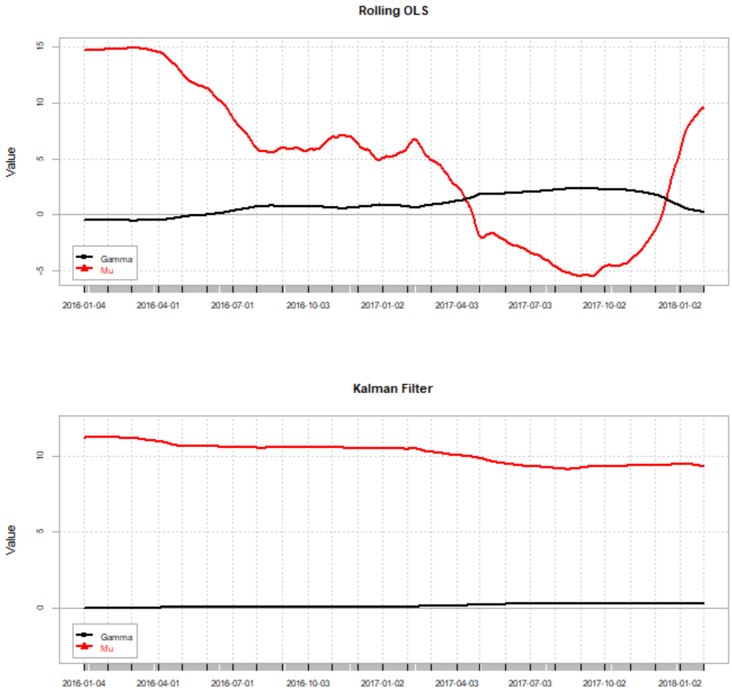
$$v_t = ln(P_t^A) - \gamma_t ln(P_t^B) - \mu_t \tag{3.8}$$

The Kalman filter is generally preferred over the rolling OLS approach since it provides more accurate and less volatile estimates of the actual parameters as proved by simulation studies where the data generating process is known in advance (Das and Ghoshal, 2010). The advantage of the Kalman filter is its ability to adjust quickly to new information and converge towards the true value without introducing higher uncertainty into the estimates. Figure 3.5 illustrates this phenomenon visually, where the Kalman filtered estimates are obviously less sporadic around the "true" parameter values. The true values are of course unobservable in the case of real market data, but it seems as though the OLS estimate fits to noise when the premium turns negative for a few months and then suddenly becomes positive again.

Another benefit of the Kalman filter is the fact that no lookback window has to be chosen. A lookback window is unfortunately reintroduced when we normalize the spread later, so the benefit in our application is rather limited. What we have to chose is a *noise-ratio* defined as $S = \frac{\sigma^2_{\mu,\gamma}}{\sigma^2_{v_t}}$, being the ratio between the state noise and the measurement noise.[1] The noise-ratio regulates how adaptive the Kalman filter will be to noisy observations. We chose estimates for $\sigma^2_{\mu,\gamma}$ and $\sigma^2_{v_t}$ yielding a noise ratio of $S = 10^{-5}$ that has been showed to be in optimal territory for pairs trading in simulation studies (Burgess, 2000). This is where the correlation between estimated and true deviations from the common trend are close to their maximum. We do not optimize the noise-ratio for each pair separately but use this estimate for all pairs.

[1]The state noise σ^2_μ and σ^2_γ have the same value and are thus denoted as $\sigma^2_{\mu,\gamma}$.

FIGURE 3.5: Time-varying estimates of cointegrating relationship, rolling OLS vs Kalman filter



Although the Kalman filter is regarded a superior method to estimate cointegrating relationships, one has to keep in mind that there is still a trade-off between adaption and over-sensitivity, and that this is largely governed by the choice of noise-ratio.

3.5 Trading

The third process of the strategy, running in parallel to the pairs selection and time-adaptive modeling, is described below. When the spread is calculated according to equation 3.6 or 3.8 we normalize it to be able to generate consistent trading signals.

3.5.1 Tracking the Normalized Spread

A Z-score is computed using the calculated spread ϵ_t as described in Caldeira and Moura, (2013). The Z-score is dimensionless and measures the distance to the mean of the spread in units of standard deviations, as seen in equation 3.9.

$$z_t = \frac{\epsilon_t - \mu_\epsilon}{\sigma_\epsilon} \tag{3.9}$$

Our trading signal is generated when the normalized spread (Z-score) reaches certain thresholds where we enter and exit trades, as seen in figure 3.6. The signal can be seen as a vector where 1 means that we are long the spread, -1 that we are short the spread and 0 that we do not have any position, also seen in figure 3.6. This signal is then lagged (one or two days) and multiplied with the return vectors for the corresponding assets in the pair. Lagging the signal vector is done to make sure that the trading strategy is viable in practice, i.e. that we do not use contemporaneous information in our trading rules. The Z-score is tracked for all individual pairs separately.

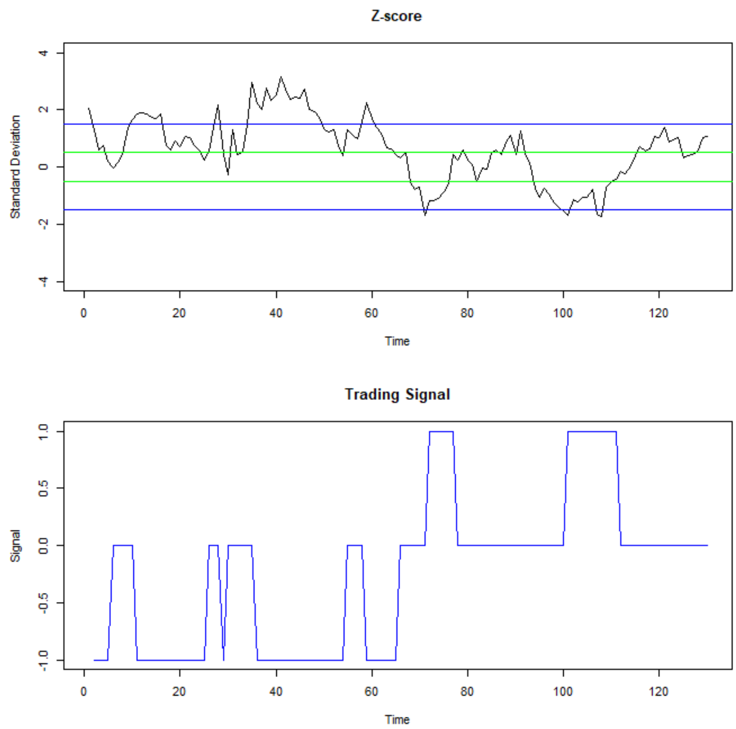
The mean μ_ϵ and the standard deviation σ_ϵ of the spread determines if the spread is in disequilibrium. This poses the question of what n-period window to measure μ_ϵ and σ_ϵ over to normalize the spread, and if the two measures should be static or time-varying. One common approach is to simply use the static mean ($\mu_\epsilon = \mu$ from equation 3.2) and corresponding standard deviation of the formation period where cointegration testing is performed as in Dunis et al., (2010). Another approach is to use a rolling estimate with the same lookback period as the length of the formation period ($\mu_\epsilon = \mu_t$ from equation 3.6 or 3.8). I.e. we assume that the estimated long-term mean from the cointegrating regression is also the average spread we will measure deviations from. While these two methods make most theoretical sense, they lead to either completely static or very slowly adapting estimates of μ_ϵ and σ_ϵ, which could lead to unnecessary losses when spreads do not converge as expected (Girma and Paulson, 1999). The phenomenon of non-converging spreads can be clearly observed in our previous example in section 3.4 (figure 3.2 and 3.3). Given a static or slow dynamic estimate of the long-run mean the mean reversion might take several months or not converge at all before the trading period has ended, which would be clearly suboptimal. However, we can relax this restriction on our Z-score and let μ_ϵ be different from the long-term mean in the cointegrating regression.

To see why this relaxation might be useful, recall the "channel" effect described earlier, which implies that a spread identified by cointegration testing can have inherent mean reverting properties on different time-frames. Specifically, a spread can be mean reverting towards a long-run mean for several months, but also fluctuate around a "short-run" mean. Put differently, the spread can drift away from its long-run mean, but still exhibit a clearly mean reverting behavior. In line with these observations, Girma and Paulson, (1999) propose a significantly shorter and adaptive lookback period for μ_ϵ and σ_ϵ and propose a

5-day and 10-day rolling window. This allows the strategy to exploit short term deviations that consequently revert to the average spread more efficiently. The number of potential trades are increased and unprofitable trades are closed out earlier. This implementation also largely removes the need for a stop-loss used in many papers that utilize a static or slowly adapting long-term estimate for the mean. Intuitively, a shorter lookback period can be described as a moving average following the spread more closely, representing a "short-run" mean.

We decide to try rolling windows of 20, 60 and 250 days for μ_ϵ and σ_ϵ in equation 3.9, where the two shorter windows are in line with the discussion in Girma and Paulson, (1999) and the longer window in line with other literature in the field using the size of the formation period to build the normalized Z-score.

FIGURE 3.6: Trading signal generated from Z-score entry thresholds $\pm$ 1.5 SD (blue outer lines) and exit thresholds $\pm$ 0.5 SD (green inner lines)



An obvious problem with the use of a normalized Z-score is the assumption that the spread is approximately normally distributed, and that we must define what constitutes an extreme spread by defining our thresholds as seen in figure 3.6. This is a strong assumption given that financial time-series often exhibit fat-tails. These limitations have lead to the development of new approaches such as the copula method as described above.

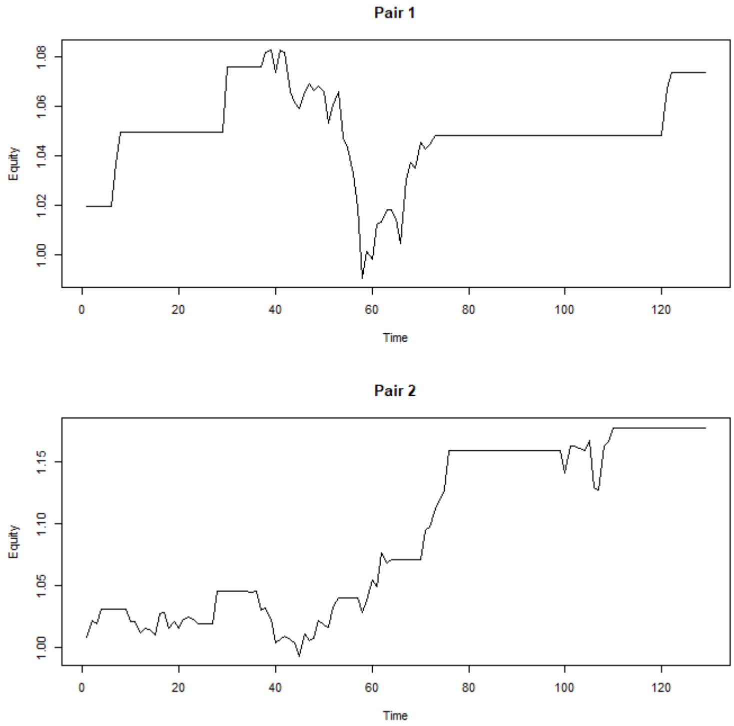
3.5.2 Trading Thresholds

The choice of threshold levels for the Z-Score is an important aspect of the strategy and tells our algorithm what constitutes an extreme observation of the spread. There is a trade-off between trade quality and number of trades when increasing or decreasing the threshold. A high threshold will generally yield trades of good quality and low drawdowns, but reduce the amount of trades significantly and possibly overall returns. A low threshold will generate trades off worse quality and higher drawdowns, but increase the amount of trades drastically and possibly overall returns. We elect to use three different thresholds commonly used in the literature $\pm 1.0SD$, $\pm 1.5SD$ and $\pm 2.0SD$. We chose a permanent exit threshold of $\pm 0.5SD$ which is also a commonly adopted threshold in the literature (Caldeira and Moura, 2013). Our trading rules for the $\pm 1.5SD$ threshold (as seen in figure 3.6) are thus defined as:

$$\begin{aligned}
&\text{Open Long spread if} && z_t < -1.5SD \\
&\text{Open Short spread if} && z_t > 1.5SD \\
&\text{Close long spread if} && z_t > -0.5SD \\
&\text{Close short spread if} && z_t < 0.5SD
\end{aligned}$$

with equivalent trading rules for the $\pm 2SD$ and $\pm 1.0SD$ specifications, where only opening thresholds are changed. The closing thresholds remain the same at $\pm 0.5SD$. It is regarded as suboptimal to wait until the deviation from the equilibrium is completely closed, even if it might do so eventually. Considering that the mean-reverting dynamics are at their max when the deviation is high, they are also likely to be low $\pm 0.5SD$ around the mean. Positions are also closed if they have not converged before the trading period has ended. Figure 3.7 depicts the equity curve of two different pairs during a trading period of 6 months. Our portfolio consists of five such pairs at each time, and one can easily see when trades are opened and closed in the figure.

FIGURE 3.7: Example of equity curves of two pairs in a trading period



3.6 Returns and Performance Measures

3.6.1 Calculation of Returns

The asset returns are calculated as first differences of log-prices, and the return of a pair is calculated as described by Dunis et al., (2010), and seen in equation 3.10:

$$r_{it} = ln(\frac{P_t^L}{P_{t-1}^L}) - ln(\frac{P_t^S}{P_{t-1}^S}) \tag{3.10}$$

where r_{it} is the return for pair i in period t, P_t^L is the price of the index we are long in period t and P_t^S is the price of the index we are short in period t. The returns and volatility can be annualized using equations A.14 and A.15 (Appendix A).

After performing the backtest using log-returns (r_{it}), we transform these to simple returns (R_{it}) using the formula in Appendix A.16. Aggregating returns across a portfolio of K pairs requires simple returns, and we use equation 3.11 to obtain daily simple returns (R_{st}) for our strategy, where w_{it} is the weight ($\frac{1}{K}$) of each pair.

$$R_{st} = \sum_{i=1}^{K} w_{it} R_{it} \tag{3.11}$$

Since our equity curves are presented using geometric chaining of simple returns, we elect to present all performance measures in the results section using simple returns for consistency. Although log-returns are time additive, they are less suitable to use when producing equity curves over long time-horizons since log-returns are only negligibly close to simple returns over short horizons such as days.

3.6.2 Transaction Costs

Transaction costs are included of 5 basis points one-way for one asset, meaning that transaction costs for one pair is 10 bp one-way and 20 bp for a round-trip trade. This is in line with the middle estimate of Bowen et al., (2010). The transaction costs are subtracted from the returns in the period when the trade is initiated and when it is closed. Naturally the number of pairs traded affects the total transaction costs of the portfolio, why we chose to settle with five pairs (10 assets in total) as proposed by Dunis et al., (2010). The total cost for opening and closing five pairs once is thus 100 bp (1% from total returns) and has to be considered a moderate estimate for transaction costs.

3.6.3 Information Ratio

In quantitative finance, the information ratio has started to gain traction compared to the more traditional Sharpe Ratio. The definition gets rid of the risk-free rate and defines the excess returns R^e relative to a relevant benchmark in the numerator. The estimated standard deviation of the excess returns is the denominator. The *ex-post* information ratio is then defined in equation 3.12 for any given period as described by Goodwin, (1998).

$$IR = \frac{\overline{R^e}}{\hat{\sigma}^e} \tag{3.12}$$

Since our returns are from a long-short strategy and thus are excess returns by definition, we can simply use the mean of the return series in the numerator and the sample standard deviation in the denominator. This redefinition removes some of the ambiguity of including the risk-free rate in the calculations. For example, questions arise on how we treat negative risk-free rates in Sharpe Ratio calculations, and if the risk-free rate is in-fact the relevant benchmark.

3.6.4 Single Factor Model

We define a single factor model as follows:

$$R_s = \alpha_s + \beta_s(R_b - R_f) \tag{3.13}$$

where R_s is the return of the strategy, R_b the return of the benchmark, R_f is a constant risk-free rate at 2% (daily $R_f = \frac{2\%}{252}$), α_s is the part of the strategy returns not explained by the specification and β_s the factor loading on the benchmark excess returns.[2] There is no need to include the risk-free rate in front of the alpha since we are dealing with a long-short strategy and thus dealing with excess returns by definition as mentioned before. The benchmark is however a long-only return stream why we have to subtract the risk-free rate to obtain excess returns.

3.6.5 Drawdown Measures

The maximum drawdown (MDD) is defined as the maximum loss from a peak to a bottom defined over a specific period of time. More formally, we can define the maximum drawdown in percentage terms over the time period T as:

$$MDD_T = \max_{\tau \in (0,T)} \left[\max_{t \in (0,\tau)} \frac{X_t - X_\tau}{X_t} \right] \tag{3.14}$$

where X is defined as the equity of the strategy, X_t is a local peak in equity and X_τ is a local bottom. The MDD can be seen as a proxy for how much "risk" investors tolerate before withdrawing their money, and thus is a very important industry measure. For example, a hedge fund with a 50% drawdown is unlikely to survive, due to client withdrawals and possible margin calls on leveraged positions.

We can also define the drawdown (D) at time τ in equation 3.15 and the average drawdown (ADD) over time T in equation 3.16 where D_j is the j :th drawdown of a total of d number of drawdowns.

[2] The risk-free rate of 2% is a constant and chosen as a conservative proxy. A better estimate derived from real data will not provide any added value to our analysis, as we are interested in the pattern of alpha and beta.

$$D(\tau) = \max\left[0, \max_{t\in(0,\tau)} \frac{X_t - X_\tau}{X_t}\right] \tag{3.15}$$

$$ADD(T) = \frac{|\sum_{j=1}^{d} D_j|}{d} \tag{3.16}$$

Finally, we define the Risk-Return-Ratio as defined by Johnsson, (2010). This ratio puts the raw return of a strategy in relation to its maximum drawdown as seen in equation 3.17.

$$RRR = \frac{\overline{R_T}}{MDD_T} \tag{3.17}$$

This definition has gained popularity due to its simplicity and intuitive interpretation.

3.6.6 Value at Risk

In addition to the risk-measures above, we report 1-month (1% and 5%) value at risk (VaR) and expected shortfall (ES) measures for the return series as proposed by Clegg and Krauss, (2018) and Krauss and Stübinger, (2017). We report the historical VaR and ES suggested by Mina and Xiao, (2001), instead of relying on assumptions about the underlying distribution. If our returns are significantly non-normal (skewed and/or kurtotic), using a Gaussian assumption and traditional Monte Carlo methods will underestimate the VaR and ES estimates. A better option is then to let the historical data dictate the shape of the distribution given the frequency that they have been observed in the past. This is a simple yet transparent method that seems to be a preferred method among practitioners as well as academics.

4. Results

4.1 Backtesting Results

We perform a rigorous performance evaluation of our strategy. Specifically, we examine the annualized performance metrics such as the information ratio, risk-return-ratio and maximum drawdown measures as discussed in section 3.6. We also compute exposure of the strategy to systematic risk using the single factor specification in 3.13. We then proceed to look at monthly return distributions for the strategy and corresponding historical value-at-risk (VaR) and expected shortfall (ES) measures.

4.1.1 Equity Graphs and Performance

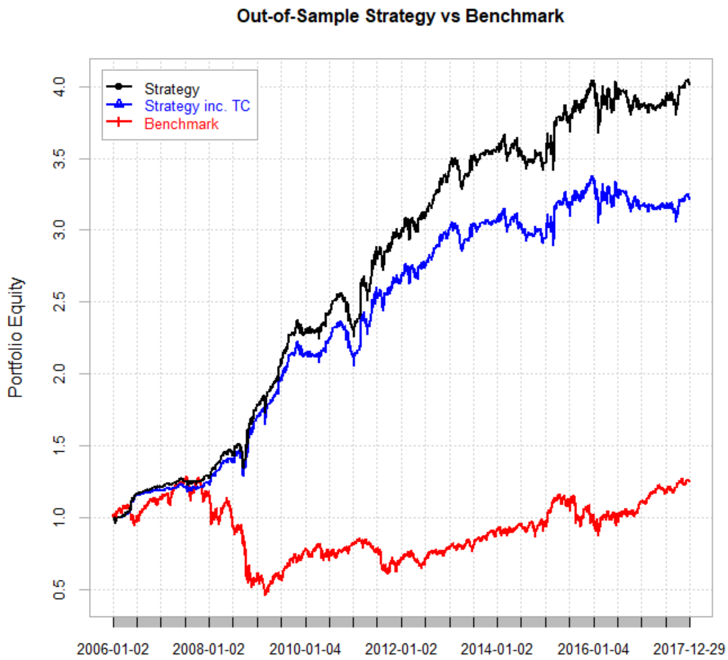
The out-of-sample performance of the long term backtest for all parameterisations can be seen in tables 4.1 and 4.2. The equity graph for one parameter set is plotted in figure 4.1 and compared to a benchmark portfolio. The benchmark portfolio is a equal weighted buy-and-hold portfolio of the 27 indices in the sample.[1] The strategy is tested with a 1-day and 2-day lag on the signal as mentioned previously. The 1-day lag means that the trades are entered at the close the day of the signal (i.e. without delay), and the 2-day lag translates to a "wait one period" delay sometimes used in the literature.[2]

For the rolling OLS specification seen in table 4.1, we can observe annualized returns of 4.6% and 16.2% before transaction costs and 3.9% to 11.8% after transaction costs. Information ratios are between 0.62 and 1.54 before transaction costs and between 0.52 and 1.13 after transaction costs. For the Kalman filter specification seen in table 4.2, we can observe annualized returns of 5.4% and 19.3% before transaction costs and 4.7% to 13.3% after

[1] This portfolio aims to proxy the broad market portfolio in Europe and is confirmed to be an almost a perfect mirror of the STOXX Europe 600, a index comprising of 600 European equities in 17 European markets.

[2] The "wait one period" approach can be regarded as overly conservative from a practical standpoint when using daily data.

FIGURE 4.1: Out-of-sample strategy performance. Threshold ± 1.5 *SD* and 60 period lookback. A 1-day lag is imposed on the signal.



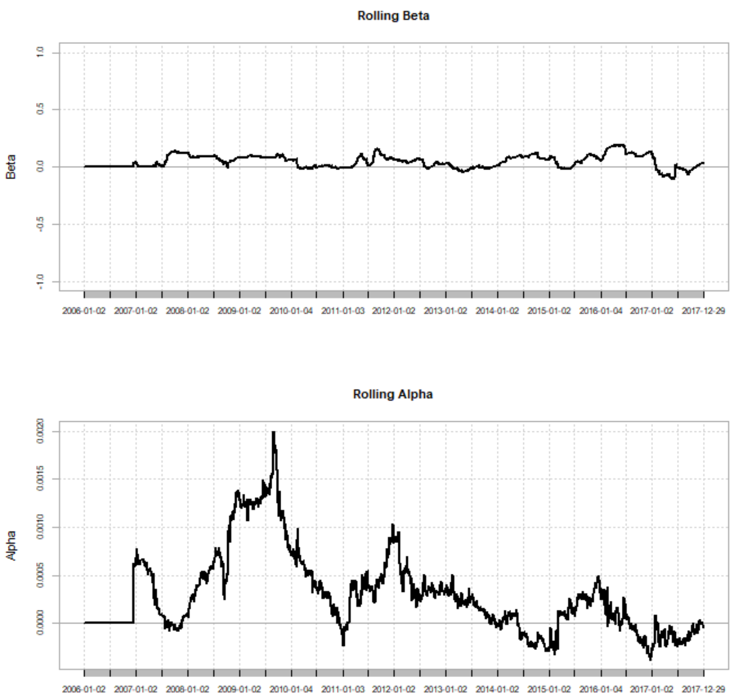
transaction costs. Information ratios are between 0.77 and 1.87 before transaction costs and between 0.66 and 1.29 after transaction costs. Consequently, the strategy clearly outperforms its benchmark portfolio that has a information ratio of 0.09 over the same period and an annualized return of 1.8%. We can also conclude that the Kalman filter appears to be superior in estimating the cointegrating relationship, yielding higher returns and information ratios. This is not surprising given its ability to adapt to new data without over-fitting to the noise as discussed previously.

The volatility of the strategy is between 7.1% and 10.5% for all parameterisations and methods in table 4.1 and 4.2, less than half of the volatility of the benchmark at 21.2%. Similarly, maximum drawdown for the strategy is between 8.9% and 19.1% for all cases and less than a third of the drawdown of the benchmark at 64.5%. A similar picture appears when comparing the average drawdown and the risk-return-ratio to that of the benchmark. Clearly, the strategy seems to have favorable characteristics for the risk-averse investor.

Analyzing the different parameter sets in table 4.1, we find that performance seems to decline when using the longer lookback period (250 days) using the OLS method, supporting the use of shorter lookback periods as discussion in section 3.5.1. The average drawdown length is significantly shorter when we use the shorter lookback periods, indicating that loosing positions are closed out earlier. Our results thus point in favor of using shorter lookback windows to normalize the spread in line with Girma and Paulson, (1999). The pattern for threshold levels are less clear cut for the OLS method. While returns increase before transaction costs using a lower threshold, the number of trades increase and the quality of trades decrease. After transaction costs, the returns and information ratios are similar across all threshold levels. The state space specification in table 4.2 does not seem to be equally sensitive to the choice of lookback period, especially when the threshold is at a lower level ($\pm 1SD$). The choice of threshold level appears important before transaction costs, but less important after transaction costs. What is clear, is that the number of trades (and thus returns) increase with a shorter lookback window and lower entry thresholds. However, more trades also mean more transaction costs, offsetting part of the higher returns. It is evident that the less selective parameter sets are much more sensitive to transaction costs comparing the performance before and after transaction costs. In the case where we have most trades (1538 in total) as seen in table 4.2, we note that returns fall from 19.3% to 13.3% after inclusion of transaction costs. The implication is that optimal parameters depend heavily on the assumptions regarding transaction costs.

Figure 4.2 depicts the rolling beta and alpha of one parameterisation according to the single factor specification in equation 3.13. The rolling beta tells us that we indeed seem to have created a strategy that is very close to market-neutral. The pattern is very similar for all parameterisations with regards to both beta and alpha. Looking at the rolling alpha we note that the performance of the strategy seems to have declined quite dramatically over recent years. Curiously, the most alpha is produced during the financial crisis in 2007-2009 indicating that the strategy seems to perform better in turbulent market environments, and that the performance is positively related to volatility. A clear pattern emerges when we see that the alpha increases during the Greek debt crises in 2012 and during the increased volatility in early 2015. This positive relationship between performance and volatility may seem counterintuitive, but makes sense given that we are harvesting alpha from relative mispricings, likely to be at their peak level in a volatile market environment. The pattern in figure 4.2 is recurring when we plot the annualized return, volatility and information ratio over time in Appendix B.3. We note that the standard deviation of the strategy increases to close to 15% during turbulent times, although returns increase even more, yielding high

FIGURE 4.2: 250-Day rolling beta and alpha of strategy.
A 1-day lag is imposed on the signal



information ratios during these times.

We lastly turn to the case where we impose a 2-day lag on the signal, as seen in tables C.3 and C.4 in Appendix C. We note that this restriction significantly reduces the performance of the strategy, much in line with what previous literature has found regarding time-delays in execution (Bowen et al., 2010; Gatev et al., 2006). Despite the significantly lower annualized returns, the strategy still beats the benchmark comfortably after transaction costs with information ratios between 0.25 and 0.62 for the OLS specification and between 0.37 and 0.75 for the Kalman filter specification. The higher information ratios compared to the benchmark seem to be primarily driven by lower standard deviations for the strategy. We can thus conclude that while profits are significantly reduced by a "wait-one-period" restriction, the inherently lower risk in a long-short strategy is still present.

TABLE 4.1: Out of Sample Results 2006-2017, Rolling OLS & 1-day signal lag

	Rolling OLS	*Excl. TC*			*Incl. TC*			
		Lookback Z-Score			**Lookback Z-Score**			
Entry	**Measure**	**20**	**60**	**250**	**20**	**60**	**250**	**Benchmark**
2.0	Annualized Returns	0.107	0.085	0.046	0.082	0.072	0.039	0.018
	Annualized SD	0.080	0.080	0.074	0.079	0.080	0.074	0.211
	Maximum Drawdown	0.089	0.107	0.109	0.122	0.108	0.113	0.645
	Average Drawdown	0.011	0.010	0.011	0.010	0.010	0.012	0.036
	Annualized IR	1.345	1.070	0.622	1.038	0.902	0.524	0.085
	Risk-Return-Ratio	1.204	0.797	0.423	0.673	0.664	0.343	0.028
	Avg. DD Length (days)	19	14	28	21	15	32	88
	Nr Trades	648	369	193	648	369	193	NA
1.5	Annualized Return	0.123	0.119	0.070	0.085	0.099	0.058	0.018
	Annualized SD	0.095	0.094	0.090	0.094	0.094	0.090	0.211
	Maximum Drawdown	0.114	0.121	0.107	0.152	0.132	0.108	0.645
	Average Drawdown	0.012	0.011	0.014	0.014	0.011	0.013	0.036
	Annualized IR	1.296	1.264	0.771	0.898	1.057	0.643	0.085
	Risk-Return-Ratio	1.071	0.981	0.648	0.555	0.749	0.535	0.028
	Avg. DD Length (days)	17	14	24	22	16	26	88
	Nr Trades	1010	538	304	1010	538	304	NA
1.0	Annualized Returns	0.162	0.150	0.075	0.106	0.118	0.058	0.018
	Annualized SD	0.105	0.105	0.102	0.104	0.105	0.101	0.211
	Maximum Drawdown	0.156	0.130	0.117	0.191	0.143	0.128	0.645
	Average Drawdown	0.013	0.013	0.015	0.015	0.014	0.014	0.036
	Annualized IR	1.544	1.428	0.740	1.013	1.130	0.577	0.085
	Risk-Return-Ratio	1.041	1.153	0.645	0.554	0.825	0.456	0.028
	Avg. DD Length (days)	15	15	19	20	16	22	88
	Nr Trades	1470	830	483	1470	830	483	NA

Results of the 24 (six-month) combined out-of-sample trading periods between January 01, 2006 and December 31, 2017. A rolling OLS regression is used to estimate the cointegrating relationship and a 1-day lag is imposed on the signal.

TABLE 4.2: Out of Sample Results 2006-2017, Kalman Filter & 1-day signal lag

	Kalman Filter	*Excl. TC*			*Incl. TC*			
		Lookback Z-Score			**Lookback Z-Score**			
Entry	**Measure**	**20**	**60**	**250**	**20**	**60**	**250**	**Benchmark**
2.0	Annualized Returns	0.106	0.092	0.054	0.081	0.078	0.047	0.018
	Annualized SD	0.078	0.079	0.071	0.078	0.079	0.071	0.211
	Maximum Drawdown	0.116	0.124	0.108	0.136	0.134	0.119	0.645
	Average Drawdown	0.010	0.009	0.012	0.009	0.009	0.011	0.036
	Annualized IR	1.352	1.163	0.766	1.040	0.993	0.663	0.085
	Risk-Return-Ratio	0.915	0.742	0.505	0.598	0.584	0.394	0.028
	Avg. DD Length (days)	17	14	27	20	16	30	88
	Nr Trades	653	385	227	653	385	227	NA
1.5	Annualized Return	0.149	0.122	0.084	0.109	0.102	0.070	0.018
	Annualized SD	0.095	0.093	0.083	0.094	0.093	0.083	0.211
	Maximum Drawdown	0.109	0.118	0.108	0.135	0.132	0.122	0.645
	Average Drawdown	0.012	0.010	0.010	0.011	0.010	0.010	0.036
	Annualized IR	1.570	1.307	1.008	1.161	1.097	0.845	0.085
	Risk-Return-Ratio	1.361	1.036	0.782	0.811	0.773	0.575	0.028
	Avg. DD Length (days)	16	15	18	18	17	21	88
	Nr Trades	1073	582	387	1073	582	387	NA
1.0	Annualized Returns	0.193	0.164	0.141	0.133	0.129	0.116	0.018
	Annualized SD	0.103	0.104	0.100	0.103	0.104	0.099	0.211
	Maximum Drawdown	0.109	0.141	0.093	0.135	0.164	0.104	0.645
	Average Drawdown	0.012	0.011	0.011	0.014	0.013	0.011	0.036
	Annualized IR	1.872	1.567	1.421	1.298	1.238	1.173	0.085
	Risk-Return-Ratio	1.774	1.159	1.514	0.986	0.784	1.122	0.028
	Avg. DD Length (days)	14	13	13	17	17	14	88
	Nr Trades	1538	923	667	1538	923	667	NA

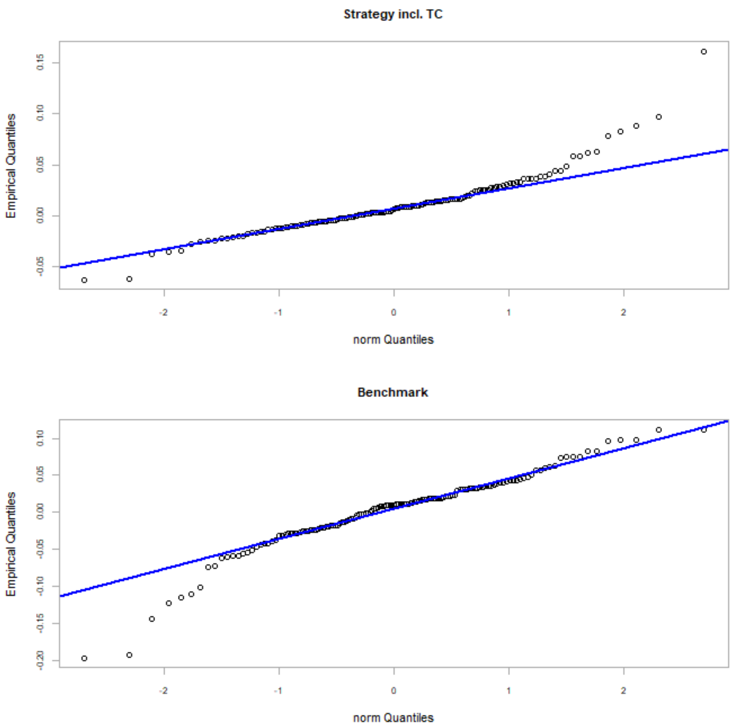
Results of the 24 (six-month) combined out-of-sample trading periods between January 01, 2006 and December 31, 2017. A state space regression model is defined and the Kalman filter algorithm is used to estimate the cointegrating relationship. A 1-day lag is imposed on the signal.

4.1.2 Return Distributions

We present descriptive statistics of the monthly return distributions of the strategy in table 4.3 and 4.4. For the rolling OLS specification, the mean return for the different parameterisations varies between 0.40% and 1.38% before transaction costs and 0.34% and 1.03% after transaction costs. For the Kalman filter specification, the mean return for the different parameterisations varies between 0.47% and 1.61% before transaction costs and 0.40% and 1.16% after transaction costs. These numbers can be compared to the benchmark with an average monthly return of 0.26%. We compute Newey-West (NW) standard errors of the means to account for autocorrelation and heteroskedasticity in the return series as proposed by Clegg and Krauss, (2018) and Krauss and Stübinger, (2017). We observe t-statistics well above 2 and often around 3 and 4 for all the parameterisations of the strategy, as opposed to a t-statistic of 0.53 for the benchmark. We note that the worst month (minimum) is between -4% and -7% for the strategy, while the worst month is -19.7% for the benchmark. This finding confirms that the biggest benefit of our strategy is the limited downside risk. Positive skewness of the returns are found to be between 0.56 and 2.87 for the OLS approach, and between 1.15 and 3.83 for the Kalman filter approach. This can be compared to a negative skewness of the benchmark of -1.01 over the period. The high kurtosis observed in the strategy returns suggests leptokurtic distributions, especially for the Kalman filter approach, where we can see values as high as 27. At face value, this would imply heavier tail risk in the strategy compared to the benchmark. However, when we plot QQ-plots (figure 4.3) of the strategy returns and the benchmark returns a more nuanced picture appears. The high leptokurtic tendencies in the strategy are driven by large *positive* outliers affecting the kurtosis heavily in some parameterisations, caused by the 4th power term in the definition (see equation A.18 in Appendix A). In contrast, general leptokurtic tendencies in the benchmark and normal financial returns are driven by *negative* outliers. As noted by Auer, (2018), outliers in a sample may potentially massively increase the kurtosis and drive a behavior where the kurtosis fluctuates quite radically between cases even if the underlying distribution is similar. This is exactly what we observe in our different parameterisations. We also report a Skewness-kurtosis ratio as proposed by Bacon, (2011), where higher values here are better. All our strategy parameterisations have positive values between 0.10 and 0.20, while the benchmark has a negative value of

-0.18. Finally we report the share positive months ranging from 0.62 to 0.70 before transaction costs and from 0.61 to 0.66 after transaction costs for the strategy. This can be compared to the benchmark where the share positive months is 0.59.

FIGURE 4.3: QQ-plot of Strategy vs Benchmark, blue line is the normal distribution



Looking at tables C.5 and C.6 (Appendix C), where we impose a 2-day lag on the signal, we note once again that returns are significantly reduced. The t-statistics remain high for most of the parameterisations due to a low standard deviation and consequently low standard errors compared to the benchmark. Once again the interpretation is that we are reducing downside risk rather than producing high excess returns in the case of a 2-day signal lag, resulting in a risk adjusted outperformance.

TABLE 4.3: Out of Sample Monthly Return Distributions 2006-2017, Rolling OLS & 1-day signal lag

Rolling OLS		*Excl. TC* **Lookback Z-Score**			*Incl. TC* **Lookback Z-Score**			
Entry	**Measure**	**20**	**60**	**250**	**20**	**60**	**250**	**Benchmark**
2.0	Mean return	0.0092	0.0073	0.0040	0.0073	0.0062	0.0034	0.0026
	SE (NW)	0.0024	0.0015	0.0012	0.0024	0.0015	0.0012	0.0050
	t-stat. (NW)	3.7772	4.9358	3.3937	3.0279	4.2745	2.8648	0.5316
	Minimum	-0.0419	-0.0395	-0.0467	-0.0429	-0.0401	-0.0479	-0.1971
	Quartile 1	-0.0028	-0.0018	-0.0048	-0.0043	-0.0026	-0.0052	-0.0227
	Median	0.0063	0.0044	0.0029	0.0044	0.0032	0.0022	0.0102
	Quartile 3	0.0183	0.0170	0.0122	0.0170	0.0155	0.0115	0.0321
	Maximum	0.1153	0.0811	0.0725	0.1117	0.0785	0.0721	0.1123
	SD	0.0210	0.0187	0.0162	0.0207	0.0184	0.0158	0.0504
	Skewness	1.1314	0.5750	0.6626	1.1199	0.5737	0.6034	-1.0089
	Kurtosis	4.1314	1.5000	2.6650	4.1816	1.5687	2.7312	2.7179
	Sk-Ku Ratio	0.1581	0.1273	0.1166	0.1554	0.1251	0.1049	-0.1758
	Share Pos.	0.6923	0.6763	0.6377	0.6364	0.6643	0.6259	0.5874
1.5	Mean return	0.0107	0.0101	0.0061	0.0076	0.0085	0.0051	0.0026
	SE (NW)	0.0027	0.0023	0.0021	0.0027	0.0023	0.0020	0.0050
	t-stat. (NW)	3.8955	4.3531	2.9668	2.8521	3.7077	2.5666	0.5316
	Minimum	-0.0501	-0.0520	-0.0681	-0.0526	-0.0525	-0.0683	-0.1971
	Quartile 1	-0.0025	-0.0031	-0.0079	-0.0051	-0.0040	-0.0083	-0.0227
	Median	0.0081	0.0088	0.0054	0.0037	0.0061	0.0048	0.0102
	Quartile 3	0.0204	0.0189	0.0167	0.0180	0.0171	0.0157	0.0321
	Maximum	0.1517	0.1509	0.1144	0.1469	0.1475	0.1117	0.1123
	SD	0.0263	0.0252	0.0236	0.0257	0.0251	0.0229	0.0504
	Skewness	1.4412	1.4041	0.7944	1.4612	1.3943	0.7542	-1.0089
	Kurtosis	5.3609	6.7089	3.2995	5.5911	6.4688	3.4574	2.7179
	Sk-Ku Ratio	0.1718	0.1441	0.1257	0.1695	0.1467	0.1164	-0.1758
	Share Pos.	0.6853	0.6901	0.6294	0.6084	0.6364	0.6154	0.5874
1.0	Mean return	0.0138	0.0127	0.0067	0.0094	0.0103	0.0053	0.0026
	SE (NW)	0.0036	0.0032	0.0021	0.0035	0.0032	0.0020	0.0050
	t-stat. (NW)	3.8251	3.9360	3.1929	2.7062	3.2419	2.6283	0.5316
	Minimum	-0.0514	-0.0558	-0.0706	-0.0534	-0.0563	-0.0710	-0.1971
	Quartile 1	-0.0050	-0.0052	-0.0068	-0.0086	-0.0074	-0.0069	-0.0227
	Median	0.0086	0.0099	0.0067	0.0028	0.0071	0.0053	0.0102
	Quartile 3	0.0249	0.0241	0.0189	0.0209	0.0206	0.0171	0.0321
	Maximum	0.1693	0.2503	0.1137	0.1628	0.2435	0.1106	0.1123
	SD	0.0315	0.0327	0.0248	0.0307	0.0322	0.0241	0.0504
	Skewness	1.3453	2.8880	0.5995	1.3457	2.8658	0.5592	-1.0089
	Kurtosis	3.5977	18.2851	2.6530	3.7399	17.9017	2.8178	2.7179
	Sk-Ku Ratio	0.2032	0.1352	0.1057	0.1990	0.1366	0.0958	-0.1758
	Share Pos.	0.6853	0.6643	0.6224	0.6154	0.6294	0.6154	0.5874

Monthly Returns of the 24 (six-month) combined out-of-sample trading periods between January 01, 2006 and December 31, 2017. A rolling OLS regression is used to estimate the cointegrating relationship and a 1-day lag is imposed on the signal. Newey-West Standard Errors are used since return distributions are non-normal.

TABLE 4.4: Out of Sample Monthly Return Distributions 2006-2017,
Kalman Filter & 1-day signal lag

Kalman Filter		*Excl. TC* **Lookback Z-Score**			*Incl. TC* **Lookback Z-Score**			
Entry	**Measure**	**20**	**60**	**250**	**20**	**60**	**250**	**Benchmark**
2.0	Mean return	0.0091	0.0078	0.0047	0.0072	0.0067	0.0040	0.0026
	SE (NW)	0.0024	0.0016	0.0018	0.0024	0.0015	0.0018	0.0050
	t-stat. (NW)	3.7505	5.0245	2.5723	3.0036	4.3546	2.3012	0.5316
	Minimum	-0.0626	-0.0448	-0.0395	-0.0636	-0.0458	-0.0401	-0.1971
	Quartile 1	-0.0023	-0.0028	-0.0018	-0.0040	-0.0041	-0.0021	-0.0227
	Median	0.0063	0.0054	0.0019	0.0042	0.0048	0.0013	0.0102
	Quartile 3	0.0187	0.0153	0.0090	0.0170	0.0141	0.0083	0.0321
	Maximum	0.1296	0.1188	0.1548	0.1258	0.1154	0.1507	0.1123
	SD	0.0227	0.0193	0.0188	0.0225	0.0192	0.0184	0.0504
	Skewness	1.1934	1.4848	3.8518	1.1526	1.4442	3.8281	-1.0089
	Kurtosis	6.1620	6.9595	27.7758	5.9447	6.5655	27.6858	2.7179
	Sk-Ku Ratio	0.1298	0.1486	0.1247	0.1284	0.1504	0.1243	-0.1758
	Share Pos.	0.6643	0.6835	0.6357	0.6224	0.6525	0.6279	0.5874
1.5	Mean return	0.0127	0.0104	0.0072	0.0097	0.0088	0.0061	0.0026
	SE (NW)	0.0033	0.0024	0.0023	0.0032	0.0024	0.0022	0.0050
	t-stat. (NW)	3.9021	4.3872	3.1338	3.0556	3.7250	2.7288	0.5316
	Minimum	-0.0618	-0.0448	-0.0427	-0.0629	-0.0458	-0.0438	-0.1971
	Quartile 1	-0.0037	-0.0051	-0.0036	-0.0062	-0.0066	-0.0047	-0.0227
	Median	0.0086	0.0062	0.0052	0.0068	0.0048	0.0040	0.0102
	Quartile 3	0.0262	0.0194	0.0162	0.0206	0.0175	0.0150	0.0321
	Maximum	0.1672	0.1804	0.1735	0.1618	0.1764	0.1682	0.1123
	SD	0.0287	0.0258	0.0226	0.0283	0.0255	0.0221	0.0504
	Skewness	1.4609	2.3579	2.8468	1.4511	2.3474	2.8049	-1.0089
	Kurtosis	5.7909	12.0400	19.1450	5.8223	11.8608	18.7579	2.7179
	Sk-Ku Ratio	0.1656	0.1562	0.1281	0.1639	0.1574	0.1285	-0.1758
	Share Pos.	0.6783	0.6620	0.6596	0.6294	0.6084	0.6241	0.5874
1.0	Mean return	0.0161	0.0137	0.0118	0.0116	0.0110	0.0099	0.0026
	SE (NW)	0.0039	0.0034	0.0027	0.0038	0.0033	0.0026	0.0050
	t-stat. (NW)	4.1056	4.0083	4.4158	3.0720	3.2962	3.8040	0.5316
	Minimum	-0.0567	-0.0497	-0.0480	-0.0578	-0.0506	-0.0490	-0.1971
	Quartile 1	-0.0038	-0.0050	-0.0029	-0.0078	-0.0076	-0.0052	-0.0227
	Median	0.0102	0.0097	0.0115	0.0064	0.0067	0.0076	0.0102
	Quartile 3	0.0317	0.0229	0.0210	0.0281	0.0203	0.0185	0.0321
	Maximum	0.1924	0.2016	0.2016	0.1865	0.1964	0.1942	0.1123
	SD	0.0331	0.0311	0.0275	0.0323	0.0306	0.0268	0.0504
	Skewness	1.3948	1.9953	2.4496	1.3998	2.0032	2.4375	-1.0089
	Kurtosis	4.7722	8.6738	14.4275	5.0649	8.6405	14.3992	2.7179
	Sk-Ku Ratio	0.1788	0.1703	0.1401	0.1730	0.1715	0.1396	-0.1758
	Share Pos.	0.6993	0.6434	0.6993	0.6434	0.6014	0.6643	0.5874

Monthly Returns of the 24 (six-month) combined out-of-sample trading periods between January 01, 2006 and December 31, 2017. A state space regression model is defined and the Kalman filter algorithm is used to estimate the cointegrating relationship. A 1-day lag is imposed on the signal and Newey-West Standard Errors are used since return distributions are non-normal.

4.1.3 Value at Risk Measures

We also present monthly historical value at risk (VaR) and expected shortfall (ES) measures for our strategy compared to the benchmark in tables 4.5 and 4.6. This analysis confirms our findings from the two previous sections that the tail risk on the downside is actually lower for our pairs trading strategy than for the benchmark. Looking at the monthly 95% VaR for our strategy, we get values ranging from -1% to -3%, compared to -7% for the benchmark. This pattern is even more pronounced in the case of historical ES, where the monthly 95% ES for our strategy is consistently between -2% and -4%, while the corresponding value for the benchmark is -13%. A very similar pattern, although even more pronounced, appears in the case of the 99% VaR and ES. These results once again confirm that the strategy seems to have inherently lower risk than a buy-and-hold strategy of the benchmark, not surprising given that the strategy is close to beta neutral over the entire sample period.

TABLE 4.5: OOS Monthly Historical VaR and ES estimates 2006-2017, Rolling OLS & a 1-day signal lag

Rolling OLS		*Excl. TC*			*Incl. TC*			
		Lookback Z-Score			**Lookback Z-Score**			
Entry	**Measure**	**20**	**60**	**250**	**20**	**60**	**250**	**Benchmark**
2.0	Hist. 95% VaR	-0.0206	-0.0215	-0.0209	-0.0222	-0.0223	-0.0211	-0.0734
	Hist. 95% ES	-0.0304	-0.0287	-0.0269	-0.0321	-0.0295	-0.0272	-0.1323
	Hist. 99% VaR	-0.0373	-0.0329	-0.0289	-0.0392	-0.0340	-0.0292	-0.1726
	Hist. 99% ES	-0.0412	-0.0369	-0.0379	-0.0424	-0.0376	-0.0388	-0.1951
1.5	Hist. 95% VaR	-0.0260	-0.0256	-0.0245	-0.0292	-0.0274	-0.0253	-0.0734
	Hist. 95% ES	-0.0345	-0.0373	-0.0379	-0.0369	-0.0384	-0.0385	-0.1323
	Hist. 99% VaR	-0.0376	-0.0462	-0.0457	-0.0390	-0.0476	-0.0466	-0.1726
	Hist. 99% ES	-0.0442	-0.0498	-0.0587	-0.0461	-0.0509	-0.0593	-0.1951
1.0	Hist. 95% VaR	-0.0287	-0.0273	-0.0294	-0.0325	-0.0293	-0.0294	-0.0734
	Hist. 95% ES	-0.0355	-0.0388	-0.0423	-0.0387	-0.0401	-0.0432	-0.1323
	Hist. 99% VaR	-0.0408	-0.0462	-0.0459	-0.0444	-0.0479	-0.0468	-0.1726
	Hist. 99% ES	-0.0462	-0.0518	-0.0585	-0.0495	-0.0527	-0.0592	-0.1951

Monthly historical VaR and ES estimates of the 24 (six-month) combined out-of-sample trading periods between January 01, 2006 and December 31, 2017. A rolling OLS regression is used to estimate the cointegrating relationship and a 1-day lag is imposed on the signal.

TABLE 4.6: OOS Monthly Historical VaR and ES estimates 2006-2017, Kalman Filter & a 1-day signal lag

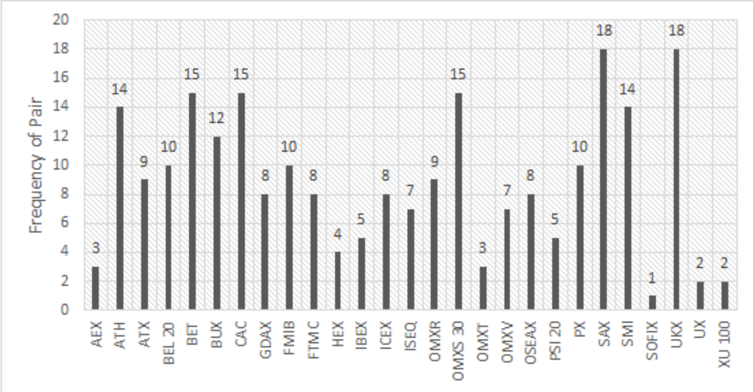
Kalman Filter		*Excl. TC* Lookback Z-Score			*Incl. TC* Lookback Z-Score			
Entry	**Measure**	**20**	**60**	**250**	**20**	**60**	**250**	**Benchmark**
2.0	Hist. 95% VaR	-0.0193	-0.0162	-0.0129	-0.0212	-0.0170	-0.0138	-0.0734
	Hist. 95% ES	-0.0357	-0.0268	-0.0248	-0.0373	-0.0275	-0.0255	-0.1323
	Hist. 99% VaR	-0.0466	-0.0319	-0.0330	-0.0480	-0.0322	-0.0330	-0.1726
	Hist. 99% ES	-0.0579	-0.0391	-0.0368	-0.0593	-0.0397	-0.0371	-0.1951
1.5	Hist. 95% VaR	-0.0210	-0.0174	-0.0210	-0.0246	-0.0193	-0.0211	-0.0734
	Hist. 95% ES	-0.0364	-0.0276	-0.0311	-0.0390	-0.0288	-0.0316	-0.1323
	Hist. 99% VaR	-0.0496	-0.0324	-0.0350	-0.0521	-0.0336	-0.0355	-0.1726
	Hist. 99% ES	-0.0612	-0.0394	-0.0391	-0.0628	-0.0403	-0.0402	-0.1951
1.0	Hist. 95% VaR	-0.0270	-0.0240	-0.0212	-0.0315	-0.0257	-0.0216	-0.0734
	Hist. 95% ES	-0.0369	-0.0344	-0.0319	-0.0402	-0.0358	-0.0329	-0.1323
	Hist. 99% VaR	-0.0460	-0.0411	-0.0431	-0.0485	-0.0425	-0.0439	-0.1726
	Hist. 99% ES	-0.0551	-0.0473	-0.0470	-0.0570	-0.0483	-0.0479	-0.1951

Monthly historical VaR and ES estimates of the 24 (six-month) combined out-of-sample trading periods between January 01, 2006 and December 31, 2017. A state space regression model is defined and the Kalman filter algorithm is used to estimate the cointegrating relationship. A 1-day lag is imposed on the signal.

4.1.4 Frequently Traded Pairs

We now briefly turn to investigate if there are any patterns in the pairs that are selected by our pairs-selection algorithm. Figure 4.4 plots the number of occurrences for each index in a traded pair. The first obvious result is that all 27 indices are included at least once in a pair selected for trading, and that there is no obvious pattern in the data. The results are driven by different pairs throughout the sample, indicating that the likelihood of data snooping is low. The Bulgarian SOFIX index is traded in only one pair, and the Ukrainian UX index and the Turkish XU 100 is traded in two pairs. We also note that the five pairs that are traded the most are the Romanian BET, the French CAC, the Swedish OMXS30, the Slovakian SAX and the British UKX (FTSE250). All these indices are aligned with the London trading session (08:00 to 16:30 UTC), with the exception of the Slovakian SAX index opening at 10:00 UTC and closing at 14:30 UTC. The Baltic indices, were the time discrepancies are most prominent, are some of the least traded. We can thus conclude that our results are not contingent on differences in trading hours or time-zones.

FIGURE 4.4: Number of occurrences in a traded pair

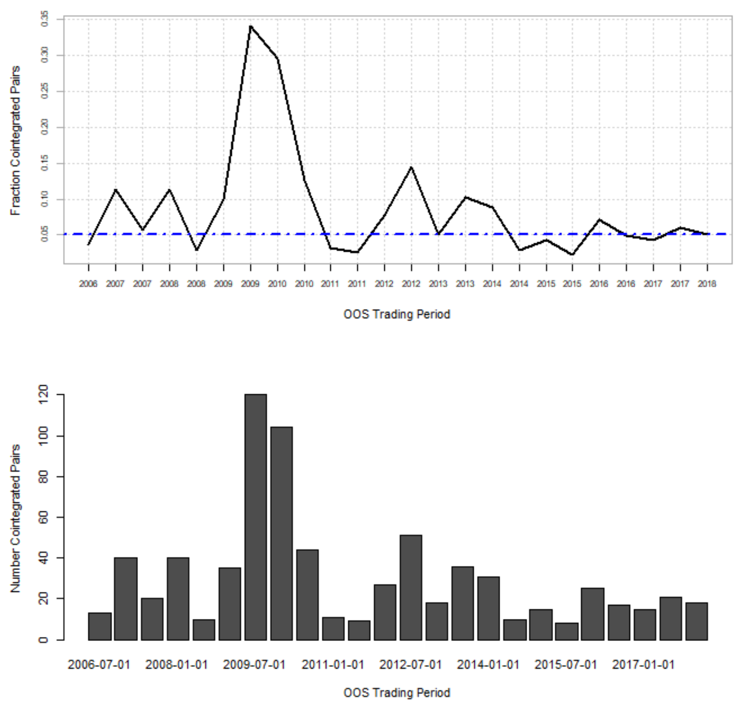


4.1.5 Fraction Cointegrated Pairs over Time

Investigating the fraction and number of pairs passing through the pairs-selection algorithm throughout the sample reveals some interesting results (figure 4.5). There is a distinct pattern of an increased amount of cointegrated pairs during the financial crisis in 2007-2009. A third (120 pairs) of all the pairs are classified as cointegrated during the peak in this period. The next highest peak (51 pairs) occurs in 2012 during the Greek sovereign debt crisis. This result would indicate that common stochastic trends are more pronounced during market crashes. The finding is also in line with the conclusions by Baele, (2005), Jiang et al., (2012) and Gagnon et al., (2016) regarding volatility spillover effects and interdependencies of higher order moments. Gagnon et al., (2016) note that interdependencies between markets increased in persistence and speed of adjustment during the financial crisis. Given that the common stochastic trends are stronger and that the speed of adjustment (of deviations) increase during market turbulence, we have a reasonable explanation as to why our strategy produces significant alpha during these times.

An apparent issue inherent in our methodology is the multiple hypothesis problem. The highlighted blue dotted line in figure 4.5 marks the fraction of false positives that we would expect given the significance level of our cointegration testing. Most of the time, we have only slightly more cointegrated pairs than we would expect false positives. During recent years, the fraction has been close to or under the 5% level, which could potentially explain the decay in alpha we noted earlier. This highlights the importance of ranking the pairs tested for cointegration according to some criterion. Trading all pairs will likely result in poor results since many pairs will appear cointegrated by chance alone.

FIGURE 4.5: Fraction / number of Cointegrated pairs 2006-2017



4.2 Robustness of the Strategy

We have now presented the main results of the long term backtest, and will now perform a battery of robustness tests to see how sensitive the results are to changes in certain key assumptions affecting the profitability. We use a entry threshold of $\pm$ 2.0SD and a lookback period for the normalized spread of 20 periods for this analysis (with exception for weekly case when we use a lookback of 10 periods). A condensed performance summary of the robustness tests can be found in table 4.7.

4.2.1 Random Pairs

To check our results against data snooping as well as to confirm the added value of our selection algorithm, we test our strategy by feeding it five random pairs for each trading period, meaning that no cointegration testing is conducted. It is quite obvious by simple visual inspection of figure B.4 in Appendix B that the strategy breaks when we do not test our pairs for cointegration before trading them. The annualized returns are negative after transaction costs and even worse than a buy-and-hold strategy as seen in table 4.7. An interesting observation however, is that our strategy still seems to have a lower standard deviation than the benchmark. This result has been noted before and is not surprising given that the strategy is still market-neutral, even if pairs are selected at random.

4.2.2 Euro Area Subsample

To examine the possibility that short-term currency fluctuations could be a main driver of the results we have found, we narrow down the sample to the 15 countries in our sample that have their indices denominated in euro (see table C.1 in Appendix C). We apply the strategy in the exact same manner, with a slight practical adjustment. Our selection algorithm now only performs a test for stationarity on the residuals and then picks the five pairs with the lowest p-value, i.e. it does not fit a $AR(1)$-model to the spread as previously to formally test for cointegration and sort by the AR-coefficient ρ. This simplification is made since only 15 assets significantly reduces the amount of possible pairs (105) and thus the number of suitable candidates that passes through our selection algorithm. Even with this simplification in the selection process, we see that our strategy performs quite well as seen in figure B.5 in Appendix B. The strategy has a annualized information ratio of 1.03 before transaction costs and 0.77 after transaction costs as seen in table 4.7.

4.2.3 Weekly Returns

Given the sensitivity of the strategy to the time-lag between signal and execution on the daily time-frame, we transform our data series to weekly observations and run the strategy. On a weekly time-frame it is hard to argue that one would not have time to execute the trade in a timely manner, and waiting one period simply does not make sense. However, we still perform the cointegration testing on daily data to make sure that there is consistency in the pairs that are selected and to make sure that our cointegration tests have enough statistical power. The lookback periods naturally have to be modified somewhat to fit our lower frequency data. The year-long 250-day lookback period for the regression is changed to 52 weeks and the moving average used for the Z-score is changed to 10. Looking at figure B.6 in Appendix B we can conclude that the strategy seems to work just as well on a weekly time frame as on the daily. The strategy has an annualized information ratio of 0.81 before transaction costs and 0.70 after transaction costs as seen in table 4.7.

TABLE 4.7: Robustness tests

	Random Pairs		**EUR Subsample**		**Weekly Returns**		**Benchmark**
Measure	*Excl. TC*	*Incl. TC*	*Excl. TC*	*Incl. TC*	*Excl. TC*	*Incl. TC*	
Annualized Returns	0.013	-0.007	0.088	0.066	0.048	0.041	0.018
Annualized SD	0.076	0.076	0.086	0.085	0.060	0.059	0.211
Maximum Drawdown	0.191	0.258	0.126	0.142	0.065	0.071	0.644
Annualized IR	0.165	-0.095	1.030	0.774	0.809	0.699	0.088
Risk-Return-Ratio	0.066	-0.028	0.700	0.467	0.742	0.584	0.029

5. Conclusion

This paper explores the area of cross-market pairs trading using daily observations of 27 European equity indices between 2006 and 2017. The cointegration approach for pairs trading as first proposed by Vidyamurthy, (2004) is used in the pairs selection and formation of the spread. The cointegrating relationship is estimated using a rolling OLS approach and a Kalman filter approach.

High annualized returns (between 3.9% and 13.3%) and information ratios (between 0.52 and 1.29) are observed for all the parameter specifications after transaction costs. The strategy clearly outperforms the benchmark portfolio - the latter with an annualized return of 1.8% and information ratio of 0.09. Most notably, risk seems to be inherently lower in the pairs trading strategy compared to the benchmark. This is not surprising given that the returns are close to beta-neutral across the entire out-of-sample period. The strategy also produces significant alpha controlling for a systematic risk factor defined as the excess returns of the benchmark portfolio. Most alpha is produced during the financial crisis of 2007-2009, and we observe a pattern where outperformance seems to be positively related to volatility. We find that the number of cointegrated pairs are significantly higher during this period, potentially indicating that the persistence and strength of a common stochastic trend is increased during market crashes. If this is the case, we would expect our chosen pairs to have better performance during such times. The performance of the strategy does however seem to be decaying during recent years, perhaps as more arbitrageurs employ similar strategies and markets become more efficient. The monthly return distributions are positively skewed and leptokurtic, although large positive outliers rather than negative outliers seem to drive the kurtosis in the sample. Given this fact, historical value at risk and expected shortfall estimates indicate lower tail risk in the strategy than in the benchmark. Lastly, the results confirm that pairs trading returns are highly sensitive to the time-lag between the signal and the executed trade. We do however show using lower frequency (weekly) data that the strategy is not contingent on instant execution, and is viable when trading only once a week.

Future studies could further investigate the sensitivity of cross-market pairs trading to time-related discrepancies such as opening hours and time-zones. Constructing a dataset that carefully aligns all observations in time would certainly be an improvement. Closer investigation of how foreign exchange fluctuations affects the results would also be desirable. In addition, it is possible to extend the analysis using the Johansen, (1988) method to a case where we form portfolios rather than pairs to trade, where multiple cointegrating vectors are present.

References

Alexander, Carol (2001). *Market models: A guide to financial data analysis*. John Wiley & Sons.

Auer, Martin (2018). "Expected Shortfall". *Hands-On Value-at-Risk and Expected Shortfall*. Springer, pp. 39–40.

Avellaneda, Marco and Jeong-Hyun Lee (2010). "Statistical arbitrage in the US equities market". *Quantitative Finance* 10.7, pp. 761–782.

Bacon, Carl R (2011). *Practical portfolio performance measurement and attribution*. Vol. 568. John Wiley & Sons.

Baele, Lieven (2005). "Volatility spillover effects in European equity markets". *Journal of Financial and Quantitative Analysis* 40.2, pp. 373–401.

Bertram, William K (2010). "Analytic solutions for optimal statistical arbitrage trading". *Physica A: Statistical Mechanics and its Applications* 389.11, pp. 2234–2243.

Bowen, David, Mark Hutchinson, and Niall O'Sullivan (2010). "High frequency equity pairs trading: transaction costs, speed of execution and patterns in returns".

Burgess, Andrew Neil et al. (2000). "A computational methodology for modelling the dynamics of statistical arbitrage". PhD thesis. University of London.

Caldeira, João and Guilherme Moura (2013). "Selection of a portfolio of pairs based on cointegration: A statistical arbitrage strategy".

Clegg, Matthew and Christopher Krauss (2018). "Pairs trading with partial cointegration". *Quantitative Finance* 18.1, pp. 121–138.

Cochrane, John H (1988). "How big is the random walk in GNP?" *Journal of political economy* 96.5, pp. 893–920.

Cummins, Mark and Andrea Bucca (2012). "Quantitative spread trading on crude oil and refined products markets". *Quantitative Finance* 12.12, pp. 1857–1875.

Das, Atanu and Tapan Kumar Ghoshal (2010). "Market risk beta estimation using adaptive Kalman filter". *International Journal of Engineering Science and Technology* 2.6, pp. 1923–1934.

Do, Binh and Robert Faff (2010). "Does simple pairs trading still work?" *Financial Analysts Journal* 66.4, pp. 83–95.

Do, Binh and Robert Faff (2012). “Are pairs trading profits robust to trading costs?” *Journal of Financial Research* 35.2, pp. 261–287.

Do, Binh, Robert Faff, and Kais Hamza (2006). “A new approach to modeling and estimation for pairs trading”. *Proceedings of 2006 Financial Management Association European Conference*. Citeseer, pp. 87–99.

Dunis, Christian L and Richard Ho (2005). “Cointegration portfolios of European equities for index tracking and market neutral strategies”. *Journal of Asset Management* 6.1, pp. 33–52.

Dunis, Christian L et al. (2010). “Statistical arbitrage and high-frequency data with an application to Eurostoxx 50 equities”. *Liverpool Business School, Working paper*.

Elliott, Robert J, John Van Der Hoek*, and William P Malcolm (2005). “Pairs trading”. *Quantitative Finance* 5.3, pp. 271–276.

Engle, Robert F and Clive WJ Granger (1987). “Co-integration and error correction: representation, estimation, and testing”. *Econometrica: journal of the Econometric Society*, pp. 251–276.

Froot, Kenneth A and Emil M Dabora (1999). “How are stock prices affected by the location of trade?” *Journal of Financial economics* 53.2, pp. 189–216.

Gagnon, Marie-Hélène, Gabriel J Power, and Dominique Toupin (2016). “International stock market cointegration under the risk-neutral measure”. *International Review of Financial Analysis* 47, pp. 243–255.

Gatev, Evan, William N Goetzmann, and K Geert Rouwenhorst (2006). “Pairs trading: Performance of a relative-value arbitrage rule”. *The Review of Financial Studies* 19.3, pp. 797–827.

Girma, Paul Berhanu and Albert S Paulson (1999). “Risk arbitrage opportunities in petroleum futures spreads”. *Journal of Futures Markets* 19.8, pp. 931–955.

Goodwin, Thomas H (1998). “The information ratio”. *Financial Analysts Journal* 54.4, pp. 34–43.

Hong, Gwangheon and Raul Susmel (2012). “Pairs-Trading in the Asian ADR Market: Returns and Convergence”. *University of Houston, working paper*.

Huck, Nicolas (2009). “Pairs selection and outranking: An application to the S&P 100 index”. *European Journal of Operational Research* 196.2, pp. 819–825.
— (2010). “Pairs trading and outranking: The multi-step-ahead forecasting case”. *European Journal of Operational Research* 207.3, pp. 1702–1716.
— (2015). “Pairs trading: does volatility timing matter?” *Applied economics* 47.57, pp. 6239–6256.

Huck, Nicolas and Komivi Afawubo (2015). "Pairs trading and selection methods: is cointegration superior?" *Applied Economics* 47.6, pp. 599–613.

Jegadeesh, Narasimhan (1990). "Evidence of predictable behavior of security returns". *The Journal of finance* 45.3, pp. 881–898.

Jegadeesh, Narasimhan and Sheridan Titman (1993). "Returns to buying winners and selling losers: Implications for stock market efficiency". *The Journal of finance* 48.1, pp. 65–91.

Jiang, George J, Eirini Konstantinidi, and George Skiadopoulos (2012). "Volatility spillovers and the effect of news announcements". *Journal of Banking & Finance* 36.8, pp. 2260–2273.

Johansen, Søren (1988). "Statistical analysis of cointegration vectors". *Journal of economic dynamics and control* 12.2-3, pp. 231–254.

Johnsson, Richard CB (2010). "A Simple Risk-Return Ratio".

Jurek, Jakub W and Halla Yang (2007). "Dynamic portfolio selection in arbitrage".

Kasa, Kenneth (1992). "Common stochastic trends in international stock markets". *Journal of monetary Economics* 29.1, pp. 95–124.

Krauss, Christopher (2017). "Statistical arbitrage pairs trading strategies: Review and outlook". *Journal of Economic Surveys* 31.2, pp. 513–545.

Krauss, Christopher and Johannes Stübinger (2017). "Non-linear dependence modelling with bivariate copulas: Statistical arbitrage pairs trading on the S&P 100". *Applied Economics* 49.52, pp. 5352–5369.

Liew, Rong Qi and Yuan Wu (2013). "Pairs trading: A copula approach". *Journal of Derivatives & Hedge Funds* 19.1, pp. 12–30.

Litterman, Bob et al. (2004). *Modern investment management: an equilibrium approach*. Vol. 246. John Wiley & Sons.

Liu, Jun and Allan Timmermann (2013). "Optimal convergence trade strategies". *The Review of Financial Studies* 26.4, pp. 1048–1086.

Ljung, Greta M and George EP Box (1978). "On a measure of lack of fit in time series models". *Biometrika* 65.2, pp. 297–303.

Lo, Andrew W and A Craig MacKinlay (1988). "Stock market prices do not follow random walks: Evidence from a simple specification test". *The review of financial studies* 1.1, pp. 41–66.

Mina, Jorge, Jerry Yi Xiao, et al. (2001). "Return to RiskMetrics: the evolution of a standard". *RiskMetrics Group*.

Moura, Carlos Eduardo de, Adrian Pizzinga, and Jorge Zubelli (2016). "A pairs trading strategy based on linear state space models and the Kalman filter". *Quantitative Finance* 16.10, pp. 1559–1573.

Mylonidis, Nikolaos and Christos Kollias (2010). "Dynamic European stock market convergence: Evidence from rolling cointegration analysis in the first euro-decade". *Journal of Banking & Finance* 34.9, pp. 2056–2064.

Neumaier, Arnold and Tapio Schneider (1998). "Multivariate autoregressive and Ornstein-Uhlenbeck processes: estimates for order, parameters, spectral information, and confidence regions". *ACM Transactions in Mathematical Software*.

Pantula, Sastry G, Graciela Gonzalez-Farias, and Wayne A Fuller (1994). "A comparison of unit-root test criteria". *Journal of Business & Economic Statistics* 12.4, pp. 449–459.

Rad, Hossein, Rand Kwong Yew Low, and Robert Faff (2016). "The profitability of pairs trading strategies: distance, cointegration and copula methods". *Quantitative Finance* 16.10, pp. 1541–1558.

Stander, Yolanda, Daniël Marais, and Ilse Botha (2013). "Trading strategies with copulas". *Journal of Economic and Financial Sciences* 6.1, pp. 83–107.

Vidyamurthy, Ganapathy (2004). *Pairs Trading: quantitative methods and analysis*. Vol. 217. John Wiley & Sons.

Xie, Wenjun and Yuan Wu (2013). "Copula-based pairs trading strategy". *Asian Finance Association (AsFA) 2013 Conference. doi*. Vol. 10.

A. Mathematical Appendix

A.1 Variance Ratio Test

We define the most rudimentary relative price series possible for a cointegrated pair, namely the ratio of log-prices BEL 20 / FTSE 250. A variance ratio test is designed to identify when a time series deviates from random walk characteristics, i.e. when it does not follow the non-stationary dynamics where observations are completely uncorrelated. The logic is in many ways similar to autocorrelation tests such as the Ljung-Box test. In fact, the variance ratio can be expressed as a linear combination of autocorrelation coefficients in a special case (Lo and MacKinlay, 1988). We define the variance ratio statistic in equation A.1 (Cochrane, 1988; Lo and MacKinlay, 1988). The ratio can be seen as the normalized ratio of long term variance (calculated over period τ) to a one period variance (calculated over period t) scaled by τ.

$$VR(\tau) = \frac{\sum_t (\Delta^\tau y_t - \overline{\Delta^\tau y})^2}{\tau \sum_t (\Delta y_t - \overline{\Delta y})^2} \tag{A.1}$$

The variance ratio statistics can be viewed collectively using a variance ratio function (VRF). A positive (negative) gradient of the VRF indicates positive (negative) autocorrelation and trending (mean reverting or cyclical) behavior. For a random walk, variance will grow linearly with the period τ, giving us a ratio that should be close to one over time. A trending time series grows at a non-linear rate where the VRF increases above one over time. A mean-reverting series declines at a non-linear rate where the VRF decreases below one over time. More intuitively, a variance ratio under one shows us that volatility present in short-term price dynamics is not reflected in long term volatility, telling us that there has to be a mean reverting component in the price series. The VRFs in figure B.1 tells us that the relative log-price series BEL 20 / FTSE 250 indeed exhibits strong mean-reverting characteristics, and that the two price series individually exhibit trending behavior.

A.2 The Kalman Filter Algorithm

Define $\theta_t = (\mu_t, \gamma_t)'$ as the state vector from equation 3.7. I also define $y_t = ln(P_t^A)$ and $x_t = ln(P_t^B)$ for simplicity in the notation. The equation system can then be rewritten as:

$$y_t = \begin{pmatrix} 1 & x_t \end{pmatrix} \theta_t + v_t \tag{A.2}$$

$$\begin{pmatrix} \mu_t \\ \gamma_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \gamma_{t-1} \end{pmatrix} + \begin{pmatrix} w_{\mu,t} \\ w_{\gamma,t} \end{pmatrix} \tag{A.3}$$

The system matrices in our state space specification is then:

$$\mathbf{F}_t = \begin{pmatrix} 1 & x_t \end{pmatrix}, \quad \mathbf{V}_t = \sigma_v^2 \tag{A.4}$$

$$\mathbf{G} = \mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} \sigma_\mu^2 & 0 \\ 0 & \sigma_\gamma^2 \end{pmatrix} \tag{A.5}$$

where $\mathbf{F}_t$ is time varying. Our state space model can now be written as:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{F}_t\theta_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim N(\mathbf{0}, \mathbf{V}_t) \\ \theta_t &= \mathbf{G}_t\theta_{t-1} + \mathbf{w}_t, \qquad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t) \end{aligned} \tag{A.6}$$

We also need to defined the initial distribution for θ_t:

$$\theta_0 \sim N(m_0, C_0) \tag{A.7}$$

$$\mathbf{m}_0 = \mathbf{0}_{2,1} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \mathbf{C}_0 = k \cdot \mathbf{I}_2, \quad k = 10^7 \tag{A.8}$$

In the state space model, the state vector θ_t is the signal (including μ_t and γ_t), and $\mathbf{w}_t$ is the measurement noise. Given the noisy observations of $y_1, y_2, ..., y_T$ we want to extract the optimal signal (estimate of θ_t) given the information available at t, $\mathbf{I}_t = (y_1, y_2, ..., y_T)$. The Kalman filter is a recursive algorithm determining the optimal estimates for θ_t using a set of *prediction equations* and a set of *updating equations*. We first have to define:

$$\begin{aligned} \mathbf{m}_t &= E[\theta_t \mid \mathbf{I}_t] = \text{filtered (optimal) estimate of } \theta_t \\ \mathbf{C}_t &= E[(\theta_t - \mathbf{m}_t)(\theta_t - \mathbf{m}_t)' \mid \mathbf{I}_t] = \text{MSE matrix of } m_t \end{aligned} \tag{A.9}$$

The prediction equations are presented below in A.10-A.12. Given $\mathbf{m}_{t-1}$ and $\mathbf{C}_{t-1}$ at time $t-1$, the optimal prediction of θ_t and its corresponding MSE matrix are:

$$\begin{aligned} \mathbf{m}_{t|t-1} &= E[\theta_t \mid \mathbf{I}_t] = \mathbf{G}_t\mathbf{m}_{t-1} \\ \mathbf{C}_{t|t-1} &= E[(\theta_t - \mathbf{m}_t)(\theta_t - \mathbf{m}_t)' \mid \mathbf{I}_t] = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t' + \mathbf{W}_t \end{aligned} \tag{A.10}$$

The associated optimal predictor of $\mathbf{y}_t$ given information at $t-1$ is:

$$\mathbf{y}_{t|t-1} = E[\mathbf{y}_t \mid \mathbf{I}_{t-1}] = \mathbf{F}_t\mathbf{m}_{t|t-1} \tag{A.11}$$

The prediction error and its MSE matrix are then:

$$\begin{aligned} \mathbf{e}_t &= \mathbf{y}_t - \mathbf{y}_{t|t-1} = \mathbf{y}_t - \mathbf{F}_t\mathbf{m}_{t|t-1} = \mathbf{F}_t(\theta_t - \mathbf{m}_{t|t-1}) + \mathbf{v}_t \\ E[\mathbf{e}_t\mathbf{e}_t'] &= \mathbf{Q}_t = \mathbf{F}_t\mathbf{C}_{t|t-1}\mathbf{F'}_t + \mathbf{V}_t \end{aligned} \tag{A.12}$$

The updating equations are presented below in A.13. When an observation $\mathbf{y}_t$ is available, the prediction and its MSE matrix can be updated using:

$$\begin{aligned}\mathbf{m}_t &= \mathbf{m}_{t|t-1} + \mathbf{C}_{t|t-1}\mathbf{F}'_t\mathbf{Q}_t^{-1}(\mathbf{y}_t - \mathbf{F}_t\mathbf{m}_{t|t-1}) = \mathbf{m}_{t|t-1} + \mathbf{C}_{t|t-1}\mathbf{F}'_t\mathbf{Q}_t^{-1}\mathbf{v}_t \\ \mathbf{C}_t &= \mathbf{C}_{t|t-1} - \mathbf{C}_{t|t-1}\mathbf{F}'_t\mathbf{Q}_t^{-1}\mathbf{F}_t\mathbf{C}_{t|t-1}\end{aligned} \tag{A.13}$$

The *Kalman gain* matrix is $\mathbf{K}_t = \mathbf{C}_{t|t-1}\mathbf{F}'_t\mathbf{Q}_t^{-1}$, giving a weight to the new information $\mathbf{e}_t = \mathbf{y}_t - \mathbf{F}_t\mathbf{m}_{t|t-1}$ to update the equation for $\mathbf{m}_t$.

A.3 Formulas

Annualized returns and volatility

$$r^A = 252 * \frac{1}{N}\sum_{t=1}^{N} r_{it} \tag{A.14}$$

$$\hat{\sigma}^A = \sqrt{252} * \sqrt{\frac{1}{N-1} * \sum_{t=1}^{N}(r_{it} - \bar{r})^2} \tag{A.15}$$

where N is the total number of periods, 252 scales the daily returns and $\sqrt{252}$ scales the daily volatility to annualized measures.

Transform log-returns to simple returns

$$R_{it} = e^{r}_{it} - 1 \tag{A.16}$$

Skewness and kurtosis measures

$$Skewness = \frac{1}{n}\sum_{i=1}^{n}(\frac{r_i - \bar{r}}{\sigma_p})^3 \tag{A.17}$$

$$Kurtosis = \frac{1}{n}\sum_{i=1}^{n}(\frac{r_i - \bar{r}}{\sigma_p})^4 \tag{A.18}$$

$$SkewnessKurtosisRatio = \frac{Skewness}{Kurtosis} \tag{A.19}$$

B. Figures

FIGURE B.1: Variance Ratio Function of BEL 20, FTSE 250 and relative log-price series BEL 20 / FTSE 250

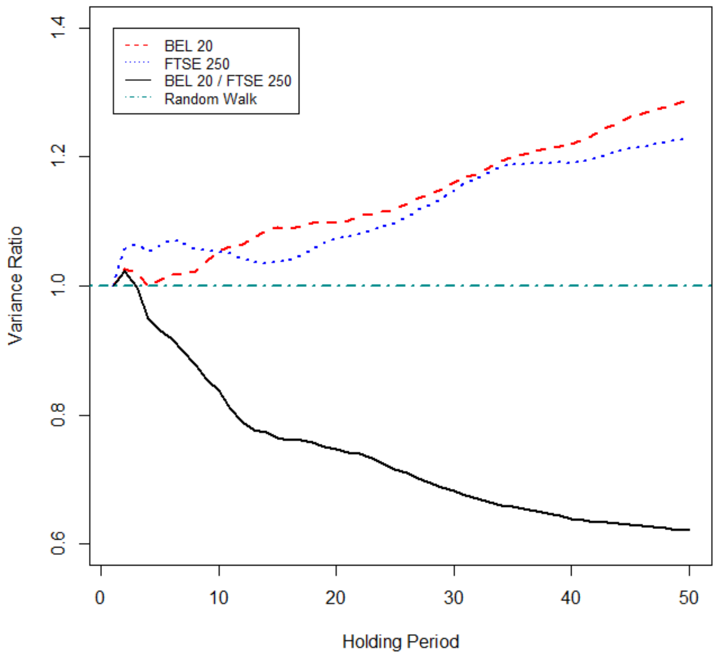


FIGURE B.2: Risk-Return characteristics of strategy vs benchmark. Threshold $\pm$ 1.5 *SD*, 60 period lookback and a 1-day signal lag. Grey lines represent Information Ratio 1 and 2.

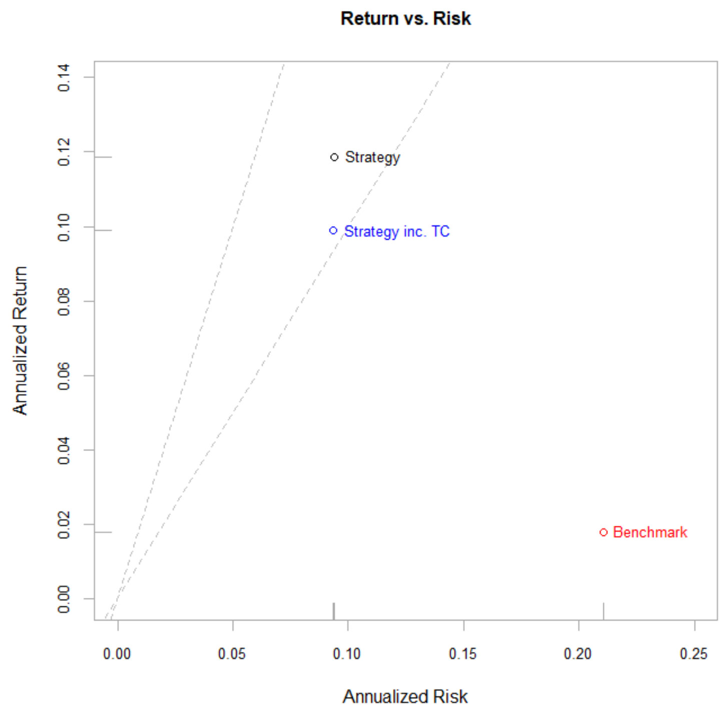


FIGURE B.3: 250-Day rolling performance of strategy. Threshold $\pm$ 1.5 *SD*, 60 period lookback and a 1-day signal lag.

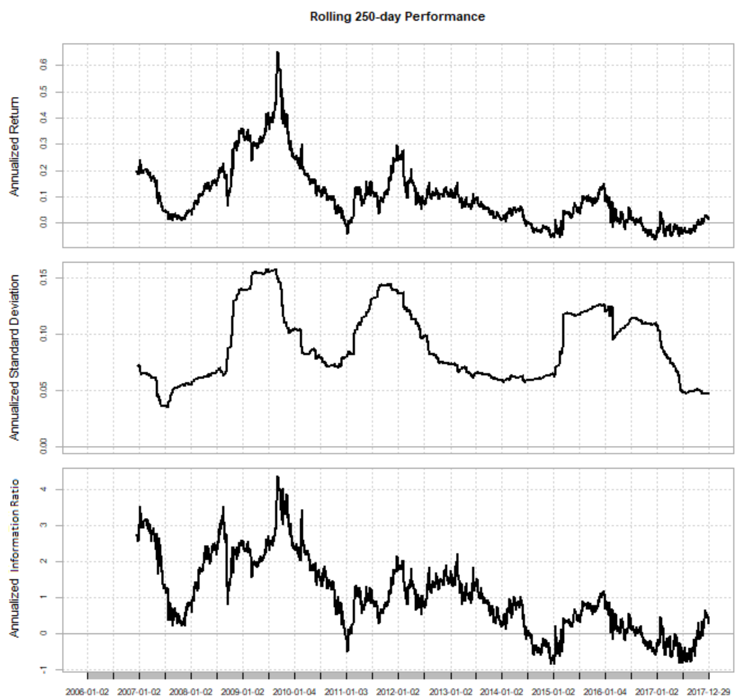


FIGURE B.4: **Random Pairs:** Out-of-sample strategy performance including transaction costs. Threshold $\pm$ 2 *SD* and 20 period lookback. A 1-day lag is imposed on the signal

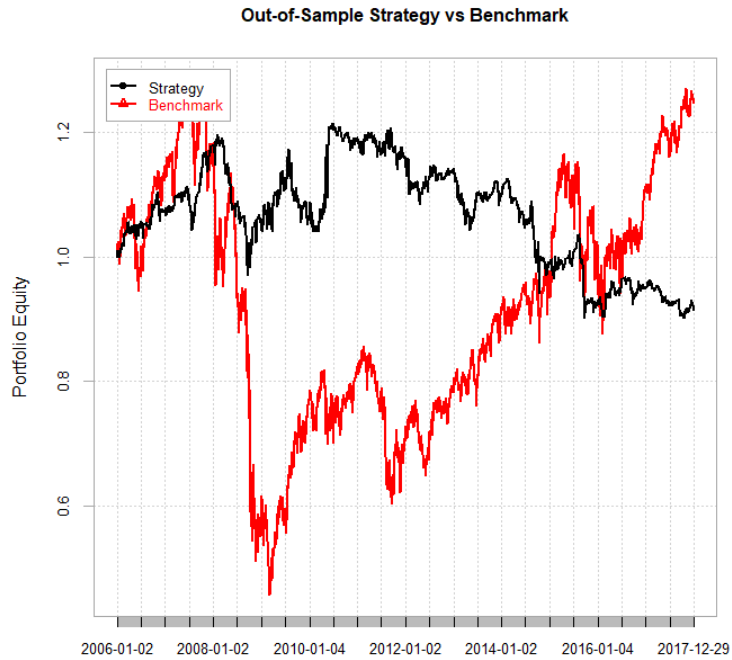


FIGURE B.5: **EUR subsample:** Out-of-sample strategy performance including transaction costs. Threshold $\pm$ 2 *SD* and 20 period lookback. A 1-day lag is imposed on the signal

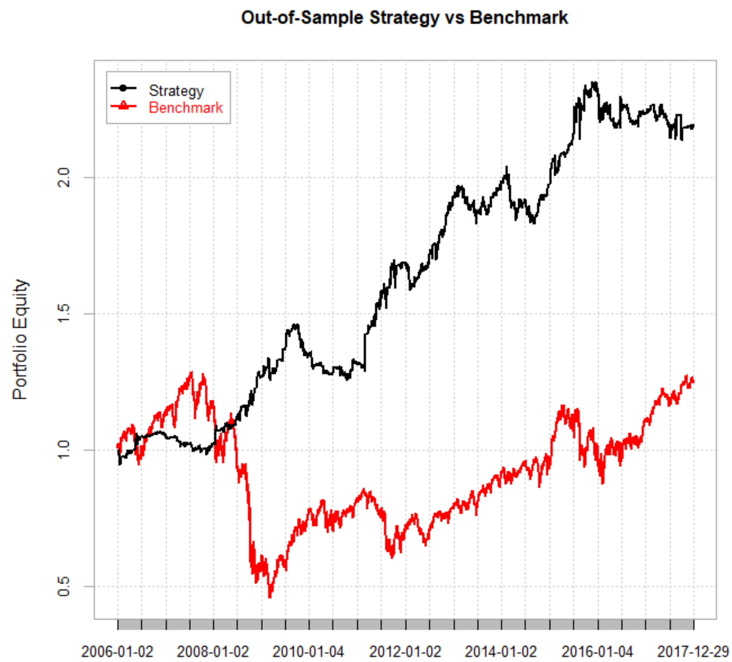
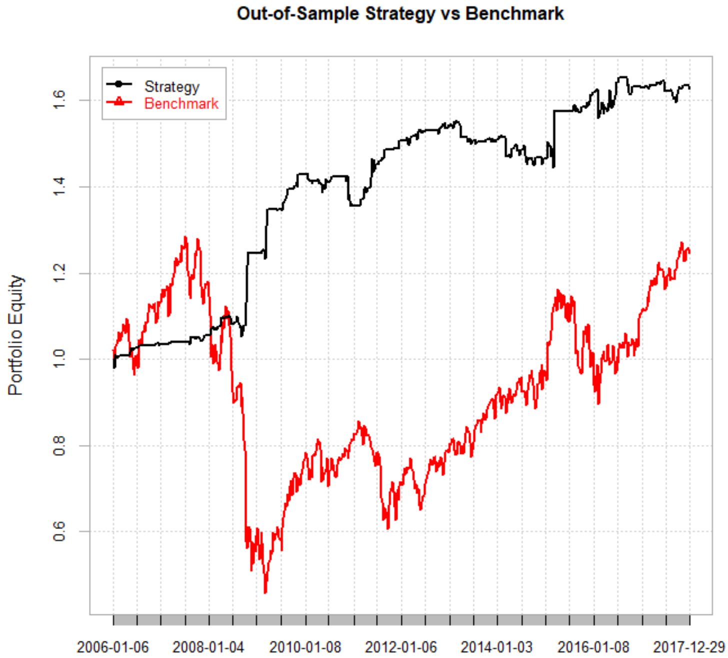


FIGURE B.6: **Weekly:** Out-of-sample strategy performance including transaction costs. Threshold $\pm$ 2 *SD* and 10 period lookback.
A 1-day lag is imposed on the signal



C. Tables

TABLE C.1: Indices included in the dataset

		Full Sample		
Nr	**Symbol**	**Name**	**Country**	**Currency**
1	AEX	Amsterdam Exchange Index	Netherlands	EUR
2	ATH	ATHEX Composite Index	Greece	EUR
3	ATX	Austrian Traded Index	Austria	EUR
4	BEL 20	Belgium 20	Belgium	EUR
5	BET	BET Index	Romania	RON
6	BUX	BUX Index	Hungary	HUF
7	CAC	CAC 40	France	EUR
8	DAX	DAX Index	Germany	EUR
9	FMIB	FTSE MIB Index	Italy	EUR
10	FTMC	FTSE 250 Index	United Kingdom	GBP
11	HEX	OMX Helsinki Index	Finland	EUR
12	IBEX	IBEX Index	Spain	EUR
13	ICEX	OMX Iceland All Share Index	Iceland	ISK
14	ISEQ	Irish Stock Exchange Overall Index	Ireland	EUR
15	OMXR	OMX Riga Index	Latvia	EUR
16	OMXS 30	OMX Stockholm 30 Index	Sweden	SEK
17	OMXT	OMX Talinn Index	Estonia	EUR
18	OMXV	OMX Vilnius Index	Lithuania	EUR
19	OSEAX	OSE All Share Index	Norway	NOK
20	PSI 20	PSI 20 Index	Portugal	EUR
21	PX	PX Index	Czech Republic	CZK
22	SAX	SAX Index	Slovakia	EUR
23	SMI	Swiss Market Index	Switzerland	CHF
24	SOFIX	SOFIX Index	Bulgaria	BGN
25	UKX	FTSE 100 Index	United Kingdom	GBP
26	UX	UX Index	Unkraine	UAH
27	XU 100	XU 100 Index	Turkey	TRY

TABLE C.2: Opening Hours, European Exchanges included in the Sample

			Local Time		*UTC*	
Name	**UTC**	**Summer UTC**	**Open**	**Close**	**Open**	**Close**
Turkey Borsa Istanbul	+3	Mar–Oct	09:30	17:30	07:30	15:30
Bulgarian Stock Exchange	+2	Mar–Oct	10:00	17:00	08:00	15:00
Finland Helsinki Stock Exchange	+2	Mar–Oct	10:00	18:30	08:00	16:30
Ukraine Ukrainian Exchange	+2	Mar–Oct	10:00	17:30	08:00	15:30
Latvia Riga Stock Exchange	+2	Mar–Oct	10:00	16:00	08:00	14:00
Estonia Tallinn Stock Exchange	+2	Mar–Oct	10:00	16:00	08:00	14:00
Lithuania NASDAQ OMX Vilnius	+2	Mar–Oct	10:00	16:00	08:00	14:00
Greece Athens Stock Exchange	+2	Mar–Oct	10:00	17:20	08:00	15:20
Romania Bucharest Stock Exchange	+1	Mar–Oct	09:45	18:00	07:45	16:00
Germany Frankfurt Stock Exchange (Xetra)	+1	Mar–Oct	08:00	20:00	07:00	19:00
Austria Wiener Börse AG	+1	Mar–Oct	08:55	17:35	07:55	16:35
Belgium Euronext Brussels	+1	Mar–Oct	09:00	17:30	08:00	16:30
Hungary Budapest Stock Exchange	+1	Mar–Oct	09:00	17:00	08:00	16:00
France Euronext Paris	+1	Mar–Oct	09:00	17:30	08:00	16:30
Switzerland Swiss Exchange	+1	Mar–Oct	09:00	17:30	08:00	16:30
Spain Spanish Stock Exchange	+1	Mar–Oct	09:00	17:30	08:00	16:30
Italy Milan Stock Exchange	+1	Mar–Oct	09:00	17:35	08:00	16:35
Netherlands Euronext Amsterdam	+1	Mar–Oct	09:00	17:40	08:00	16:40
Czech Republic Prague Stock Exchange	+1	Mar–Oct	09:00	16:30	08:00	15:30
Sweden Stockholm Stock Exchange	+1	Mar–Oct	09:00	17:30	08:00	16:30
Norway Oslo Stock Exchange	+1	Mar–Oct	09:00	16:30	08:00	15:30
Slovakia Bratislava Stock Exchange	+1	Mar–Oct	11:00	15:30	10:00	14:30
United Kingdom London Stock Exchange	0	Mar–Oct	08:00	16:30	08:00	16:30
Irish Stock Exchange	0	Mar–Oct	08:00	16:30	08:00	16:30
Iceland Stock Exchange	0	No	09:30	15:30	09:30	15:30
Portugal Euronext Lisbon	0	Mar–Oct	08:00	16:30	08:00	16:30

TABLE C.3: Out of Sample Results 2006-2017, Rolling OLS & 2-day lag

	Rolling OLS	*Excl. TC*			*Incl. TC*			
		Lookback Z-Score			**Lookback Z-Score**			
Entry	**Measure**	**20**	**60**	**250**	**20**	**60**	**250**	**Benchmark**
2.0	Annualized Returns	0.064	0.054	0.023	0.044	0.043	0.018	0.018
	Annualized SD	0.079	0.079	0.072	0.079	0.079	0.072	0.211
	Maximum Drawdown	0.103	0.102	0.130	0.142	0.103	0.135	0.645
	Average Drawdown	0.012	0.013	0.013	0.012	0.013	0.013	0.036
	Annualized IR	0.804	0.686	0.322	0.556	0.549	0.247	0.085
	Risk-Return-Ratio	0.618	0.531	0.179	0.309	0.420	0.132	0.028
	Avg. DD Length (days)	28	25	48	33	27	51	88
	Nr Trades	648	369	193	648	369	193	NA
1.5	Annualized Return	0.071	0.082	0.042	0.038	0.064	0.033	0.018
	Annualized SD	0.095	0.096	0.090	0.095	0.095	0.089	0.211
	Maximum Drawdown	0.136	0.129	0.124	0.255	0.135	0.132	0.645
	Average Drawdown	0.014	0.014	0.015	0.015	0.016	0.016	0.036
	Annualized IR	0.743	0.854	0.468	0.403	0.671	0.370	0.085
	Risk-Return-Ratio	0.517	0.632	0.338	0.150	0.472	0.251	0.028
	Avg. DD Length (days)	26	19	34	33	24	38	88
	Nr Trades	1010	538	304	1010	538	304	NA
1.0	Annualized Returns	0.079	0.093	0.048	0.033	0.066	0.033	0.018
	Annualized SD	0.104	0.106	0.101	0.103	0.106	0.101	0.211
	Maximum Drawdown	0.147	0.137	0.131	0.270	0.149	0.162	0.645
	Average Drawdown	0.017	0.017	0.015	0.022	0.020	0.016	0.036
	Annualized IR	0.765	0.876	0.472	0.315	0.624	0.330	0.085
	Risk-Return-Ratio	0.539	0.676	0.362	0.121	0.442	0.205	0.028
	Avg. DD Length (days)	25	23	27	39	31	32	88
	Nr Trades	1470	830	483	1470	830	483	NA

Results of the 24 (six-month) combined out-of-sample trading periods between January 01, 2006 and December 31, 2017. A rolling OLS regression is used to estimate the cointegrating relationship and a 2-day lag is imposed on the signal.

TABLE C.4: Out of Sample Results 2006-2017, Kalman Filter & 2-day lag

	Kalman Filter	*Excl. TC*			*Incl. TC*			
		Lookback Z-Score			**Lookback Z-Score**			
Entry	**Measure**	**20**	**60**	**250**	**20**	**60**	**250**	**Benchmark**
2.0	Annualized Returns	0.062	0.064	0.034	0.041	0.052	0.026	0.018
	Annualized SD	0.079	0.078	0.072	0.079	0.078	0.072	0.211
	Maximum Drawdown	0.154	0.141	0.140	0.179	0.150	0.156	0.645
	Average Drawdown	0.012	0.011	0.012	0.012	0.011	0.013	0.036
	Annualized IR	0.784	0.813	0.466	0.518	0.662	0.367	0.085
	Risk-Return-Ratio	0.402	0.452	0.239	0.228	0.345	0.169	0.028
	Avg. DD Length (days)	26	22	40	31	27	51	88
	Nr Trades	653	385	227	653	385	227	NA
1.5	Annualized Return	0.097	0.083	0.046	0.064	0.062	0.033	0.018
	Annualized SD	0.096	0.091	0.082	0.091	0.096	0.082	0.211
	Maximum Drawdown	0.149	0.141	0.134	0.154	0.196	0.161	0.645
	Average Drawdown	0.014	0.011	0.011	0.012	0.015	0.012	0.036
	Annualized IR	1.013	0.908	0.565	0.708	0.648	0.398	0.085
	Risk-Return-Ratio	0.654	0.585	0.344	0.418	0.317	0.202	0.028
	Avg. DD Length (days)	20	18	25	25	28	33	88
	Nr Trades	1073	582	387	582	1073	387	NA
1.0	Annualized Returns	0.108	0.109	0.102	0.057	0.078	0.079	0.018
	Annualized SD	0.104	0.104	0.098	0.103	0.104	0.098	0.211
	Maximum Drawdown	0.136	0.171	0.113	0.223	0.221	0.132	0.645
	Average Drawdown	0.015	0.013	0.012	0.017	0.015	0.015	0.036
	Annualized IR	1.046	1.046	1.039	0.549	0.751	0.811	0.085
	Risk-Return-Ratio	0.796	0.640	0.902	0.254	0.354	0.602	0.028
	Avg. DD Length (days)	20	19	16	29	25	21	88
	Nr Trades	1538	923	667	1538	923	667	NA

Results of the 24 (six-month) combined out-of-sample trading periods between January 01, 2006 and December 31, 2017. A state space regression model is defined and the Kalman filter algorithm is used to estimate the cointegrating relationship. A 2-day lag is imposed on the signal.

TABLE C.5: Out of Sample Monthly Return Distributions 2006-2017, Rolling OLS & 2-day signal lag

Rolling OLS		*Excl. TC* Lookback Z-Score			*Incl. TC* Lookback Z-Score			
Entry	**Measure**	**20**	**60**	**250**	**20**	**60**	**250**	**Benchmark**
2.0	Mean return	0.0056	0.0047	0.0021	0.0040	0.0039	0.0016	0.0026
	SE (NW)	0.0019	0.0013	0.0012	0.0019	0.0013	0.0011	0.0050
	t-stat. (NW)	2.8911	3.5884	1.7942	2.0980	2.9339	1.4200	0.5316
	Minimum	-0.0484	-0.0430	-0.0427	-0.0501	-0.0436	-0.0436	-0.1971
	Quartile 1	-0.0065	-0.0059	-0.0070	-0.0081	-0.0068	-0.0071	-0.0227
	Median	0.0033	0.0031	0.0010	0.0022	0.0018	0.0006	0.0102
	Quartile 3	0.0138	0.0145	0.0110	0.0118	0.0141	0.0096	0.0321
	Maximum	0.0822	0.0675	0.0719	0.0787	0.0664	0.0714	0.1123
	SD	0.0193	0.0176	0.0158	0.0190	0.0173	0.0159	0.0504
	Skewness	0.7649	0.4415	0.7120	0.7824	0.4227	0.7900	-1.0089
	Kurtosis	2.2173	1.1166	2.4990	2.3160	1.1627	2.9336	2.7179
	Sk-Ku Ratio	0.1461	0.1069	0.1290	0.1467	0.1012	0.1327	-0.1758
	Share Pos.	0.6014	0.5857	0.5612	0.5455	0.5532	0.5468	0.5874
1.5	Mean return	0.0064	0.0072	0.0038	0.0037	0.0057	0.0031	0.0026
	SE (NW)	0.0023	0.0020	0.0017	0.0023	0.0019	0.0017	0.0050
	t-stat. (NW)	2.7579	3.6369	2.2142	1.6034	2.9644	1.8222	0.5316
	Minimum	-0.0477	-0.0569	-0.0643	-0.0498	-0.0586	-0.0648	-0.1971
	Quartile 1	-0.0057	-0.0054	-0.0091	-0.0076	-0.0067	-0.0097	-0.0227
	Median	0.0057	0.0053	0.0025	0.0024	0.0038	0.0019	0.0102
	Quartile 3	0.0191	0.0185	0.0144	0.0159	0.0172	0.0128	0.0321
	Maximum	0.1043	0.1586	0.0784	0.0995	0.1551	0.0762	0.1123
	SD	0.0237	0.0250	0.0221	0.0235	0.0247	0.0219	0.0504
	Skewness	0.7588	1.5727	0.4990	0.7712	1.5566	0.4969	-1.0089
	Kurtosis	2.3742	8.8844	1.6990	2.3088	8.8126	1.7307	2.7179
	Sk-Ku Ratio	0.1407	0.1319	0.1058	0.1448	0.1313	0.1047	-0.1758
	Share Pos.	0.6084	0.6364	0.5734	0.5664	0.6154	0.5594	0.5874
1.0	Mean return	0.0071	0.0082	0.0044	0.0033	0.0060	0.0032	0.0026
	SE (NW)	0.0023	0.0023	0.0019	0.0022	0.0022	0.0018	0.0050
	t-stat. (NW)	3.1250	3.6177	2.3234	1.4605	2.7330	1.7095	0.5316
	Minimum	-0.0418	-0.0544	-0.0775	-0.0445	-0.0561	-0.0782	-0.1971
	Quartile 1	-0.0062	-0.0070	-0.0077	-0.0114	-0.0091	-0.0085	-0.0227
	Median	0.0044	0.0063	0.0041	0.0007	0.0035	0.0033	0.0102
	Quartile 3	0.0191	0.0204	0.0173	0.0150	0.0180	0.0164	0.0321
	Maximum	0.1273	0.1983	0.0877	0.1210	0.1920	0.0846	0.1123
	SD	0.0259	0.0295	0.0235	0.0254	0.0289	0.0233	0.0504
	Skewness	1.1402	2.0587	0.0762	1.1911	2.0399	0.0815	-1.0089
	Kurtosis	3.3248	11.3926	1.7270	3.3020	11.2468	1.7608	2.7179
	Sk-Ku Ratio	0.1796	0.1425	0.0161	0.1883	0.1427	0.0171	-0.1758
	Share Pos.	0.6154	0.6294	0.5944	0.5245	0.5874	0.5594	0.5874

Monthly Returns of the 24 (six-month) combined out-of-sample trading periods between January 01, 2006 and December 31, 2017. A rolling OLS regression is used to estimate the cointegrating relationship and a 2-day lag is imposed on the signal. Newey-West Standard Errors are used since return distributions are non-normal.

TABLE C.6: Out of Sample Monthly Return Distributions 2006-2017, Kalman Filter & 2-day signal lag

Kalman Filter		*Excl. TC* **Lookback Z-Score**			*Incl. TC* **Lookback Z-Score**			
Entry	**Measure**	**20**	**60**	**250**	**20**	**60**	**250**	**Benchmark**
2.0	Mean return	0.0054	0.0056	0.0029	0.0037	0.0046	0.0023	0.0026
	SE (NW)	0.0020	0.0015	0.0015	0.0019	0.0015	0.0015	0.0050
	t-stat. (NW)	2.7691	3.7104	1.9443	1.9115	3.0894	1.5824	0.5316
	Minimum	-0.0438	-0.0582	-0.0460	-0.0459	-0.0591	-0.0471	-0.1971
	Quartile 1	-0.0051	-0.0044	-0.0034	-0.0073	-0.0049	-0.0037	-0.0227
	Median	0.0048	0.0024	0.0010	0.0030	0.0017	0.0004	0.0102
	Quartile 3	0.0151	0.0121	0.0073	0.0128	0.0116	0.0069	0.0321
	Maximum	0.0739	0.1533	0.1146	0.0713	0.1500	0.1104	0.1123
	SD	0.0192	0.0213	0.0177	0.0188	0.0211	0.0173	0.0504
	Skewness	0.5398	2.4812	2.1778	0.5501	2.4589	2.0734	-1.0089
	Kurtosis	1.6868	15.6178	12.2694	1.7393	15.4570	11.7237	2.7179
	Sk-Ku Ratio	0.1148	0.1328	0.1421	0.1157	0.1328	0.1403	-0.1758
	Share Pos.	0.6224	0.6357	0.5891	0.5734	0.5929	0.5659	0.5874
1.5	Mean return	0.0085	0.0073	0.0040	0.0057	0.0058	0.0029	0.0026
	SE (NW)	0.0024	0.0019	0.0017	0.0024	0.0019	0.0016	0.0050
	t-stat. (NW)	3.4738	3.8342	2.4201	2.3406	3.0772	1.7859	0.5316
	Minimum	-0.0457	-0.0582	-0.0534	-0.0487	-0.0591	-0.0552	-0.1971
	Quartile 1	-0.0078	-0.0049	-0.0079	-0.0101	-0.0062	-0.0096	-0.0227
	Median	0.0059	0.0043	0.0036	0.0027	0.0035	0.0023	0.0102
	Quartile 3	0.0171	0.0158	0.0145	0.0141	0.0142	0.0135	0.0321
	Maximum	0.1291	0.1940	0.1494	0.1237	0.1899	0.1438	0.1123
	SD	0.0265	0.0255	0.0221	0.0258	0.0251	0.0217	0.0504
	Skewness	1.3178	2.9590	1.9403	1.3733	2.9509	1.8431	-1.0089
	Kurtosis	4.2661	19.0639	12.2006	4.3265	19.2204	11.7072	2.7179
	Sk-Ku Ratio	0.1807	0.1336	0.1272	0.1868	0.1323	0.1249	-0.1758
	Share Pos.	0.6294	0.6014	0.5816	0.5664	0.5804	0.5603	0.5874
1.0	Mean return	0.0094	0.0096	0.0087	0.0053	0.0072	0.0069	0.0026
	SE (NW)	0.0023	0.0026	0.0021	0.0023	0.0026	0.0021	0.0050
	t-stat. (NW)	4.1124	3.6547	4.1516	2.3201	2.7383	3.3539	0.5316
	Minimum	-0.0474	-0.0772	-0.0580	-0.0497	-0.0815	-0.0588	-0.1971
	Quartile 1	-0.0058	-0.0077	-0.0073	-0.0099	-0.0100	-0.0088	-0.0227
	Median	0.0057	0.0052	0.0075	0.0017	0.0028	0.0051	0.0102
	Quartile 3	0.0226	0.0215	0.0208	0.0186	0.0188	0.0189	0.0321
	Maximum	0.1265	0.2260	0.1628	0.1204	0.2199	0.1547	0.1123
	SD	0.0270	0.0334	0.0269	0.0265	0.0328	0.0264	0.0504
	Skewness	0.9073	2.4973	1.3637	0.9522	2.4884	1.2825	-1.0089
	Kurtosis	2.4535	13.3171	6.4591	2.4847	13.4771	5.8557	2.7179
	Sk-Ku Ratio	0.1658	0.1525	0.1437	0.1730	0.1505	0.1443	-0.1758
	Share Pos.	0.6573	0.5874	0.6014	0.5524	0.5455	0.5734	0.5874

Monthly Returns of the 24 (six-month) combined out-of-sample trading periods between January 01, 2006 and December 31, 2017. A state space regression model is defined and the Kalman filter algorithm is used to estimate the cointegrating relationship. A 2-day lag is imposed on the signal and Newey-West Standard Errors are used since return distributions are non-normal.

TABLE C.7: OOS Monthly Historical VaR and ES estimates 2006-2017, Rolling OLS & a 2-day signal lag

Kalman Filter		*Excl. TC* **Lookback Z-Score**			*Incl. TC* **Lookback Z-Score**			
Entry	**Measure**	**20**	**60**	**250**	**20**	**60**	**250**	**Benchmark**
2.0	Hist. 95% VaR	-0.0199	-0.0207	-0.0210	-0.0207	-0.0215	-0.0210	-0.0734
	Hist. 95% ES	-0.0306	-0.0307	-0.0269	-0.0313	-0.0315	-0.0273	-0.1323
	Hist. 99% VaR	-0.0383	-0.0353	-0.0287	-0.0395	-0.0364	-0.0292	-0.1726
	Hist. 99% ES	-0.0446	-0.0398	-0.0358	-0.0460	-0.0408	-0.0366	-0.1951
1.5	Hist. 95% VaR	-0.0317	-0.0325	-0.0280	-0.0343	-0.0331	-0.0281	-0.0734
	Hist. 95% ES	-0.0402	-0.0422	-0.0400	-0.0424	-0.0432	-0.0405	-0.1323
	Hist. 99% VaR	-0.0432	-0.0465	-0.0466	-0.0447	-0.0475	-0.0474	-0.1726
	Hist. 99% ES	-0.0458	-0.0530	-0.0567	-0.0475	-0.0542	-0.0574	-0.1951
1.0	Hist. 95% VaR	-0.0302	-0.0315	-0.0330	-0.0327	-0.0322	-0.0344	-0.0734
	Hist. 95% ES	-0.0361	-0.0447	-0.0464	-0.0389	-0.0460	-0.0474	-0.1323
	Hist. 99% VaR	-0.0390	-0.0510	-0.0509	-0.0418	-0.0527	-0.0520	-0.1726
	Hist. 99% ES	-0.0406	-0.0535	-0.0649	-0.0434	-0.0554	-0.0659	-0.1951

Monthly historical VaR and ES estimates of the 24 (six-month) combined out-of-sample trading periods between January 01, 2006 and December 31, 2017. A rolling OLS regression is used to estimate the cointegrating relationship and a 2-day lag is imposed on the signal.

TABLE C.8: OOS Monthly Historical VaR and ES estimates 2006-2017, Kalman Filter & a 2-day signal lag

Kalman Filter		*Excl. TC* **Lookback Z-Score**			*Incl. TC* **Lookback Z-Score**			
Entry	**Measure**	**20**	**60**	**250**	**20**	**60**	**250**	**Benchmark**
2.0	Hist. 95% VaR	-0.0236	-0.0181	-0.0182	-0.0246	-0.0188	-0.0192	-0.0734
	Hist. 95% ES	-0.0344	-0.0316	-0.0305	-0.0350	-0.0323	-0.0311	-0.1323
	Hist. 99% VaR	-0.0389	-0.0346	-0.0375	-0.0398	-0.0349	-0.0378	-0.1726
	Hist. 99% ES	-0.0418	-0.0467	-0.0429	-0.0433	-0.0472	-0.0438	-0.1951
1.5	Hist. 95% VaR	-0.0249	-0.0209	-0.0254	-0.0266	-0.0215	-0.0258	-0.0734
	Hist. 95% ES	-0.0391	-0.0330	-0.0372	-0.0389	-0.0341	-0.0379	-0.1323
	Hist. 99% VaR	-0.0435	-0.0364	-0.0472	-0.0453	-0.0381	-0.0481	-0.1726
	Hist. 99% ES	-0.0448	-0.0491	-0.0530	-0.0473	-0.0499	-0.0543	-0.1951
1.0	Hist. 95% VaR	-0.0299	-0.0282	-0.0248	-0.0324	-0.0300	-0.0256	-0.0734
	Hist. 95% ES	-0.0386	-0.0432	-0.0393	-0.0416	-0.0451	-0.0406	-0.1323
	Hist. 99% VaR	-0.0451	-0.0535	-0.0503	-0.0480	-0.0547	-0.0522	-0.1726
	Hist. 99% ES	-0.0471	-0.0682	-0.0557	-0.0491	-0.0709	-0.0573	-0.1951

Monthly historical VaR and ES estimates of the 24 (six-month) combined out-of-sample trading periods between January 01, 2006 and December 31, 2017. A state space regression model is defined and the Kalman filter algorithm is used to estimate the cointegrating relationship. A 2-day lag is imposed on the signal.