Quantitative Finance

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rquf20

Pairs trading with wavelet transform

Burak Alparslan Eroğlu, Haluk Yener & Taner Yiğit

To cite this article: Burak Alparslan Eroğlu, Haluk Yener & Taner Yiğit (2023) Pairs trading with wavelet transform, Quantitative Finance, 23:7-8, 1129-1154, DOI: 10.1080/14697688.2023.2230249

To link to this article: https://doi.org/10.1080/14697688.2023.2230249

View supplementary material

Published online: 10 Jul 2023.

Submit your article to this journal

Article views: 595

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rquf20

Quantitative Finance, 2023
Vol. 23, Nos. 7–8, 1129–1154, https://doi.org/10.1080/14697688.2023.2230249

Routledge
Taylor & Francis Group
Check for updates

Pairs trading with wavelet transform

BURAK ALPARSLAN EROĞLU †, HALUK YENER *‡ and TANER YIĞIT§

†Department of Economics, Bakırçay University, Menemen, Izmir 35660, Turkey
‡Department of Business Administration, Istanbul Bilgi University, Eyup, Istanbul 34060, Turkey
§Department of Economics, Bilkent University, Bilkent, Ankara 06533, Turkey

(*Received 25 January 2022; accepted 22 June 2023; published online 10 July 2023*)

We show that applying the wavelet transform to S&P 500 constituents' prices generates a substantial increase in the returns of the pairs-trading strategy. Pairs trading strategy is based on finding prices that move together, but if there is shared noise in the asset prices, the co-movement, on which one base the trades, might be caused by this common noise. We show that wavelet transform filters away the noise, leading to more profitable trades. The most notable change occurs in the parameter estimation stage, which forms the weights of the assets in the pairs portfolio. Without filtering, the parameters estimated in the training period lose relevance in the trading period. However, when prices are filtered from common noise, the parameters maintain relevance much longer and result in more profitable trades. Particularly, we show that more precise parameter estimation is reflected on a more stationary and conservative spread, meaning more mean reversion in opened pairs trades. We also show that wavelet filtering the prices reduces the downside risk of the trades considerably.

Keywords: Pairs trading; Wavelet transform; Minimum distance method; Cointegration method; Statistical arbitrage

JEL Classification: C1, C22, C53, G11, G12, G14

1. Introduction

Pairs trading involves taking opposing positions in two risky securities which are found to be related, usually based on statistical assessments. If an assessment suggests a relation, an equilibrium level between two securities is assumed to exist. Whenever the securities' prices diverge sufficiently from the equilibrium, an investor aims to benefit from this divergence with the hope that it is transitory. To this end, the investor takes a long position in the cheap risky security and shorts the expensive one. Once the diverged securities' prices start to return toward the equilibrium level, the strategy generates profit. Clearly, foreseeing the potential of profits is difficult to achieve because the noisy evolution of financial assets is detrimental to their predictability. Many studies show that pairs trading has started out as a profitable strategy, yet the earnings of the existing methodologies have lately waned. In this study, we show that this declining performance is due to noisy data and filtering the price data with a wavelet transform improves earnings substantially. More importantly, we examine the filtered-out noise and investigate why these benefits occur.

We consider two commonly utilized pairs selection methods from the literature as benchmarks, namely the minimum distance method (Gatev *et al.* 1999, 2006), and the cointegration method (Vidyamurthy 2004). These two methods capture a statistical relation between two risky financial assets based on past historical data. The underlying idea of the minimum distance method is to capture the statistical relation between the stocks based on 'the law of one price' of two assets with similar pay-off profile (Ingersoll 1987). In the cointegration method, first, the well-known cointegration tests (Engle and Granger 1987, Johansen 1988) are applied to evaluate the existence of a long-run relationship between two risky securities. If cointegration exists, an investor can take long-short positions in a diverged pair accordingly to generate profits once pairs converge at the long-run equilibrium. Our underlying motivation for using the most common pairing methodologies and their simplest forms is to focus on the benefits of filtering pairs trading even in a bare bones environment.

In the pairs literature, the pioneering work of Gatev *et al.* (2006) shows that abnormal profits to pairs trading indeed exist. The study applies the minimum distance method to a universe of CRSP stocks and documents an average annual return of around 11% before transaction costs for the period between 1962 and 2002. Provided the level of returns

*Corresponding author. Email: haluk.yener@bilgi.edu.tr

© 2023 Informa UK Limited, trading as Taylor & Francis Group

in an earlier study in 1999 (Gatev *et al.* 1999), the successful outcome of pairs trades in Gatev *et al.* (2006) cannot be the result of data snooping, and the returns have low exposure to the systemic factors of a Fama-French type model. However, as extensions of Gatev *et al.* (2006), later studies by Do and Faff (2010, 2012) show that the returns to pairs are on a declining path and are more modest than previously thought after the inclusion of transaction costs. Thus, if abnormal returns exist, they may not be robust to transaction costs, or at best, offer low-risk and low-return opportunities to institutional investors (Do and Faff 2012).

In a similar vein, there are some notable studies, related to cointegration, providing support to declining returns in pairs trading of stocks in the US financial markets. One such study is Clegg and Krauss (2018), which uses the partial cointegration method over S&P 500 for the period between 1990 and 2015. The study documents an annual return of around 12% after the transaction costs by also confirming the declining trend in profits, especially the negative level of returns during the period between 2010 and 2015. As a more comprehensive study, Rad *et al.* (2016), considering the CRSP dataset extending between 1962 and 2014, compares the above-mentioned selection methods along with a copula method. This study also confirms the declining return levels after 2009 for the minimum distance and cointegration methods. However, the returns in the copula method remain stable after the same year but at a much lower level, albeit significant as others. Similarly, Stübinger *et al.* (2018), employing vine copulas on S&P 500 stocks, shows stable annual returns of 9.25% between the years 1992 and 2015. Moreover, the study shows that the returns are not subject to the negative results observed in the recent part of the dataset utilized in the above-mentioned studies.

Thus, the results regarding the decline of the abnormal returns in the US financial markets are variable, especially post-2010. It seems the level of returns crucially depends on a model's ability to extract information from the noisy stock returns. That is our motivation for using (wavelet) filters to separate out the 'relevant' information. Specifically, the existence of market-wide noise will pollute the picture when one tries to find co-moving pairs of securities as the common noise might be mistaken for common movement. Filtering out the common noise helps us find much better pairs that lead to more successful trades. In the finance literature, wavelets are frequently utilized for noise reduction because the presence of noise can cause estimation-induced biases due to irregularities and roughness of the data (Sun and Meinl 2012). Furthermore, another issue, which matters in financial trading, is the erratic structure of the financial series. For instance, Uddin *et al.* (2019) points out that the behavior of oil series can include drifts, spikes, and other non-stationary effects that common frequency-domain decomposition methods, such as Fourier transform, are unable to reveal or miss out completely. Nonetheless, wavelets are quite helpful in alleviating such problems. On the technical side, wavelets can distinguish between short-run singular events and common long-run behavior of the financial data (Conlon *et al.* 2018). Accordingly, wavelets can isolate long-run co-movement in the presence of financial and economic frictions. To sum up, filtering the empirical properties of the financial data with the mentioned technical features of wavelets contributes towards profitable pairs trading.

Our results show that the declining performance of the above methodologies is mainly caused by the poor fit of the pairing mechanism between the training and trading periods. In other words, the pairs and the coefficients chosen for the spreads in the training period end up performing poorly in the trading period. We show that the noise in the price series mainly causes poor performance †, and the application of the wavelet filter improves the earnings substantially. To accurately identify the source of the benefits, we first filter the price series before pairing, observing an improvement in the quality of pairing, hence the returns. We then examine the impact of filtering the prices after using the existing methodologies for pairing, only to see even a more considerable improvement in returns. The main contribution of our research starts here, which is deciphering the reasons for the improvement caused by filtering. We find that the significant change is in the coefficient estimation stage. De-noising the price series leads to (i) the removal of common noise to better selection of pairs, (ii) a more accurate estimation of the parameters (due to the removal of measurement error) and a better proportion of the long/short assets in pairs trade, (iii) a more stationary spread, hence a better success ratio between the training and trading periods and a larger proportion of successfully closed trades, and (iv) a more *conservative* spread, leading to lower losses in the force-closed trades at the end of the trading period (also evidenced by higher Sharpe ratios). Although the benefit of filtering occurs both in the pairing (pairs selection stage) *and* trading (after pairing) stages, we briefly mention the benefits in the pairing stage in an online appendix and focus our analysis on the latter coefficient estimation part as the major part of improvement happens in that stage. In this way, we also keep the analysis more focused on identifying the source of benefit. We elaborate on these benefits with a variety of exercises.

Our empirical results are based on seven different yearly out-of-sample periods. Particularly, we undertake our study between the years 2010 and 2018 for the constituents of S&P 500. We consider a relatively short period since most studies already cover the period up to 2015. We do not select data prior to the year 2010 to prevent the effect of the 2008/09 global financial crisis because, as Do and Faff (2010) shows, these periods correspond to times when pairs yield 'remarkable returns'. When selecting the pairs, we do not restrict them to the same sector as in some notable studies such as Do and Faff (2010, 2012) and Clegg and Krauss (2018). Instead, we select pairs mechanistically, implying that stocks from two different sectors may also end up as a pair. Our underlying motivation behind this selection criterion is the co-movement of the constituents of S&P 500 as documented in Vijh (1994) and Barberis *et al.* (2005). For example, Cohen and Frazzini (2008) and Menzly and Ozbas (2010) show the effect of economic linkages in cross-predicting the returns of the firms. Thus, it would be natural to think that the sectoral relations would also hold within the constituents of S&P 500. In addition, since wavelet filtering the data gets rid of the common

† We find a negative correlation between the noise we separate and the returns to standard methods of unfiltered pairing.

(market or sectoral) noise we are more unrestricted in finding the co-moving pairs.

Our results without the application of the wavelet transformation confirm the continuation of persistent negative returns observed in Rad *et al.* (2016) after 2009 and Clegg and Krauss (2018) for 2010–2015. However, with the application of the wavelet transformation, even after pairing with the standard techniques, the before (after) transaction cost average return levels increase noticeably from -1.81% (-1.88%) to 9.66% (9.01%) under the cointegration method, and from -0.55% (-1.76%) to 11.82% (10.98%) under the minimum distance method. These returns indicate abnormal profits based on four asset pricing models, which are from Petkova (2006), Fama and French (2015) and Hou *et al.* (2015) and the abnormal returns are robust to transaction costs. The models introduced by Petkova (2006) and Hou *et al.* (2015) are alternatives that are shown to perform better than the three factor model of Fama and French (1993). To the best of our knowledge, our study is the first one to provide a comparison of these models.

While the returns in our study compare well with that in Stübinger *et al.* (2018), they also provide support to those in a recent study by Flori and Regoli (2021) applying a deep learning approach, also known as Long Short-Term Memory (LSTM) networks, to pairs trading. By referring to reversal effect and depending on the employed method, that study reports sample return levels of around 2.5% to 14.30% between the years 2011 and 2018 with average Sharpe ratio values in between 0.25–1.34 (per our calculations from table 4, p.11 in Flori and Regoli 2021). In our case, we report slightly lower return levels with average Sharpe ratio values more than twice the highest value of 1.34. Given the comparison, although our wavelet-filtering approach may look like an alternative to studies that employ copula, machine learning or neural network methods, it is rather complementary. Even with the most basic selection methods, such as cointegration and minimum distance, it leads to substantial improvements. Our goal here is to introduce the benefits of filtering the noise with the wavelet transform, leading to the contention that its application with more complex approaches such as copula methods or deep learning might lead to even higher profits.

The outline of the paper is as follows: We first provide the description of the selection methods employed in this study in Section 2. We describe the wavelet transform in Section 3. In Section 4, we explain the details of the data and the trading rules. We display and interpret the results and show their robustness in Section 5 . We conclude the study with Section 6 by providing a summary of results. In the online appendix, we provide details on the cointegration test employed in the study along with further technicalities on wavelet transforms as well as details on some key results and our simulation study.

2. Methods

2.1. *The minimum distance method*

One of the frequently applied pair selection techniques in the literature is the distance method. The underlying intuition behind this method, as its name suggests, is to select pairs based on the distance between two risky securities. In this study, we name the risky securities stocks. If the paired stocks based on the distance method diverge from each other, a pairs trader could short the expensive stock and long the cheap one to make a profit with the expectation that the divergence is temporary within a perceived horizon.

For the application of this method, we consider average squared deviation of the normalized prices of each stock as in Gatev *et al.* (2006). The normalization involves dividing the price series of a stock in the sample with the initial price of the same stock. More clearly, let $S_{i,t}$ and $S_{j,t}$ be the time t prices of two distinct stocks i and j respectively. Then, the normalization procedure leads to the division $\tilde{S}_{i,t} = S_{i,t}/S_{i,t_0}$ and $\tilde{S}_{j,t} = S_{j,t}/S_{j,t_0}$, where $t \in \{t_0, \dots, t_0 + T\}$ with t_0 being the initial time and T being the maturity time for the period of application. As we will explain in Subsection 4.2 with more detail, the period of application is our *training* (in-sample) period, which we utilize for determining the pairs in the *trading* (out-of-sample) period. For the selection of pairs, we construct the mean squared distance between two stocks as:

$$D_{i,j} = \frac{1}{T}\sum_{t=1}^{T}(\tilde{S}_{i,t} - \tilde{S}_{j,t})^2. \tag{1}$$

Assuming N stocks, we compute equation (1) for each $i, j \in \{1, \dots, N\}$ with $i \neq j$, and sort these values in ascending order to build up a pairs list. In our analysis, we consider 1000 pairs from the top of the list. The results for smaller quantities of pairs are similar and available upon request.

2.2. *The cointegration method*

Another frequently applied pair selection technique in pairs trading is the cointegration method (Vidyamurthy 2004), which is an econometric assessment of the long-run equilibrium relationship between two stocks. Under this method, the relationship between two stocks $S_{i,t}$ and $S_{j,t}$ is considered to be of the form,

$$S_{i,t} = \alpha + \beta S_{j,t} + \epsilon_t, \tag{2}$$

where $t \in \{t_0, \dots, T + t_0\}$ is the time within the above-mentioned training period and the values of the unknown parameters α and β are determined based on a cointegration regression. For the cointegration test, the stock price series should be integrated of order one, denoted by $I(1)$. Then, if two integrated price series are cointegrated, the error term ϵ_t is stationary, which is also called $I(0)$ process (Engle and Granger 1987). In other words, two price series are cointegrated if they evolve together in equilibrium. The error term ϵ_t is the spread, because we can write from equation (2) the cointegrating relation $S_{i,t} - \beta S_{j,t} = \alpha + \epsilon_t$. Here, α is the long-run mean of this cointegration relation, and consequently, the spread is a mean-reverting process. As a result, any temporary deviation from the long-run mean, α, might yield profitable trading strategies. In this case, similar to what we explained under the distance method, an investor aims to generate profits from these temporary deviations by longing the cheaper stock and shorting the expensive one. Once these price series converge again around the long-run mean, the investor makes a profit. For this study, we consider all possible cointegrated pairs among the constituents of S&P 500. The details on the number of pairs selected are reported in Subsection 5.1.

From the above explanation, we may see that our first goal is to find the appropriate pairs to be traded. Such a result is not guaranteed as the stocks might be spuriously related. To address this problem, we apply the widely used Johansen test on the price series of paired stocks (Johansen 1988).

Finally, in the selection of the pairs with the cointegration method, we utilize the Johansen's (1995) trace test. First, we choose the optimal lag length for the given series by using Schwarz Information Criterion. Next, we conduct the cointegration tests, which rely on the p-values generated from the asymptotic distribution. These values can be found in MacKinnon (1996). Moreover, we allow the intercept in the cointegrating relations, but there is no deterministic trend in the levels of the data. All tests are conducted at 5% significance level.

3. Wavelet transform

In time series analysis, wavelet transform is a time-frequency domain filtering method that decomposes time series into low-frequency (long-run) and high-frequency (short-run) components by using filters called wavelets. These filters are mathematical functions that oscillate in a finite time and eventually fade away. Such decomposition of economic/financial data is essential since the long-run component of a time series may be more helpful in revealing the underlying economic relationships between variables than does the short-term component (Conlon *et al.* 2018). Our goal with the wavelet application is to capture the time-series properties of the price series under different frequency domains. That is, we divide the evolution of a price process in between noise and trend structure to establish a profitable trading strategy by relying on the trend. In this section, we briefly outline the wavelet techniques utilized in the analysis. We undertake a simpler exposition than that in Gençay *et al.* (2001), Fan and Gençay (2006), and Conlon *et al.* (2018) to ease understanding. Furthermore, we provide an example regarding the application of wavelet transforms to pairs trading in Subsection 4.4.

3.1. The maximum overlap discrete wavelet transform

Originally, wavelet functions (filters) are designed as continuous functions to analyze time series sampled in continuous intervals. However, the discrete wavelet decomposition provides a better framework for analyzing economic/financial data sampled in discrete intervals. In addition, as Gençay *et al.* (2001) mentions, some of the fundamental properties of continuous wavelet transforms are shared by its discrete version. In this study, we consider one of the discrete wavelet decompositions, namely the Maximum Overlap Discrete Wavelet Transform.

In the wavelet literature, the most frequently used discrete wavelet decompositions are the Discrete Wavelet Transform (DWT) and Maximum Overlap Discrete Wavelet Transform (MODWT). The DWT is not suitable for our analysis because it suffers from a basic problem: In every application of the wavelet filters, the sample size is halved. This feature is known as down-sampling and leads to fewer and fewer observations in higher 'levels' of the transformation. However, down-sampling is not a desirable property when we design a real-time pairs trading rule with wavelets, since we need the sample size of the trading signal series to be equal to the original sample size. In order to avoid the sample size loss, we utilize the MODWT, which shares many properties of the DWT, but is free from the down-sampling problem.

Like DWT, MODWT requires that the analyzed series should have a dyadic length, implying that the sample size should be of the form $T = 2^U$, where U is some positive integer, and T stands for the sample size. For financial (or economic) time series, this requirement, which leads to the boundary/border effect, is often too restrictive, as the sample size may not be of the aforementioned form. That is, there may not be enough sample size to realize an appropriate decomposition. However, as a remedy, some notable signal extension techniques such as zero padding, symmetric padding (symmetrization), smooth padding, periodic padding techniques are applied before the use of MODWT method (Strang and Nguyen 1996). In our analysis, we utilize the symmetrization technique.

After symmetrization, MODWT can be applied to the extended series straightforwardly. The essence of the application lies in undertaking two types of filters which are namely high- and low-pass filters. The high-pass filter is for modelling the short-run fluctuations, whereas the low-pass filter is for the long-run fluctuations. In terms of mathematical formulation, let $\tilde{h} = (\tilde{h}_0, \tilde{h}_1, \ldots, \tilde{h}_{L-1})$ be the high-pass filter, and $\tilde{g} = (\tilde{g}_0, \tilde{g}_1, \ldots, \tilde{g}_{L-1})$ be the low pass filter, each with filter length L. In addition, the low-pass filter coefficients can be considered as weight functions that smooth the original process. To denote the decomposition of a time series of the form $(Z_t)_{t\in\{1,\ldots,T\}}$ (with dyadic length $T = 2^U$), we write

$$\tilde{V}_{1,t} = \sum_{l=0}^{L-1} \tilde{g}_l Z_{t-l \bmod T}; \quad \tilde{W}_{1,t} = \sum_{l=0}^{L-1} \tilde{h}_l Z_{t-l \bmod T}, \tag{3}$$

where $\tilde{V}_{1,t}$ and $\tilde{W}_{1,t}$ are the long-run and short-run transformations of $(Z_t)_{t\in\{1,\ldots,T\}}$, respectively. That is, $\tilde{V}_{1,t}$ is the *Level 1* long-run component, and $\tilde{W}_{1,t}$ is the *Level 1* short-run component/noise. The value of a level determines the level of detail the filtering algorithm reveals by zooming into a process, hence its resolution. The algorithm is named as the 'Pyramid Algorithm' after Mallat (1989), and we present a depiction of it in the figure that follows.

In figure 1, we see that going up in the level (which is moving downwards in the figure) implies further resolution of the low-frequency (long-run) component of the times series. As we see, the level of filtration continues up to level J^*, whose value can be determined based on built-in functions in many statistical software packages and programing languages. In this study, due to computational power concerns, we consider Level 1 approximation. Even at this basic level, we indeed obtain strong enough results.

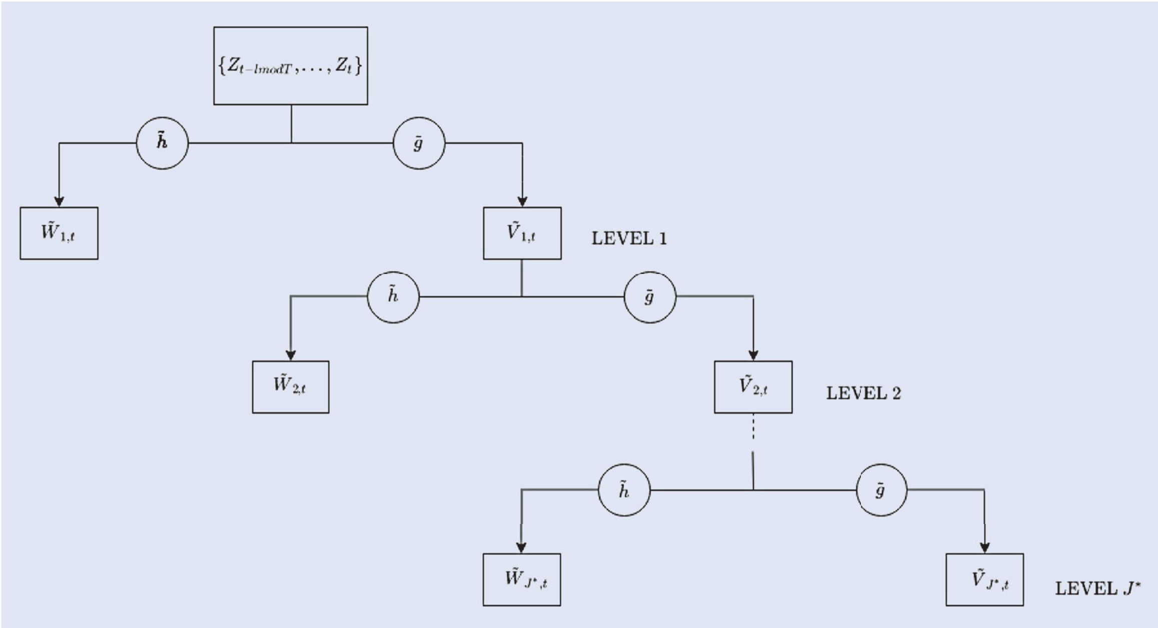

Figure 1. Pyramid algorithm of Mallat.

Two crucial aspects of the wavelet analysis are the choice of the filters $\tilde{h}_l, \tilde{g}_l$ for $l = 0, \ldots, L-1$, and the filter length L. There are numerous different filter types in the literature. All of these families are expressed with a term that corresponds to the number of vanishing moments in the wavelet filters. The value of the number of vanishing moments is equal to half of the filter length, that is $L/2$. With this value, the degree of polynomial fitted to extract information from time-series data is determined. In other words, the lag length is determined.

4. Data and trading rules

4.1. The data

The data consist of the constituents of S&P 500 for the period that extends between March 5th, 2010 and March 15th, 2018. We divide this time period into eight equal financial years. Each year contains 252 days, excluding holidays and non-business days. The starting period is selected as the year 2010 to avoid the effect of the recent global financial crisis on the dataset. A detailed demonstration of the data split can be found in table 1, where the date structure follows month/day/year.

We obtain the data from the Bloomberg terminal. It is composed of the stock prices that are adjusted for dividends and splits. We consider 415 stocks because only these stocks were traded in two consecutive years, which is required for appropriate pairs trading. We do not necessarily select stocks that were consistently present in the index for the whole observation period. Some basic statistics regarding the returns obtained from the constituents of our data are provided in table 2.

We select the constituents of S&P 500 because the index consists of liquid universe of stocks whose total market value captures almost 80% of the US stock market capitalization. It has extensive market analyst coverage in most financial news and is also utilized as a proxy for the market portfolio in numerous empirical asset pricing studies.

4.2. Formation and trading periods

For the pairs selection, we employ the cointegration and minimum distance methods under the training period in $t = 0, \ldots, T$, also known as the formation period. We consider all possible pairs in our dataset. The pairs are then traded by using the out-of-sample data for $t = T+1, \ldots, 2T$, and this period is called the trading period. We let both the training and trading periods consist of 252 daily observations, as 252 is the number of financial days in a calendar year; thus, we set $T = 252$. In the trading exercises, the opening and closing positions in pairs are decided based on a signal from the spread between the wavelet transformed stock price series. We also consider the standard application where price series are not wavelet transformed for comparison.

We may see in table 1 of Subsection 4.1 that we have seven formation and trading periods. With this reporting format, our aim is to observe the level of variability in our results in different periods. Moreover, while our selection of the formation period is standard, the trading period is different, as it is longer than six months selection encountered in the literature (Gatev *et al.* 2006, Rad *et al.* 2016). We keep the trading period long with the expectation of increasing the uncertainty in our analysis. In this way, we hope to add more robustness to our findings. Keeping the trading period long works against the pairing done in the previous (training period) year and emphasizes the benefit of the wavelet transform. For further robustness, in Subsection 5.5, we provide average return results for shorter trading periods consisting of three, six, and nine months of trading.

Regardless of the type of pairs selection method considered in the study, we use the same spread specification for trade initiation under all methods. For example, under the cointegration method, we estimate the cointegrating regression with

Table 1. Description of the data sample.

	Period	1	2	3	4	5	6	7
Training	Start	03/05/2010	03/04/2011	03/06/2012	03/11/2013	03/11/2014	03/13/2015	03/16/2016
	End	03/03/2011	03/05/2012	03/08/2013	03/10/2014	03/12/2015	03/15/2016	03/15/2017
Trading	Start	03/04/2011	03/06/2012	03/11/2013	03/11/2014	03/13/2015	03/16/2016	03/16/2017
	End	03/05/2012	03/08/2013	03/10/2014	03/12/2015	03/15/2016	03/15/2017	03/15/2018

Table 2. Basic statistics for the return series obtained from the S&P 500 stocks.

Industry	# of Firms	Mean	Stdev.	Skewness	Kurtosis
Communication Services	15	19.25%	16.90%	– 0.4116	5.9928
Consumer Discretionary	51	17.71%	17.97%	– 0.3172	6.2322
Consumer Staples	26	13.16%	12.24%	– 0.3128	5.3612
Energy	24	8.61%	25.87%	– 0.1321	5.5094
Financials	60	14.91%	20.59%	– 0.3279	7.9465
Health Care	55	20.01%	16.23%	– 0.4824	6.0463
Industrials	57	16.97%	18.05%	– 0.3298	7.1194
Information Technology	48	19.70%	17.87%	– 0.2552	6.1649
Materials	23	14.24%	19.42%	– 0.3281	5.8789
Real Estate	32	12.59%	18.15%	– 0.0857	8.8627
Utilities	24	9.37%	14.60%	– 0.3464	5.4850
TOTAL	415	15.91%	16.00%	– 0.3924	8.0039

Note: Std. dev. stands for standard deviation.

the training data and obtain the coefficient estimates $\hat{\alpha}$ and $\hat{\beta}$. Then, we generate the standard spread as $\hat{\epsilon}_{s,t} = S_{i,t} - \hat{\alpha} - \hat{\beta} S_{j,t}$ for $t = T+1, \ldots, 2T$. Note that in the formation of the spread we use the original price series; we do not consider the normalized price series mentioned under the description of the minimum distance method in Subsection 2.1.

For the computation of wavelet transformed spread, we apply the wavelet filter to the price series first, then we compute the spread. We note that the aforementioned operation is not equivalent to the application of wavelet transformation to the spread directly. To this end, we let $\tilde{V}_{i,1,t}$ and $\tilde{V}_{j,1,t}$ be the Level 1 MODWT long-run component coefficients of $S_{i,t}$ and $S_{j,t}$, respectively. Then, we estimate the (cointegrating) relation between $\tilde{V}_{i,1,t}$ and $\tilde{V}_{j,1,t}$ and obtain the parameter estimates $\hat{\alpha}_w$ and $\hat{\beta}_w$. Similarly, we derive the spread as $\hat{\epsilon}_{w,t} = \tilde{V}_{i,1,t} - \hat{\alpha}_w - \hat{\beta}_w \tilde{V}_{j,1,t}$ for $t = T+1, \ldots, 2T$. We will later show that the differences in coefficient estimates play a key role in improving returns.

Under the minimum distance method, we express the spread in the same manner to keep the similar specification across the two methods. Consequently, the parameter estimates $\hat{\alpha}_w$ and $\hat{\beta}_w$ are due to simple regression of the price series of two stocks rather than their cointegrating relation. Such consideration may raise the issue of spurious relation among the stocks because we are regressing two non-stationary series against each other. However, we conduct our trading exercises by considering a large number of trials. That is, the successful results under the wavelet spread may not be by chance due to spurious relations as such must have been manifested under the standard spread as well. In fact, we will show that the regression coefficients are essential to the results under the wavelet spread, and we provide a brief discussion so as to why they are so in Subsections 5.2 and 5.3.

4.3. *Threshold selection & trading strategy*

In the pairs trading literature, the trading signal is initiated after the estimated spread exceeds some predetermined threshold. In this context, the threshold we use relies on the standard deviation of the spread in the training period. We initiate the trading if the absolute value of the distance or spread exceeds two historical standard deviation. This is similar to the trading rule employed in Gatev *et al.* (2006). We say similar because we compute the value of σ from the training periods and use this (constant) value for the trading period. In this way, we use a similar threshold rule as in Bertram (2010) and Endres and Stübinger (2019). The difference is that those studies endogenously determine the thresholds based on an objective function and a priori mathematical assumption on the evolution of the spread. In our case, we do not follow the aforementioned analytical approach to determine the value of the thresholds; they are straightforwardly computed by using variability of the in-sample values of the spread.

To initiate the trade, for a given level of symmetric lower and upper threshold values $\ell < 0 < u = -\ell = 2\sigma$ respectively, we short one dollar of stock $S_{i,\cdot}$ and long β dollars of stock $S_{j,\cdot}$ whenever $\hat{\epsilon}_{\cdot,\cdot} > u$. In the instance when $\hat{\epsilon}_{\cdot,\cdot} < \ell$, we then short β dollars of stock $S_{j,\cdot}$ and long one dollar of $S_{i,\cdot}$. Regarding the latter case, the general approach in the literature involves longing one dollar of $S_{j,\cdot}$ and shorting $1/\beta$ dollars of $S_{i,\cdot}$ when $\hat{\epsilon}_{\cdot,\cdot} < \ell$ (Rad *et al.* 2016). We consider our slightly modified approach to make our return calculations shown in Appendix A.1 easier. Moreover, our long/short rule also differs from an alternative approach (Endres and Stübinger 2019) that involves financing one dollar long position in the cheap stock with one dollar short position in the expensive one. We report the results for this (latter) type of trading in the remark that follows.

We do not initiate any pairs trade in the region $\ell < \hat{\epsilon}_{\cdot,\cdot} < u$. Furthermore, we let the trade positions be closed whenever the spread is zero or changes its sign the first time after the trade is initiated. In the analysis, we also provide the mean returns computed from values obtained one day after the opening and closing signal. In the real data, the spread value is almost never zero; thus, we look for the first sign change in the spread to close the trade position. Finally, at the end of the trading period, any open positions are automatically (force-) closed, and their corresponding returns are included in our computations.

REMARK 4.1 We considered alternative trading strategies under similar wavelet functions. Some of the results are not promising and therefore omitted in the sequel. We briefly mention these trading strategies to show the effect of appropriate 'spread specification' in generating profits from pairs trading. To this end, we write:

- We set $(\alpha_w, \beta_w) = (0, 1)$ to obtain a 'spread' by shorting one dollar of the overvalued stock and longing one dollar of the undervalued one. This selection also relates to the case of using *standardized* price series as explained in the description of the minimum distance method. The results are quite close to the results under the standard spread (we see in the text) and are thus omitted due to their weakness;
- Regression with returns instead of price series also yield weak results and are not applied for the cointegration method to be viable;
- We also consider the original spread series $\hat{\epsilon}_{w,t} = \tilde{V}_{i,1,t} - \hat{\alpha}_w - \hat{\beta}_w \tilde{V}_{j,1,t}$ for trade signals, while generating trades by shorting one dollar of the overvalued stock and longing one dollar of the undervalued stock (as mentioned above). In this case, the results are quite successful, yet lower than what we observe here. For the cointegration method, we obtained returns that are on average lower by around 0.62% and for the minimum distance 2.7%. Despite the decline, the essence of our findings still remains the same. The only difference is that the average yearly returns under longing and shorting a dollar amount strategy retracts to levels around 9-to-10%, which are in line with the one reported in Stübinger *et al.* (2018).

4.4. An example of a pair trade

We continue with showing an application of the concepts introduced in Subsections 4.2–4.3 and Appendix A.1. In addition, we provide a note on the practical application of our algorithm in Appendix A.2. To this end, we consider the data of Oneok and Estee Lauder stocks. In figure 2, we present the evolution of the spreads and the corresponding threshold values computed both under the standard approach (dashed) and *sym*22 wavelet filter approach (solid). The values are depicted together for a clearer comparison. The evolution of the spreads is computed by using the price data from the first *trading* period that extends (as previously mentioned) between March 4th, 2011 and March 5th, 2012. However, in the computation of the spreads, the estimates $\hat{\alpha}, \hat{\alpha}_w$ and $\hat{\beta}, \hat{\beta}_w$ are calculated based on the data from the first *training* period that extends between March 5th 2010 and March 3rd 2011. Similarly, the constant 2σ threshold values are also computed with the use of the first *training* period data.

In figure 2, we use black and orange vertical lines for denoting trade opening and closing times for wavelet filtered series, respectively, and light blue and dark blue for the standard prices; in the figure 'o' denotes trade opening and 'c' denotes trade closing. We remind that the wavelet spread is denoted by the green color, while the standard one corresponds to the dashed red color. For example, under the standard spread, we have two *full-turn trades* and one *non-convergent* trade that was force-closed at the end. As we utilize for our analysis in Subsection 5.2.1, this is an example of a *partially convergent* trade. The same pair turns into a *fully convergent* trade case with six full-turn trades after the application of wavelets. The trading category definitions provided here are important in our understanding of the pairs trading performance after the application of wavelet-filtering.

The ensuing figure, figure 3 shows the resultant returns of the pair both in the standard and wavelet filtered cases. The red (dashed) color is utilized for reporting the level of cumulative returns in the filtered case, and the solid blue one shows the returns progression for the standard one. In this example, trading under the wavelet spread turns out to be more profitable, both because there are more (and profitable) trades in the filtered series, and in the standard approach, the last trade causes a large loss as its *non-convergent* part is force-closed. Under the wavelet spread, however, all trades are naturally closed after they cross the solid 'zero' threshold.

The final point is that the stock selection was made from the paired stocks in our trading dataset and is representative of the issues we would like to touch upon in the following section.

5. Results & analysis

5.1. Preliminary findings

Initially, we start with a detailed analysis and comparison of the results from standard trading versus *symN* wavelet filter trading approaches, followed by the interpretation of these results in Subsection 5.3. For the wavelet results, we only report the results from the application of *sym*22 wavelet filter under Level 1. The results regarding the application of other filter classes is shown in the robustness section (please see Subsection 5.5.1). There, we observe that Symlet filters with $N = 12$, 14, 18 have comparable results. To keep the exposition simple, we only report the details regarding *sym*22 filter.

In addition to the wavelet filters, we also considered (types of) Hodrick Prescott filter, which is a well-known filtering method in the economics literature. Although it is not a widely encountered filtering method in the finance literature, our aim with its use is to show that wavelet-filtering is a much more effective filtering method. Because the results from Hodrick

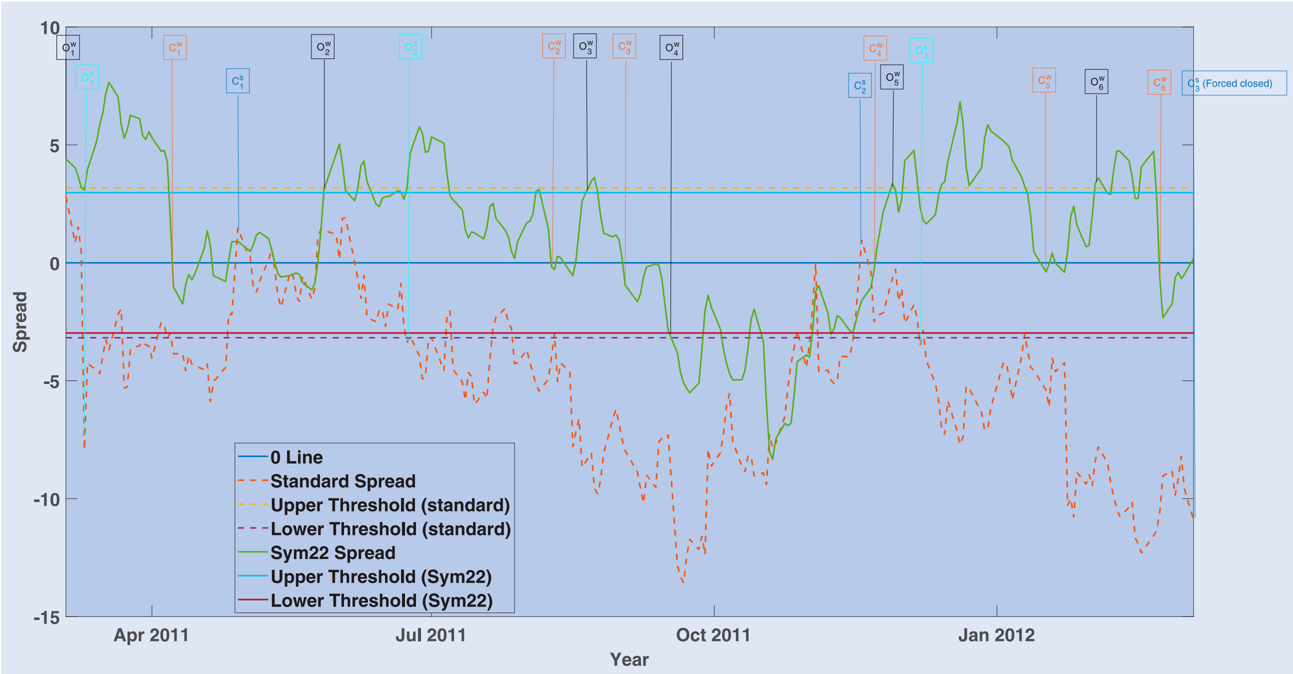

Figure 2. Oneok & Estee Lauder stocks: standard vs. *sym*22 spread series and trades.
Note: In the above the estimates are as follows: $\hat{\alpha} = -8.0866$, $\hat{\alpha}_w = 2.4757$ and $\hat{\beta} = 3.5760$, $\hat{\beta}_w = 2.7771$. The threshold for the standard spread is 3.1770, and for the wavelet spread is 2.9753.

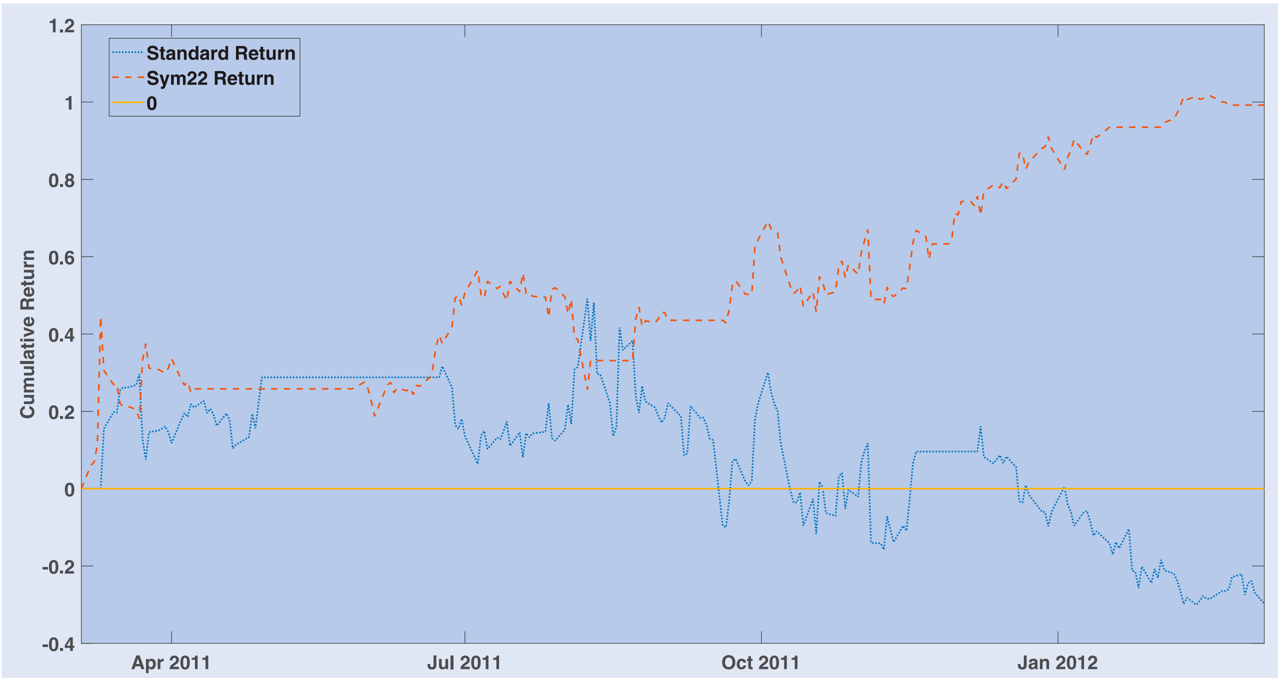

Figure 3. Oneok & Estee Lauder stocks: standard vs. *sym*22 return series.

Prescott filtering is unpromising, they are not reported in the main text† .

As previously mentioned, we consider 1000 pairs for the minimum distance method. The results on the number of cointegrated pairs are reported in table 3:

We select trades from a set amounting to a total number of $(415 \times 414)/2 = 85\,905$ pairs. Since the number of pairs selected is large under the cointegration method, we keep the number of pairs selected for the minimum distance method also (through relatively moderate) large.

5.2. Results

In this section, we report the basic and general results from our pairs trading exercise. In fact, the first two subsections focus on understanding the difference generated by wavelet filtering

† We refer the readers to the online appendix mentioned in the introductory section of this study.

Table 3. Number of cointegrated pairs.

Period	1	2	3	4	5	6	7
Start	03/05/2010	03/04/2011	03/06/2012	03/11/2013	03/11/2014	03/13/2015	03/16/2016
End	03/03/2011	03/05/2012	03/08/2013	03/10/2014	03/12/2015	03/15/2016	03/15/2017
# of cointegrated pairs	4739	5258	2720	3614	5692	4817	3600

the prices. In table 4, the results from the standard pairs trading and the wavelet filtered pairs trading are reported. Note that the values in these tables are computed based on the total profits at the end of each trading period.

From the summary of results, we see a clear improvement in the key measures of the return distributions after the application of the *sym*22 wavelet filters. Particularly, the sample averages of the mean returns during each trading period increase from negative and zero values to a range between 9 and 12%. In the first two rows of the table, we report two types of return results. The first row corresponds to the time when the trades are realized on the day of the signal, while the second one corresponds to trades one day later. We see that the difference between the mean returns computed at two different times are quite small. Consequently, we continue our analysis based on the returns computed at the day of the signal.

As further diagnostics, wavelet-filtering reduces the standard deviations of the trades. Skewness is considerably improved from negative values under the standard spread approach to positive values under the wavelet filter approach. According to Cont (2001), positively skewed returns lead to a more advantageous investment environment. This, in turn, implies more reliance on the Sharpe ratio values. We also see an improvement in the kurtosis values as well, while it is limited under the minimum distance method. Given these improvements, we see in figure 4 the clear divergence of the cumulative returns obtained under the standard and *sym*22 wavelet filter approaches. In that figure, we also demonstrate the returns from two alternative benchmark strategies. These are the cumulative returns from S&P 500 Index and buy & hold strategy of the stocks in pairs devised by the cointegration and minimum distance method. For the buy & hold strategies we assume equal split of a unit dollar investment into the stocks of a selected pair in the training part of our analysis. A comparison based on figure 4 shows that buy & hold reaches to the highest return level as they correspond to two graphs at the top towards the end of the observation period. They are then followed by the wavelet based minimum distance, S&P 500 Index and wavelet based cointegration returns. Nonetheless, one clear visual difference is that the wavelet based returns follow a much smoother path than those of the benchmark strategies. We investigate this further towards the end of this section.

Next, in table 5, we report the results regarding the risk-adjusted returns and their key downside risk measures. The first two sets of results in table 5 are the Sharpe ratio and maximum drawdown measures generated from the daily evolution of the returns. Here, we report the annualized Sharpe Ratio values and observe the substantial improvement after the application of *sym*22 wavelet filter. All mean values are above 2.8, and, as can be observed under the cointegration method, the lowest Sharpe ratio value is above 1.4; this is clearly higher than the Sharpe ratio value $15.91\%/16.00\% = 0.9944$ obtained from table 2 of Section 4.1; the table containing the summary statistics of 415 stocks. The improvement is also reflected in the maximum drawdown measures; under the *sym*22 wavelet spread, and average drawdown values are above the half value of those under the standard spread.

For the graphical display of one of our key measures, we provide figure 5 that shows the superior evolution of the *daily* Sharpe ratios of the two methods under the *sym*22 wavelet spread approach with regards to those under the benchmark strategies. We omit the graphs from the standard approach as their low level is evident from the results in figure 4. Although the supremacy of the daily Sharpe Ratios diminishes towards the end of the observation period, the clear dominance of those from the wavelet approach is apparent in most part of the figure. Such observation, in turn, justifies our earlier observation regarding the smooth path of wavelet based returns in figure 4.

The second set of downside risk measures is provided at the bottom panel of table 5. These are computed from the distributions of the profit/loss levels of the trades at the end of each trading period. After the application of the wavelet filter, we observe a substantial reduction in the level of losses versus standard pairing measured by VaR and CVar. Hence, the improvement in the (left) tail risk with the wavelet application. The improvement in the level of compensation can also be seen in the increasing level of positive returns; we see that the percentage of positive results in most of the trading periods under the standard spread approach is below 50%. However, after the application of *sym*22 wavelet filter, the percentage of positive returns increases considerably to a range between 57% and 76%.

Overall, from the basic statistic values, we see that applying *sym*22 wavelet-filtering to stock price movements and using this filtered series for signal extraction leads to superior pairs trade returns in the out-of-sample period relative to standard approach. We also see that the downside risk measures are considerably improved, and return distributions become more or less positively skewed along with much better risk-to-compensation levels. We investigate such improvement further in the next section. The reader should note at this point that the focus of our research is not finding the best performing strategy, but it is showing the impact of de-noising the prices on pairs trading returns. We inform the reader about few alternative strategies, yet as their numbers are unmanageable, it is neither possible nor our intention to undertake a comprehensive comparison. That is why, from this point onward, we focus solely on how and why filtering affects pairs trading results.

Table 4. Summary of results for basic statistics.

		Cointegration			Minimum Distance		
		Min.	Max.	Mean	Min.	Max.	Mean
Standard	Return	− 6.50%	1.61%	− 1.81%	− 3.42%	1.24%	− 0.55%
	Return $(t+1)$	− 6.77%	1.49%	− 2.06%	− 3.23%	0.85%	− 0.67%
	Std. Dev.	0.4003	1.1317	0.5785	0.2452	0.4565	0.3310
	Skewness	− 48.5924	26.4252	− 4.4250	− 7.5853	3.7158	− 0.8519
	Kurtosis	48.6905	2963.1126	686.8871	11.2337	132.7057	56.6798
Sym22	Return	4.80%	13.53%	9.66%	7.97%	14.69%	11.82%
	Return $(t+1)$	4.90%	13.20%	9.55%	8.10%	14.60%	11.79%
	Std. Dev.	0.3054	0.4655	0.3947	0.2255	0.3941	0.2939
	Skewness	− 8.3035	15.0515	0.3504	− 4.3491	8.0169	2.5257
	Kurtosis	34.7098	494.6974	174.9093	16.4836	113.7411	56.1584

Note: Mean $(t+1)$ stands for opening and closing trades one day after the signal from the spread. Std. Dev. is the abbreviation for standard deviation.

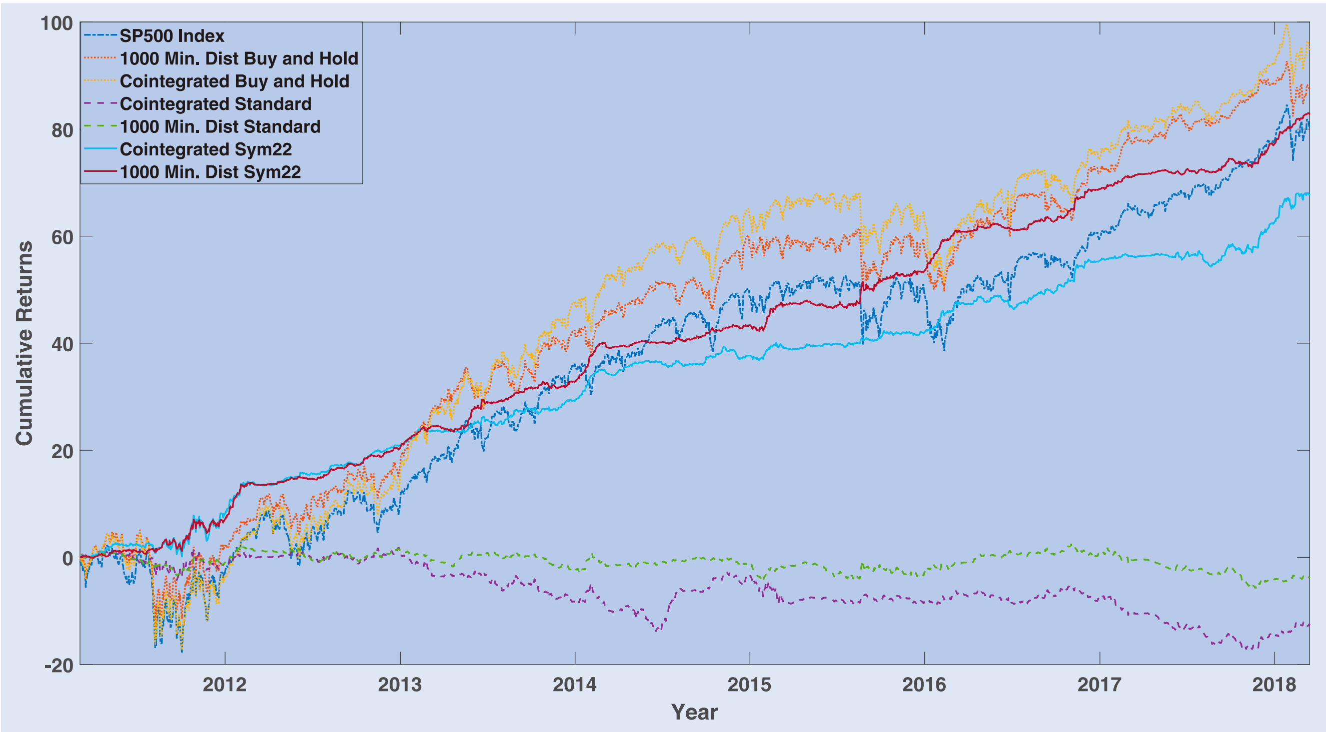

Figure 4. Cumulative excess returns from standard, wavelet and alternative strategies.

5.2.1. Digging deeper. In this section, we dig deeper to understand from where the benefit originates and decompose the (active) trading results into three categories, which are namely convergent, partially convergent, and non-convergent trades. The first two categories are already illustrated in the example provided in Subsection 4.4. The non-convergent category corresponds to trades that opened once, remained divergent until the end of the trading period, and had to be force-closed regardless of the level of the spread at the end of the trading period. The results pertaining to trade categories are reported in table 6.

In the top panel under each approach (standard and *sym*22), we report the percentage of actively traded pairs and the percentage of the category of each active pair. Note that the percentage values of the categories sum to the percentage value of the active pairs as we calculate percentages based on the number of all potential pairs; thus, 100% minus the percentage of active pairs give us the percentage of non-active or not-opened pairs. In the bottom panel, we may observe the profit levels. The key results under the standard approach are as follows: (i) Above 92% of the pairs are actively traded; (ii) the minimum distance method has a lower proportion of non-convergent pairs; (iii) under the minimum distance method, the proportion of partially convergent series is higher; (iv) however, the proportion of fully convergent pairs are almost very close across the two selection methods, so are the average profit levels within the same category.

After the application of *sym*22 wavelet-filtering (i) fewer trades are active under the cointegration method though the difference is not substantial; (ii) most notably, the proportion of fully convergent trades increases (more than double) and proportions of partially and non-convergent trades decrease; (iii) the proportion of non-convergent trades declines; (iv) the proportion of fully convergent trades are again very close

Table 5. Summary of results from the downside risk measures of the trading periods.

		Cointegration			Minimum Distance		
		Min.	Max.	Mean	Min.	Max.	Mean
Standard	Sharpe Ratio	– 1.2644	0.5659	– 0.4044	– 1.1908	0.3530	– 0.2132
	Max. Drawdown	2.06%	7.43%	5.59%	2.50%	5.43%	3.95%
	% Positive Returns	41.86%	56.19%	49.02%	47.50%	58.60%	51.97%
	VaR (5%)	– 62.49%	– 41.47%	– 51.24%	– 44.99%	– 34.04%	– 40.33%
	CVaR (5%)	– 153.99%	– 87.01%	– 107.67%	– 109.13%	– 55.71%	– 76.10%
Sym22	Sharpe Ratio	1.4652	5.4395	2.8199	2.4752	5.2853	3.6925
	Max. Drawdown	0.65%	4.75%	2.30%	0.82%	2.49%	1.48%
	% Positive Returns	57.47%	71.51%	66.00%	71.90%	79.90%	75.86%
	VaR (5%)	– 41.84%	– 16.60%	– 25.52%	– 14.53%	– 10.69%	– 12.00%
	CVaR (5%)	– 109.49%	– 40.12%	– 67.61%	– 61.70%	– 23.59%	– 32.80%
Buy & Hold	Sharpe Ratio	– 0.1668	1.8826	1.0542	0.0983	1.8057	1.0704
	Max. Drawdown	6.67%	21.94%	11.68%	5.54%	18.28%	9.59%
	% Positive Returns	41.09%	95.30%	75.58%	54.10%	97.40%	81.33%
	VaR (5%)	– 24.24%	0.36%	– 10.57%	– 18.79%	3.30%	– 7.26%
	CVaR (5%)	– 29.11%	– 5.25%	– 15.93%	– 22.73%	– 1.19%	– 12.50%

Note: VaR is the abbreviation for the value-at-risk and CVar is the abbreviation for the conditional value at risk.

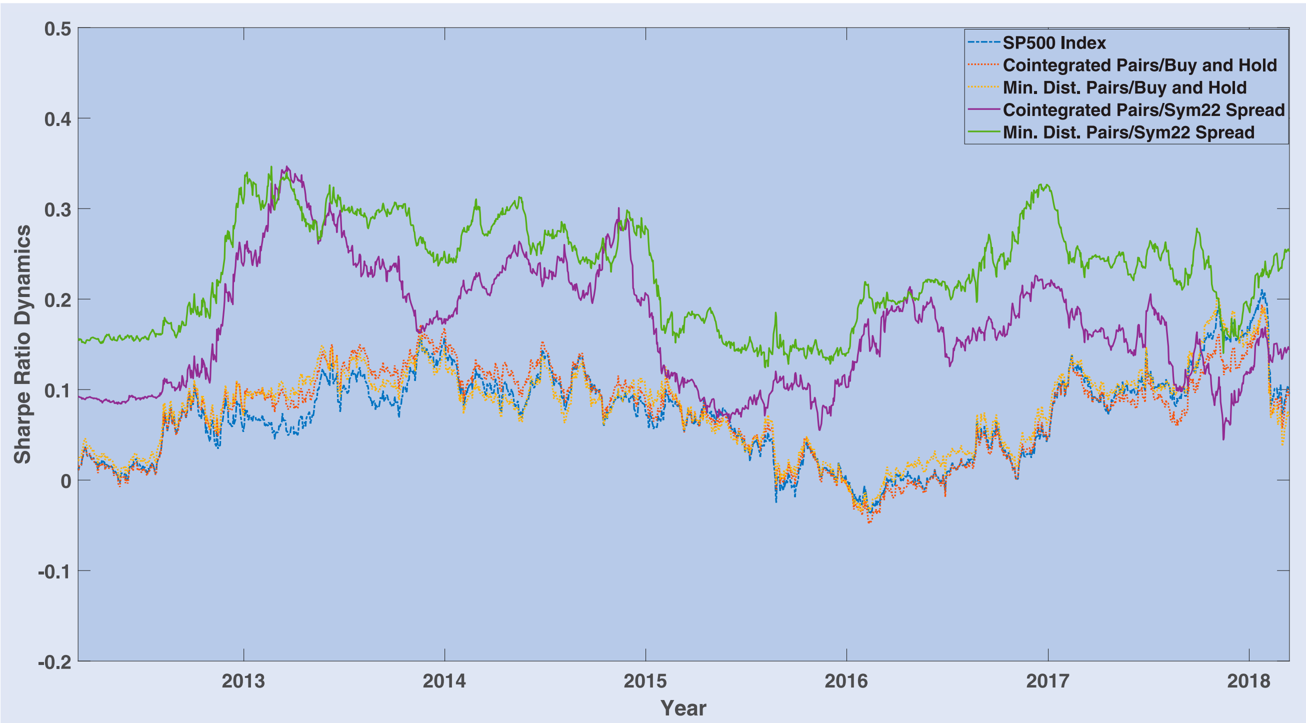

Figure 5. Time series evolution of daily sharpe ratios for the full sample period.

across the two selection methods, but their average profit levels decline; (vi) in contrast, the average loss from non-convergent trades under the standard approach improves substantially and becomes (slightly) profitable on average across the two selection methods; (vii) likewise, average profits from full-turn parts of the partially convergent trades decline, while the average loss levels of the non-convergent parts observed under the standard approach become substantially profitable.

In figures 6 and 7 we check to make sure that the difference in results are not caused by a dominating year, hence we observe the yearly evolution of proportions (left panel) and returns (right panel) per each category. The results confirm our earlier findings. Nonetheless, for the period 2014–15, wavelet filtered series also experience negative returns in the non-convergent trades of cointegration method (see figure 6). Yet, even with such negative returns from non-convergent trades, the overall profitability of filtering stays consistent with our description of the aggregated numbers in table 6.

Another way to see the breakdown of the contributions of each type of trade is to separate their probability of occurrence and their marginal profit. With the transition analysis below, we obtain details that cannot be observed in the earlier tables. We report in table 7 the values based on our calculations from the mean values in table 6. The first two set of columns of the table contain summary of the mean results from table 6, while the last set of columns of the table contain the differences in between the values in the first two. The weighted averages of returns (calculated based on category proportions (P) and

Table 6. Trade convergence & profit analysis.

		Cointegration			Minimum Distance		
		Min.	Max.	Mean	Min.	Max.	Mean
Standard	% of Active Trades	95.80%	99.49%	98.18%	99.60%	100.00%	99.83%
	% of Partial Convergent	33.06%	55.11%	42.25%	45.40%	58.10%	52.49%
	% of Full Convergent	8.64%	18.93%	11.91%	10.10%	15.30%	12.03%
	% of Non-Convergent	31.17%	54.25%	44.02%	31.40%	43.40%	35.31%
	Profit: Partial Convergent (Full-Turn)	8.48%	16.76%	11.04%	7.00%	14.03%	10.10%
	Profit: Partial Convergent (Non-Convergent)	– 7.98%	– 0.05%	– 3.35%	– 5.98%	– 1.17%	– 4.05%
	Profit: Partial Convergent (All)	5.09%	11.38%	7.69%	3.40%	9.64%	6.05%
	Profit: Full Convergent	12.42%	23.81%	19.82%	14.49%	27.73%	19.99%
	Profit: Non-Convergent	– 25.64%	– 11.63%	– 16.77%	– 19.46%	– 12.63%	– 17.14%
Sym22	% of Active Trades	85.72%	97.72%	92.90%	97.70%	99.50%	98.60%
	% of Partial Convergent	19.51%	44.38%	29.77%	27.00%	48.20%	39.20%
	% of Full Convergent	27.21%	40.37%	30.98%	25.60%	39.20%	31.84%
	% of Non-Convergent	26.12%	38.86%	32.14%	19.40%	34.40%	27.56%
	Profit: Partial Convergent (Full-Turn)	6.47%	12.59%	9.48%	7.07%	14.22%	10.70%
	Profit: Partial Convergent (Non-Convergent)	4.85%	9.67%	7.26%	4.93%	8.49%	7.01%
	Profit: Partial Convergent (All)	14.56%	22.26%	16.73%	12.00%	21.71%	17.72%
	Profit: Full Convergent	4.17%	20.44%	14.28%	10.35%	17.16%	13.68%
	Profit: Non-Convergent	– 11.95%	7.41%	0.70%	– 1.88%	5.57%	1.66%

Figure 6. Yearly evolution of proportions and returns—cointegration method.

returns (R)) for each selection method are reported in bold font. For example, under the *sym*22 approach we see that the weighted average of returns are 9.63% and 11.76% for the cointegration and minimum distance methods, respectively.

The differences between the weighted profit values ($P \times R$) show that the improvement in the non-convergent trades category is the highest. For example, under the cointegration method, of the 11.40% increase in profit, 7.61% comes from the improvement in the non-convergent trades. Likewise, under the minimum distance method, 6.51% of 12.23% of the increase is due to improvement in the non-convergent trades. The second-highest gain in profits comes from the fully convergent cases, indicating the importance of a larger number of successful trades that is the result of wavelet filtering.

In sum, with the application of *sym*22 wavelet-filtering, we observe significant improvement both in the percentage of fully convergent trades and an increase in the profits of the non-convergent ones. This increase in the non-convergent parts of the trade categories is in line with our previous finding on the improvement of the downside risk as well as the shift to positively skewed profit distributions.

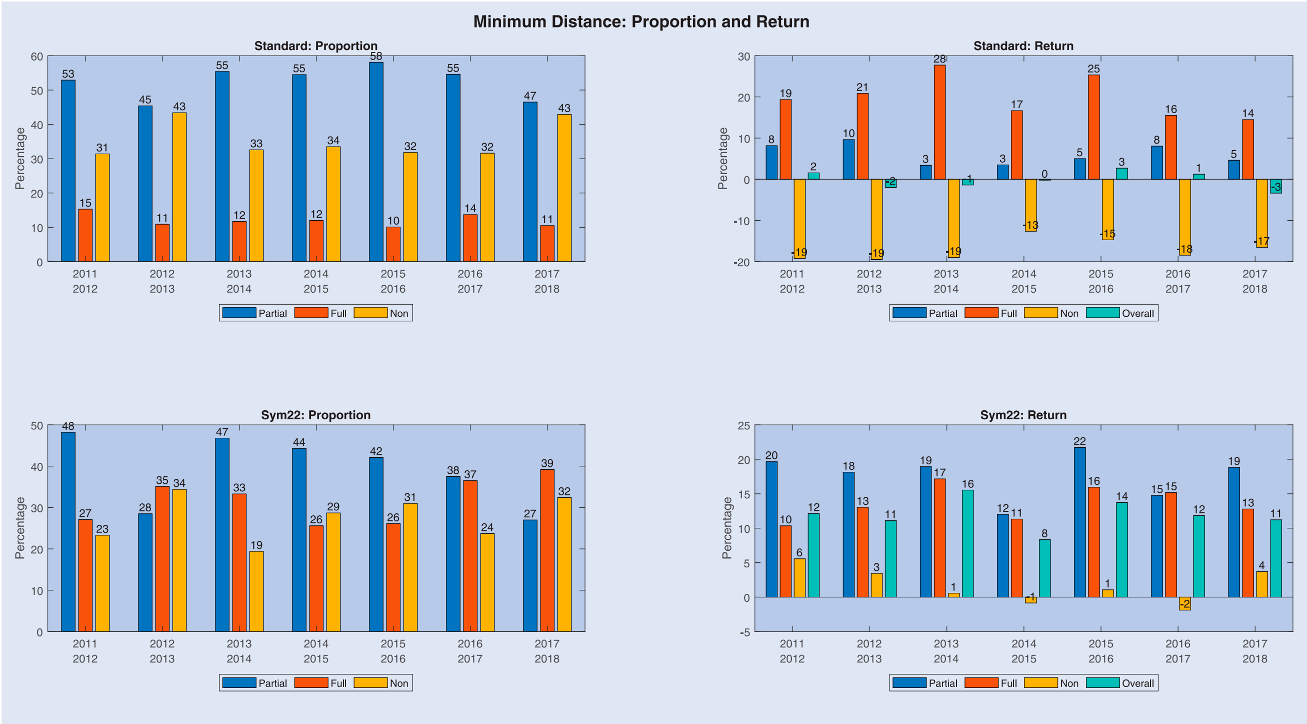

Figure 7. Yearly evolution of proportions and returns—minimum distance method.

Table 7. Proportion & return differences.

	Standard			Sym22			Sym22—Standard Difference		
	Proportion (P)	Return (R)	$P \times R$	Proportion (P)	Return (R)	$P \times R$	(P) Diff.	(R) Diff.	$P \times R$ Diff.
Cointegration			**− 1.77%**			**9.63%**			**11.40%**
Partial Convergent	42.25%	7.69%	3.25%	29.77%	16.73%	4.98%	− 12.48%	9.04%	1.73%
Full Convergent	11.91%	19.82%	2.36%	30.98%	14.28%	4.42%	19.07%	− 5.54%	2.06%
Non-Convergent	44.02%	− 16.77%	− 7.38%	32.14%	0.70%	0.22%	− 11.88%	17.47%	7.61%
Minimum Distance			**− 0.47%**			**11.76%**			**12.23%**
Partial Convergent	52.49%	6.05%	3.18%	39.20%	17.72%	6.94%	− 13.29%	11.66%	3.77%
Full Convergent	12.03%	19.99%	2.40%	31.84%	13.68%	4.36%	19.81%	− 6.31%	1.95%
Non-Convergent	35.31%	− 17.14%	− 6.05%	27.56%	1.66%	0.46%	− 7.76%	18.80%	6.51%

5.3. Interpretation of the results

5.3.1. The spread. The previous subsections show that the two major sources of improvement of de-noising the price series with wavelet filter are (i) increased frequency of successful trades, and (ii) the reduction of losses or increases in gains in the non-convergent trades. The first is due to the spread's evolution during the trading (out-of-sample) period becoming more stationary after the wavelet application. We illustrate this by examining the results in table 8.

The results in table 8 show that there is a substantial improvement in the mean reversion of the spreads in the trading period after the application of the wavelet transform. The increased stationarity not only increases the frequency of successful trades due to greater mean reversion, but also prevents the large deviation of spreads from the common trend between the pairs of stocks, in turn reducing the losses in the non-convergent trades. Recall that during the trading period, trades are executed with the coefficients that are determined in the training period of a year ago. Therefore, a lower rate of stationarity for the standard pairing method emerges since the coefficient estimates of the pair spreads in the year before do not hold next year. Removal of the noise by wavelet-filtering the price series, though, seems to pin down a more accurate relation for the pair and helps the continued use of this relation for more successful trades. The reader should note that despite keeping the trading period very long (one year) and working against the pair trading strategy, filtered series still result in respectable profit levels.

Before we discuss how this improvement comes about lets examine another feature about the spreads, this time explaining the improvement in the non-convergent trade gains in table 9. Inspection of the results shows that the (out-of-sample) standard spreads have non-zero means, indicating that the training estimation performs poorly for such a long trading period. However, wavelet-filtering the price series before estimation causes the spreads to become a lot more *conservative*: Both the means and variances of the spread series in all periods are lower in the trades done with filtered series. This, in turn, implies the following: (i) the reduction in the thresholds leads to more risk-adjusted profitable trades, and (ii) when the spread has not converged (and trade is

Table 8. Wavelet and standard spread unit root test results.

	2011–2012	2012–2013	2013–2014	2014–2015	2015–2016	2016–2017	2017–2018
				Cointegration			
Unit root rejection frequency (standard)	11.48%	3.23%	4.34%	6.12%	7.27%	4.17%	4.22%
Unit root rejection frequency (wavelet)	35.28%	13.41%	40.92%	32.48%	14.51%	32.57%	35.06%
Conversion rate	29.39%	12.48%	38.82%	28.64%	12.40%	31.24%	32.53%
				Minimum Distance			
Unit root rejection frequency (standard)	8.20%	6.20%	5.90%	5.20%	5.30%	5.30%	3.60%
Unit root rejection frequency (wavelet)	29.30%	33.10%	45.60%	37.80%	51.30%	45.30%	39.50%
Conversion rate	25.30%	28.60%	41.00%	34.30%	48.00%	41.30%	37.20%

force-closed), the distance from the common trend is reduced leading to higher gains (or lower losses). Since the non-convergent trades provide the most significant profit increase, the conservative nature of the spread that comes with wavelet transformation has a very crucial role in the success of filtered pairs trading.

5.3.2. The main culprit. As mentioned earlier in Section 1, filtering prices helps in two stages of pairs trading, namely pairing and determining the spread coefficients. Since the main bulk of the profit gain is in the spread formation stage and the two benefits necessitate separate investigations due to different origins, we just focus on the latter stage to analyze how de-noising the prices after the pairing with standard methods can generate such big differences in returns. Spread formation relies on the estimation of the (α_w and β_w) parameters. We claim that the results (above) regarding more frequent stationarity and more *conservative* spreads are caused by a more accurate estimation of these parameters after de-noising of the price series. To motivate our claim, we start by illustrating the correlation between the noise we filter out and the returns to standard pairing techniques in figure 8. Inspection of the figures shows that the higher the (variance of) filtered noise is, the lower the returns will be from unfiltered pairing. Therefore, we deduce that the noise in the data is one of the main reasons why pairs trading have been losing ground in its profitability. Although, wavelet filtering is a very non-linear method of de-noising the data, we provide a linear model in the Appendix A.3 of how filtering leads to consistent estimation of parameters in the presence of noise.

Next, we look even more closely at the noise that is being filtered out with wavelets. For this part, we conduct a principal component analysis on the filtered-out parts of the price series. Analysis of the results shows that the first principal components of the filtered noise capture a large part of the *common* noise. This could also be due to the wavelet transform cleaning out a strong noise component from the price series, which if not cleaned could affect the estimation stage and lead to erroneous α's and β's† (as shown in A.3). If wrong parameters are estimated using the training period, then the spreads will be nonstationary, as was the case in Section 5.3.1, leading to unclosed and unprofitable trades. Such representative principal components are an interesting finding, and the filtered out common noise seems to be very influential as seen in Table 10. Although we conjecture that the filtered noise is the common market noise component, its detailed analysis is beyond the scope of how wavelet-filtering improves pairs trading and is left for further research.

Now we switch the focus from the analysis of the noise and engage in further exercises to show that the *better* parameter estimates are responsible for the positive gains. The first is estimating the best fitting pair coefficients for the *trading* period by using the out-of-sample observations in sample. These estimates, in a cheating way, would give us the best possible values to maximize the pairs-trade returns. Then, we compare the values from the wavelet filtered parameter estimates and the standard parameter estimates. We use the mean squared error (MSE) to measure the proximity of the estimates from both standard and wavelet filtered series. The MSE of the standard pairs-trading coefficients are higher than the wavelet filtered ones by 2748.57 and 595.58 for the cointegration and minimum distance strategies, respectively. Such evidence shows that after filtering, the parameter estimates that we obtain are much more relevant for the trading period and lead to more successful trades (and better spreads).

The results of our second analysis are displayed in table 11. In order to investigate the source of the gains in returns, we vary the filtering, parameters and the thresholds used in the trades. The first four columns are the choices for the type of price series and coefficients utilized in the construction of different spread series. Here, filtered denotes the case of using $(\tilde{V}_{i,1,t}, \tilde{V}_{j,1,t})_{t\geq 0}$, while standard denotes the one with $(S_{i,1,t}, S_{j,1,t})_{t\geq 0}$ in the spread construction. For the case of coefficients, standard is used for $(\hat{\alpha}, \hat{\beta})$ estimates and *sym*22 for $(\hat{\alpha}_w, \hat{\beta}_w)$ estimates. The notation for the threshold is then self-explanatory. For our investigation, we use the evolution of annual Sharpe ratios and finally their average values in the last column. The first two rows of each pairing strategy, cointegration and minimum distance, contain the original results of our pairs trading exercises. After the first two rows, other rows until the final rows correspond to alternative price and coefficient selection scenarios for spread constructions. The last rows represent the hypothetical (Opt) and ideal outcome, in which we carry out our trades using parameter estimates from the *trading* not the training period. In other words, we find the (α, β) coefficient estimates that maximize the pairs' profits in the out-of-sample period.

The results demonstrate a very high level of risk-adjusted returns form the hypothetical scenario at the bottom of the list

† Our dataset does not cover any large crisis periods, yet our claim is that the removal of common market noise prior to pairing will particularly be relevant for turbulent times where all prices tend to move together and lucrative pairs are difficult to identify.

Table 9. Standard error comparison of wavelet and standard spread.

			Standard Spread		Wavelet Spread	
Model	Training	Trading	Mean	Std. Dev.	Mean	Std. Dev.
Cointegration	2010–2011	2011–2012	4.238	5.690	3.988	4.520
	2011–2012	2012–2013	3.597	3.480	3.078	2.805
	2012–2013	2014–2015	6.186	9.654	4.549	5.958
	2013–2014	2014–2015	7.127	10.938	5.175	5.438
	2014–2015	2015–2016	7.785	10.035	6.410	8.272
	2015–2016	2016–2017	7.319	7.896	6.179	6.579
	2016–2017	2017–2018	8.754	11.365	7.532	10.799
Minimum Distance	2010–2011	2011–2012	3.579	3.384	3.482	3.528
	2011–2012	2012–2013	3.238	2.531	2.828	1.973
	2012–2013	2014–2015	4.360	3.483	3.896	3.239
	2013–2014	2014–2015	4.485	3.902	3.656	3.053
	2014–2015	2015–2016	5.792	5.674	4.952	5.401
	2015–2016	2016–2017	5.469	4.698	4.829	4.146
	2016–2017	2017–2018	7.060	7.554	5.803	7.053

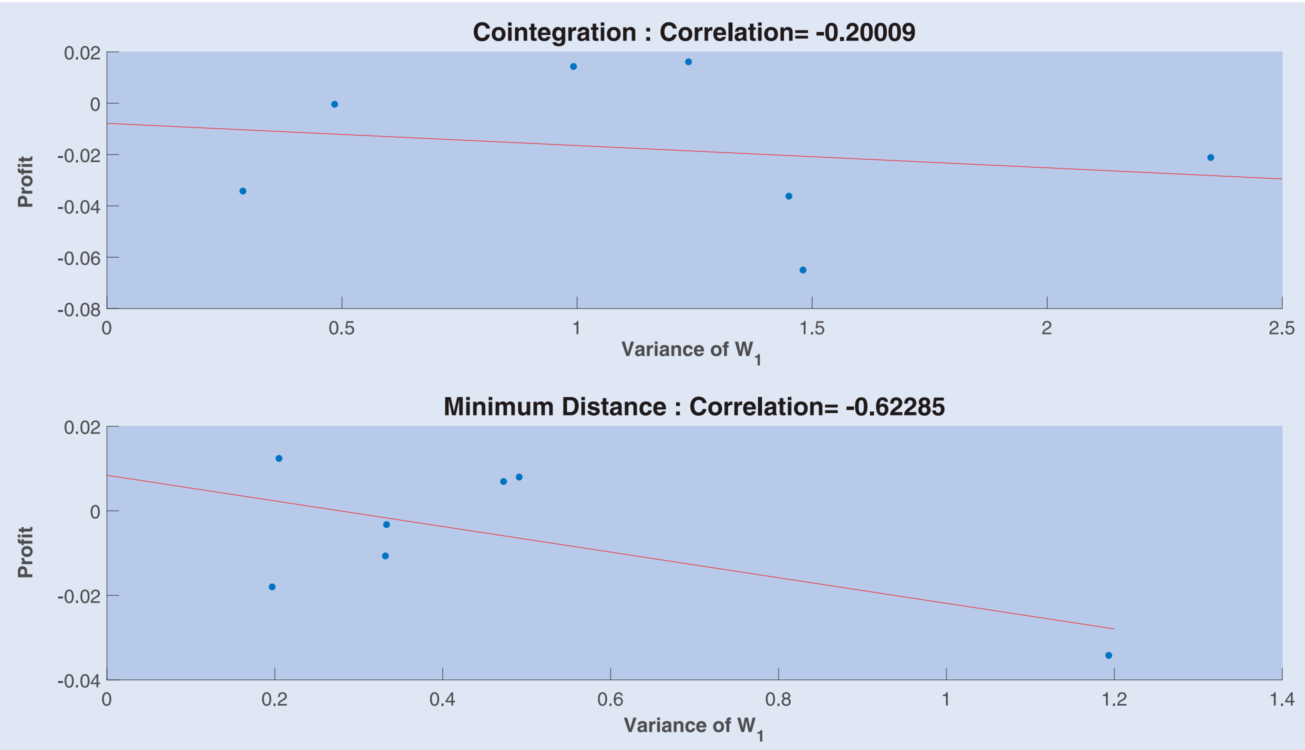

Figure 8. Correlation of filtered noise with the returns from unfiltered pairing.

of scenarios. This clearly shows the potential of high level of profits from pairs trading with better estimation of the coefficients (α, β). The results that follow belong to the construction of the wavelet spread with unfiltered series, (see the third row under each method) and our original wavelet spread. Summarizing the results yield the following notable findings: (i) our strategy is inferior to only three scenarios under the cointegration method, and to only two under the minimum distance method. The superior scenarios correspond to the results from the hypothetical scenario and those where we use the unfiltered prices along with wavelet coefficients, particularly $\hat{\beta}_w$, in spread specification (see third and fifth rows), (ii) whenever we replace $(\hat{\alpha}_w, \hat{\beta}_w)$ with $(\hat{\alpha}, \hat{\beta})$ the risk-adjusted returns deteriorate considerably, and finally (iii) $\hat{\beta}_w$ seems to be more important than $\hat{\alpha}_w$. These results display clearly that de-noising the price series leads to a superior estimation of the pairs coefficients, leading to a better fit of the model determined in the training period to the trading period. Devising the precise cause of this change is complex due to the non-linear nature of pairs trading algorithm and possibly a non-linear relation between the prices†, and is therefore left for future research.

Finally, we conduct a simulation exercise to further show that de-noising the data leads to more accurate parameter estimates. In the simulation exercise, we start by constructing two latent processes that are related to each other through a spread

† The solution of what happens to parameter variances is straightforward if the prices were to be linearly related; however our moving window (year-by-year) estimation hints towards contrary, suggesting a non-linear relation between two price series.

Table 10. Marginal variance explained of the first three principle components of the wavelet coefficients (Noise) from price series.

Training	Trading	Model	1st PC	2nd PC	3rd PC
2010–2011	2011–2012	Cointegration	73.52%	7.05%	3.10%
		Minimum Distance	69.18%	7.91%	3.40%
2011–2012	2012–2013	Cointegration	53.55%	20.01%	5.20%
		Minimum Distance	54.20%	22.34%	2.40%
2012–2013	2014–2015	Cointegration	45.29%	20.36%	8.59%
		Minimum Distance	59.16%	9.45%	3.47%
2013–2014	2014–2015	Cointegration	86.29%	2.45%	1.99%
		Minimum Distance	74.65%	6.13%	2.56%
2014–2015	2015–2016	Cointegration	54.67%	18.64%	3.75%
		Minimum Distance	71.71%	6.34%	2.86%
2015–2016	2016–2017	Cointegration	37.49%	27.63%	5.88%
		Minimum Distance	35.65%	30.15%	4.57%
2016–2017	2017–2018	Cointegration	64.09%	5.91%	5.58%
		Minimum Distance	71.90%	6.91%	5.59%
Average		Cointegration	59.27%	14.58%	4.87%
		Minimum Distance	62.35%	12.75%	3.55%

type specification considered in the study. We call this relation 'the true long-run relation' in between these processes. We later contaminate these series at high-frequency with error processes that are similar to measurement errors and see how OLS regression of the contaminated series work before and after wavelet application.

The construction of the data generating process (DGP) involves reverse engineering of what was done in the main study. We saw that the long-run components from the observed and supposedly contaminated price series are driven first, then a wavelet-based regression model is run to estimate the parameter β_w. Here, we construct the price series by assigning values to parameters exogenously (i.e. set the actual slope coefficient to β^*), then contaminate (variance of the errors are σ_1^2 and σ_2^2 for S_1 and S_2, respectively) them with the expectation that the wavelet-based regression technique will provide better performance in approximating the parameter of interest. As what we do here is similar to the measurement error problem, our a priori expectations are $\hat{\beta}_w$ being less biased than the standard slope estimator $\hat{\beta}$ and the wavelet-based slope estimator $\hat{\beta}_w$ generates less mean squared forecast error than the standard slope estimator $\hat{\beta}$.

We display a portion of the results that are the most relevant for the purpose of our study. † The sample size is set at 250, which is similar to our training period, and all of our simulations are done with 5000 iterations. The first column of table 12 displays the value of β, while the second and third columns are for σ_2^2 and σ_1^2 where they take the values $\sigma_i^2 = a * \sqrt{250}/100$ for $a = 10, 50$. The fourth and fifth columns come from averaging 5000 OLS estimates of $\hat{\beta}$ and $\hat{\beta}_w$. The last columns are mean squared forecast error (MSFE) of standard and filtered price estimates, respectively.

From the results in table 12, we observe a few important features that are consistent with our earlier findings. First, the wavelet-based estimator $\hat{\beta}_w$ suffers a minimal amount of (attenuation) bias while the standard estimator $\hat{\beta}$ deviates more from the actual value β^*. This deviation is almost monotonic with the increase in the variance of the second contamination noise (as expected from a measurement error). Second, in every case, the wavelet-based method generates less MSFE. This finding implies that within a trading period, the wavelet-based method generates less divergent behavior than the standard method. The gap between the two methods is more nuanced with higher values of the variance of the second contamination noise. Third, with higher values of β^*, we observe higher bias in the standard method. This outcome is also apparent in the MSFE.

In sum, our findings from our simulation study also show the frequency domain measurement error bias. Particularly, from the results, we see that when wavelets clear out the high frequency (problematic) components, they lead to more desirable estimates for the slope parameter. In this context, the true slope coefficient β^* governs a long-run (or low-frequency) relation. The deviation from this parameter value can also lead to a divergent behavior for the spread process. Hence, correcting the bias in this parameter leads to better trading performance in more than one way. In the next section, we conduct asset price tests to understand the risk profile of our pairs and administer later some robustness exercises to show that spurious details do not induce the results.

5.4. *Asset price tests of wavelet returns*

To have a view on the exposure of wavelet based returns on the common risk factors, we conduct two Fama-French factor models along with two other alternatives which are namely the *q*-factor model of Hou *et al.* (2015) and the model introduced in Petkova (2006). We consider alternative models to ensure more robustly the level of abnormal returns observed after the application of *sym*22 wavelet filter. In the estimation of models, we obtain the standard errors of the regression coefficients from the Newey and West (1987) heteroscedasticity robust variance covariance matrix. We start our analysis with the Fama-French factor models.

The first Fama-French factor model, in addition to the market, size and value factors of Fama and French (1993), contains the additional factors regarding profitability and

† The detailed description of the simulation is provided in the online appendix.

Table 11. Evolution of annual sharpe ratios under different spread constructions.

			Training Trading	2010–2011 2011–2012	2011–2012 2012–2013	2012–2013 2014–2015	2013–2014 2014–2015	2014–2015 2015–2016	2015–2016 2016–2017	2016–2017 2017–2018	Average
Price	$\hat{\alpha}$	$\hat{\beta}$	**Threshold**				**Cointegration**				
Sym22	Sym22	Sym22	Sym22	1.465	5.440	3.376	1.708	2.2924	2.560	2.267	2.820
Standard	Standard	Standard	Standard	− 0.005	− 0.859	− 1.264	0.227	0.566	− 0.956	− 0.539	− 0.213
Standard	Sym22	Sym22	Standard	1.364	4.679	3.959	2.774	2.948	2.787	1.942	2.922
Standard	Standard	Standard	Sym22	0.339	− 0.207	− 0.663	0.494	0.924	− 0.257	0.458	0.156
Standard	Standard	Sym22	Standard	3.224	7.238	1.415	0.988	3.078	4.593	1.929	3.209
Standard	Sym22	Standard	Standard	− 0.928	− 2.722	− 2.996	0.012	0.570	− 1.652	− 0.294	− 1.144
Sym22	Standard	Standard	Standard	0.351	0.291	− 0.186	0.690	1.109	0.488	1.099	0.549
Standard	Opt	Opt	Standard	5.172	10.224	8.157	8.959	5.450	8.033	9.397	7.913
Price	$\hat{\alpha}$	$\hat{\beta}$	**Threshold**				**Minimum Distance**				
Sym22	Sym22	Sym22	Sym22	2.475	5.285	4.322	2.860	3.150	3.882	3.873	3.693
Standard	Standard	Standard	Standard	0.353	− 0.602	− 0.325	− 0.111	0.183	0.201	− 1.191	− 0.213
Standard	Sym22	Sym22	Standard	2.562	5.846	5.377	3.797	3.412	5.056	3.635	4.241
Standard	Standard	Standard	Sym22	1.120	0.249	0.409	0.521	0.581	0.563	− 0.296	0.449
Standard	Standard	Sym22	Standard	4.695	4.650	2.392	2.216	1.175	2.782	1.117	2.718
Standard	Sym22	Standard	Standard	− 2.523	− 3.435	− 3.920	− 2.956	− 0.411	− 3.040	− 2.336	− 2.660
Sym22	Standard	Standard	Standard	0.825	1.062	1.122	0.531	1.496	1.361	1.084	1.069
Standard	Opt	Opt	Standard	7.277	13.342	12.994	9.119	8.205	10.439	9.216	10.085

Note: Price indicates whether Wavelet filtered or standard series are used to generate spread. Alpha is the intercept coefficient type, beta is the slope coefficient type. Opt indicates a hypothetical state, in which we compute the trading period profit maximizing alpha and beta pairs in spread equation.

Table 12. Simulation results with high frequency contamination.

β^*	σ_2^2	σ_1^2	$\hat{\beta}$	$\hat{\beta}_w$	$MSFE_s$	$MSFE_w$
1	1.581	1.581	0.968	0.998	1.452	0.749
		7.906	0.967	0.998	2.296	0.856
	7.906	1.581	0.862	0.991	2.328	0.858
		7.906	0.864	0.991	2.942	0.952

Table 13. Fama-French five factor models.

	FFPI		FFMR	
	Coint.	Min.Dist	Coint.	Min.Dist
Intercept	0.0004	0.0005	0.0004	0.0005
	(6.069)	(9.675)	(7.035)	(10.398)
Mkt-RF	0.0621	−0.0146	0.0892	**0.0109**
	(8.166)	(−2.303)	(13.424)	**(1.872)**
SMB	0.0934	0.0962	0.0806	0.0918
	(7.122)	(8.783)	(7.012)	(9.122)
HML	0.2189	0.1212	0.1017	0.0382
	(13.274)	(8.806)	(7.954)	(3.419)
RMW	−0.0887	−0.0705	–	–
	(−4.132)	(−3.933)		
CMA	−0.0660	−0.0745	–	–
	(−2.50)	(−3.380)		
Mom	–	–	−0.1706	−0.1052
			(−19.852)	(−13.994)
STrev	–	–	−0.0534	−0.0525
			(−5.235)	(−5.881)
R^2	0.2647	0.1187	0.3930	0.2011

investment. These are namely the robust minus weak profitability portfolios (RMW) and conservative minus aggressive investment portfolios (CMA) considered in Fama and French (2015). We name the factor analysis with these variables FFPI. In the second model, named as FFMR (Gatev *et al.* 2006), we replace RMW and CMA with the momentum (MOM) and short-term reversal (STrev) factors. We also employ other type of Fama-French models by combining all the aforementioned factors in the form of seven factor analysis as well as considering models with the long-term reversal factor (LTrev). We do not report our findings from these models here, as they do not change the main result (related to abnormal returns) of this section. Furthermore, what the original Fama-French three factor model would yield can be deduced from table 14, containing the results of alternative models. When reporting the loadings, we use bold fonts for insignificant variables as considerable number of variables is significant at 1- or 5% level. We also report the t-stats at the bottom of each loading.

The results from the two Fama-French five factor models, reported in table 13, show abnormal returns as the intercept terms are significantly positive. Annualizing the abnormal returns amount to a yearly return of comparable size to our reported returns. In FFPI the yearly return under the cointegration method is 9.16% and under the minimum distance method is 12.23%. Likewise, in FFMR the annualized returns are 9.76% under the cointegration and 12.50% under the minimum distance methods. Similarity of these results with the actual results suggests only a very small portion of wavelet based pairs returns are explained by the factors. We continue our analysis with two alternative models to see if abnormal returns persist.

As alternative, we employ one investment based asset pricing model and one Intertemporal Capital Asset Pricing Model (ICAPM), which was originally introduced by Merton (1973). The first one is the *q*-factor model proposed by Hou *et al.* (2015) with the claim that it captures the many anomalies that Fama-French three factor model cannot. The study also shows that the *q*-factor model performs better than that of Carhart (1997), which adds the momentum factor to Fama-French three factors. In essence, the *q*-factor model is an investment based asset pricing model that stems from the neoclassical *q*theory of investment (Hou *et al.* 2015). The factors of this model are similar to those employed in FFPI model; the first one is again the market factor. The remaining four factors are briefly: the size factor returns (R_ME), investment-to-assets factor returns (R_IA), return on equity factor returns (R_ROE) and expected growth factor returns (R_EG). While for the Fama-French models, we utilized the data from Professor Kenneth French's website, for the *q*-factor model, we use the data from 'global-q.org' website, which contains the data utilized in Hou *et al.* (2015). Also noted in the study, the factors are constructed 'from a triple 2-by-3-by-3 sort on size, investment-to-assets, and ROE'.

The second alternative is a discretized version of ICAPM introduced by Petkova (2006). For the model, Petkova considers the innovations in certain state variables to capture the time variations in investment opportunities. More precisely, she assumes that the excess return of a risky asset can be explained by these innovations in addition to the market factor. The innovations are, in turn, derived through the use of the vector autoregressive (VAR) model employed in Campbell (1996) over the state variables. The states variables are, in addition to the Fama-French three factor model's market, size and value factors, the dividend yield (DIV), the term spread (TERM), the default spread (DEF), and the short-term T-Bill yield (TBILL). With these variables Petkova constructs a VAR model through the triangularization and scaling procedure as in Campbell (1996), and uses the innovations from the model for the estimation of the factor model. Here, we also consider that model with the aforementioned procedure; in Petkova (2006), further analysis is conducted by employing the cross-sectional method of Fama and MacBeth (1973), which is not undertaken in this study. We use the same dataset, which Petkova (2006) utilized, with a slight difference. As in Petkova's study, for the TERM factor we use the difference between the 1-year and 10-year bond, for the DEF factor we use the difference between the yields of Baa rated corporate bond and government bond, and for the TBILL factor we use 1-month T-Bill yield. The data for these factors retrieved from the FRED database of the Federal Reserve Bank of St. Louis. For the DIV factor, we also use the dividend yield of the CRSP value-weighted portfolio. However, for this data, we take the difference between the dividend adjusted and non-adjusted daily returns to calculate the dividend yield; we did not annualize the data as the dependent variable is of daily frequency.

The results of the alternatives are reported in table 14 and show that the intercepts are still strongly significant. When annualized, as we see in table 15, the return levels are

Table 14. q-Factor & ICAPM-Petkova models.

	q-Factor			ICAPM-Petkova	
	Coint.	Min.Dist		Coint.	Min.Dist
Intercept	0.0004	0.0005	Intercept	0.0003	0.0004
	(5.9823)	(9.4577)		(5.3348)	(8.8795)
Mkt-RF	0.0759	**0.0048**	Mkt-RF	0.1150	0.0271
	(9.3181)	(– 1.5322)		(17.4399)	(4.9041)
R_ME	0.0995	0.0889	HML	0.1002	0.0484
	(7.5683)	(10.3513)		(15.1889)	(8.7655)
R_IA	0.0967	**0.0066**	SMB	0.0559	0.0562
	(4.1986)	(0.3535)		(8.4691)	(10.1613)
R_ROE	– **0.0068**	– 0.0423	DIV	**0.0015**	**0.0019**
	(– 0.3133)	(2.4558)		(0.2329)	(0.3365)
R_EG	– 0.1237	– 0.1126	TERM	**0.0087**	**0.0072**
	(– 4.8606)	(– 4.9743)		(1.3215)	(1.2988)
–	–	–	DEF.	– **0.0082**	– **0.0047**
				(– 1.2434)	(– 0.8555)
–	–	–	TBILL	0.0137	**0.0042**
				(2.0753)	(0.7543)
R^2	0.2061	0.1438	R^2	0.2593	0.1056

Table 15. Annualized abnormal returns.

	Cointegration	Minimum Distance
FFPI	9.19%	12.23%
FFMR	9.67%	12.50%
q-Factor	9.44%	11.85%
ICAPM-Petkova	8.11%	11.31%

similar to those in our Fama-French analysis. This, in turn, implies that denoising yields abnormal returns that cannot be explained by the well-known factors in the literature. In effect, as we see from the yearly evolution of the intercept terms in figure 9, the abnormal returns persist despite a decline during the trading period in between 2014–2015. In sum, we see the clear benefit of removing noise in improving returns from the pairs trading. Such noise is a latent factor that needs to be investigated in a future study.

5.5. *Robustness checks*

In this section, we continue with robustness checks. First, we check the results under different wavelet classes. Then, we change the trading (out-of-sample) period time span, and also check the effect of transaction costs. Finally, we force-close non-convergent trades of the standard price cases to see that the 22 days forward-looking behavior of *sym*22 is not the cause of the improvement.

5.5.1. Returns under different wavelet classes. In this section, we evaluate and compare the results under alternative wavelet classes. We report the mean returns and Sharpe ratios from the well-known wavelet classes which are namely Daubechies (*dbN*) wavelets (where Haar is *db*1), Symlets (*symN*) with other vanishing moments and Coiflets (*coifN*). We also tried with other classes such as Fejer-Korovkin and Bi-orthogonal wavelets, but the results of these classes are not reported because they are very unpromising. We may see the mean returns for each year in figure 10.

In that figure each curve represents the mean return at the end of a pertaining trading period of the pairs undertaken in this study. The main pattern is the increase in the return levels with increasing vanishing moments. Clearly the increase seems to be highest under Symlets, nonetheless, Coiflets also provide promising results that may as well be competitive. Such seems to be apparent from the risk-adjusted returns presented in table 16.

The results in table 16 demonstrate that Symlets possesses the highest risk-adjusted returns, followed by the Coiflets. These suggest that the symmetric filtering of the daily price series tend to provide in general better outcomes compared to those of the asymmetric *dbN* filters. Furthermore, the results improve under the wavelet class with higher filter lengths. However, the improvement seems to stabilize after a certain level.

5.5.2. Changing the length of the trading period. To check the effect of time span on the results, we run the pairs trading exercises over a trading period consisting of three, six, and nine months. We check the sensitivity of the mean returns on the length of the trading period because the number of non-convergent trades in our pairs trading exercise is high. We present our findings in table 17. As a graphical display, we also provide figure 11.

The mean returns in table 17 are annualized for a better comparison with the one-year returns provided again at the final column of the table. Overall, for the wavelet cases, we see that the return levels did not change as much, though there is some variability across the trading periods. The minimum distance method's profitability increases with the increasing time span of the trading period. Moreover, the cointegration method yields the highest level of return for a six-month trading period. This result is intuitive as there is a tradeoff between cointegration method finding long-run relations, so performance in the short term might not be as good, and the

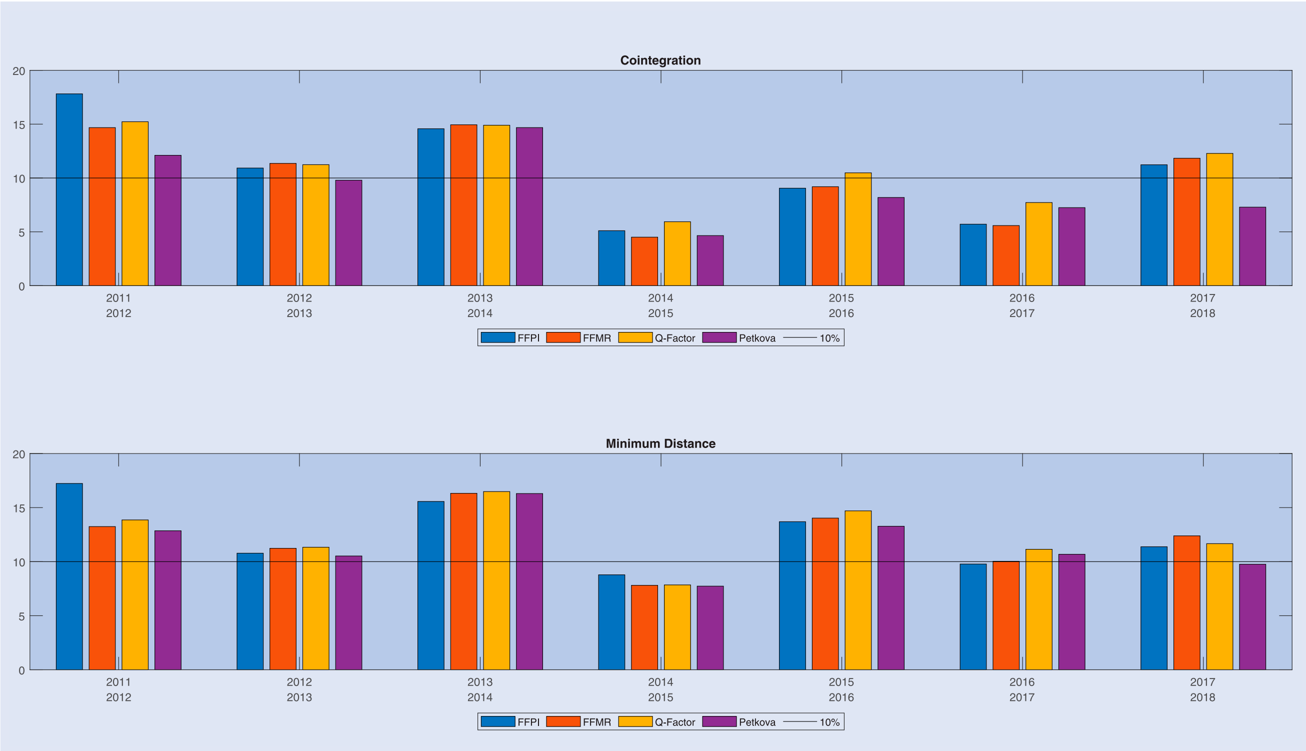

Figure 9. Yearly evolution of abnormal returns.

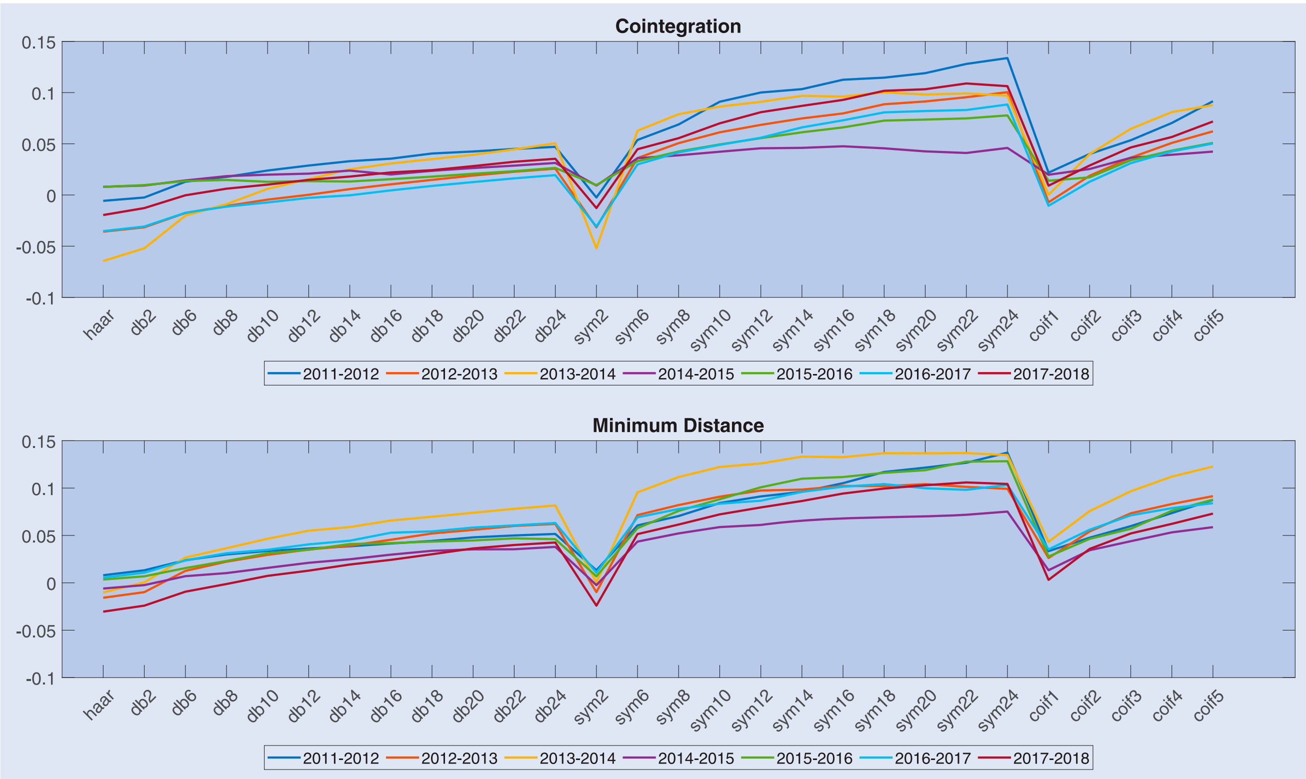

Figure 10. Returns from alternative wavelet classes.

trading period being too long for the coefficients to be relevant for a very long duration. For the standard case, results slightly improve as the trading period gets shorter, showing the poor fit of the training period coefficients (in our study) not being consistently relevant in the trading periods. However, we see that wavelet filtered coefficients' relevance persist longer.

In sum, for the wavelet filtered prices, the returns for three-, six- and nine-month periods fluctuate around 10% return level. This can also be observed from figure 11. There, we provide time series evolutions of two different selection methods to have a better sense of how returns fluctuate throughout the entire trading period between the years 2011 and 2018.

Table 16. Sharpe ratios under different wavelet classes.

	Cointegration			Minimum Distance		
	Min.	Max.	Mean	Min.	Max.	Mean
haar	−1.2759	0.2785	−0.4702	−1.0811	0.2217	−0.2434
db2	−1.0871	0.3382	−0.3717	−0.8692	0.3564	−0.0644
db6	−0.4975	0.4675	−0.0826	−0.3492	0.9567	0.4479
db8	−0.3219	0.5004	0.0574	−0.0472	1.3669	0.7130
db10	−0.2100	0.4740	0.1799	0.2798	1.7804	0.9564
db12	−0.0816	0.5322	0.2942	0.4867	2.1220	1.1539
db14	−0.0091	0.7634	0.4064	0.7311	2.2735	1.3155
db16	0.1353	0.9648	0.4969	0.9121	2.5457	1.5054
db18	0.2686	1.1205	0.6049	1.0154	2.7159	1.6544
db20	0.3884	1.2839	0.7120	1.0858	2.8537	1.7842
db22	0.4882	1.4962	0.8226	1.1111	2.9858	1.8810
db24	0.5132	1.7268	0.9321	1.1374	3.0959	1.9619
sym2	−1.0871	0.3382	−0.3717	−0.8692	0.3564	−0.0644
sym6	0.5796	2.1767	1.2149	1.2675	3.5165	2.2689
sym8	0.7391	2.8180	1.6069	1.4346	4.0470	2.6472
sym10	0.9712	3.0478	1.9077	1.6672	4.5359	2.9445
sym12	1.0647	3.3318	2.1153	1.7557	4.8520	3.1295
sym14	1.1089	3.7928	2.2859	1.8353	4.8676	3.2857
sym16	1.1965	4.1956	2.4019	1.9634	5.0699	3.3679
sym18	1.2367	4.7196	2.5523	2.1682	4.9227	3.4013
sym20	1.2860	4.8992	2.5678	2.2464	5.0206	3.4112
sym22	1.3856	5.1169	2.6106	2.3085	4.8834	3.4204
sym24	1.4852	5.3799	2.6923	2.4660	4.7576	3.4465
coif1	−0.3012	0.4703	0.1282	0.1183	1.6468	0.8681
coif2	0.3963	1.3074	0.6944	1.0657	2.9082	1.7701
coif3	0.5771	2.2470	1.2473	1.2515	3.5781	2.3042
coif4	0.7553	2.9237	1.6428	1.4807	4.0951	2.6802
coif5	0.9763	3.0869	1.9395	1.7117	4.5647	2.9602

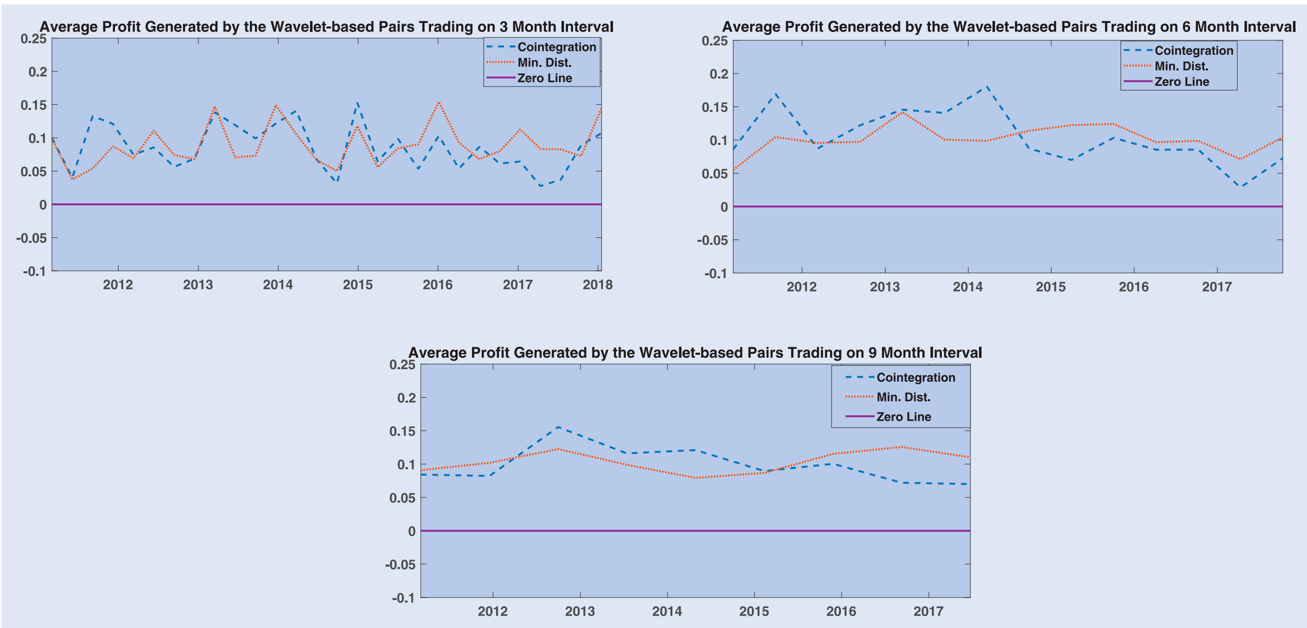

Figure 11. Profits within different trading periods.

Clearly, the graph on the top left contains a twenty-eight period as there are that many three-month observations within seven years. Others are much smoother per our argument above as the frequency of trading periods declines.

5.5.3. Key statistics & risk profile with transaction costs.

In this section, we provide the analysis of the returns obtained after the transaction costs. As the wavelet transformation makes the spread more conservative (due to higher mean reversion), one might think that the number of trades increases. With more frequent trades under conservative spreads, profits may drop due to transaction costs. To see if this is the case, we set the transaction cost parameter $\theta =$ 0.1% (or 10 bps) per share for the half-turn trading. This is equivalent to a value twice the one provided in Avellaneda

Table 17. Pairs trading returns at different trading period spans.

		3-Month	6-Month	9-Month	12-month
Cointegration	Wavelet	8.56%	10.46%	9.91%	9.66%
	Standard	0.23%	– 0.45%	– 0.91%	– 1.81%
Minimum Distance	Wavelet	8.93%	10.18%	10.35%	11.82%
	Standard	1.72%	0.29%	– 0.94%	– 0.55%

Table 18. Key statistics: No-Transaction cost vs. Transaction cost.

		Cointegration			Minimum Distance		
		Min.	Max.	Mean	Min.	Max.	Mean
No Transaction Cost	Mean Return	4.80%	13.53%	9.66%	7.97%	14.69%	11.82%
	Skewness	– 8.3035	15.0515	0.3504	– 4.3491	8.0169	2.5257
	% positive	57.47%	71.51%	66.00%	71.90%	79.90%	75.86%
	CVaR (5%)	– 109.49%	– 40.12%	– 67.61%	– 61.70%	– 23.59%	– 32.80%
	Sharpe Ratio	1.4652	5.4395	2.8199	2.4752	5.2853	3.6925
	Max. Drawdown	0.65%	4.75%	2.30%	0.82%	2.49%	1.48%
Transaction Cost	Mean Return	4.10%	12.80%	9.01%	7.18%	13.69%	10.98%
	Skewness	– 8.6169	15.0463	0.1760	– 4.7160	7.9939	2.3925
	% positive	56.62%	70.08%	65.02%	70.60%	78.50%	74.19%
	CVaR (5%)	– 110.57%	– 40.91%	– 68.87%	– 62.95%	– 24.66%	– 33.76%
	Sharpe Ratio	1.3856	5.1169	2.6106	2.3085	4.8834	3.4204
	Max. Drawdown	0.67%	4.79%	2.35%	0.84%	2.54%	1.53%
	Mean Return Diff.	–	–	0.65%	–	–	0.83%

and Lee (2010). Clegg and Krauss (2018) also uses the transaction cost figure of Avellaneda and Lee (2010) by arguing based on a series of studies that 5 bps is a large enough figure to cover the cost that might arise in the trading of S&P 500 constituents. Given the liquidity of the stocks in the Index, they assume no market impact. We assume likewise and select a higher value in case 5 bps may be considered low.

In table 18, we may observe the results on a selected list of the key measures utilized for the evaluation of the *sym*22 wavelet filter. Under the presence of the transaction costs, a downward shift in the results is natural. Particularly, the shift in the mean results is somewhat proportional to transaction costs by the value of the average number of pairs trades. To see it, we consider the mean return differences reported at the bottom of table 18 and the mean number of trades reported in table 19. The first row (Real Average) of this table consists of the average number of pairs trades of this study over the whole observation period. The bottom row named 'Implied Average' is calculated based on dividing the mean return difference values with the cost of a trade, which is 40 bps; since the half-turn trading cost is 10 bps per traded stock, and since two stocks are traded in a pair, a full-turn trade costs 40 bps. Then, for example, the implied average value under the cointegration method is calculated as $0.65\%/0.40\% = 1.63$, because 0.65% is the total return difference observed in table 18. Values in other selection methods are computed similarly.

A comparison of the 'Real Average' values in table 19 with those of the 'Implied Average' shows that the results are quite close. The difference is the highest under the minimum distance. Such result seems to be due to higher number of trades in this selection method, as the percentage of fully and partially convergent categories in table 6 is the highest under the minimum distance method after the wavelet application. Nonetheless, as we observe in table 20, wavelet application

Table 19. Average # of trades: Real vs. Transaction cost implied.

	Cointegration	Minimum Distance
	Mean	Mean
Real Average	1.59	1.88
Implied Average	1.63	2.09

does not lead to an increase in the average number of trades in each year. Moreover, we see that the standard deviations of the trades are also lower under the wavelet application. These findings demonstrate the advantage of wavelet filtering in terms of transaction costs, as similar numbers with respect to the standard scenario demonstrates that wavelets do not lead to an increase in trades that may cause substantial decline in profits due to transaction costs.

Finally, in table 21, we observe the abnormal return values after the transaction costs. Here as well, we see that even after the transaction costs, abnormal returns remain persistent. In instances, the returns amount to values just above 11%, which is the before transaction cost level reported in Gatev *et al.* (2006).

5.5.4. Closing non-convergent trades 22 days early. One might argue that the forward-looking nature of *sym*22 is responsible for its superior returns performance, especially in the non-convergent cases where the comparative benefit was the highest. The *sym*22 prices are like a forward shadow of the standard prices, and the non-convergent trades stay open longer for the standard pairs trade methods. We check the validity of this claim by closing the non-convergent standard

Table 20. Yearly evolution of average trades.

	Training	2010–2011		2011–2012		2012–2013		2013–2014	
	Trading	2011–2012		2012–2013		2013–2014		2014–2015	
		Mean	Stdev.	Mean	Stdev.	Mean	Stdev.	Mean	Stdev.
Coint.	Standard	1.9641	1.2296	1.5784	0.9474	1.7776	1.0832	2.1859	1.3790
	Sym22	1.7683	1.0229	1.2872	0.9176	1.6772	1.0512	1.6956	1.0265
Min. Dist.	Standard	2.2140	1.3130	1.8560	1.0585	2.0900	1.1168	2.0730	1.1068
	Sym22	2.0460	1.0607	1.7460	1.0052	2.0150	1.0158	1.8980	1.0003
	Training	2014–2015		2015–2016		2016–2017		Overall	
	Trading	2015–2016		2016–2017		2017–2018			
		Mean	Stdev.	Mean	Stdev.	Mean	Stdev.	Mean	Stdev.
Coint.	Standard	2.2430	1.2758	1.5013	0.8708	1.6508	0.9949	1.8430	1.1115
	Sym22	1.8914	1.0378	1.3079	0.8559	1.5078	0.9087	1.5908	0.9744
Min. Dist.	Standard	2.2140	1.1982	2.0530	1.0504	1.8950	1.1078	2.0564	1.1359
	Sym22	1.8900	0.9792	1.8490	1.0026	1.6850	0.9606	1.8756	1.0035

Table 21. Annualized abnormal returns (transaction–cost).

	Cointegration	Minimum Distance
FFPI	8.54%	11.40%
FFMR	9.10%	11.73%
q-Factor	8.79%	11.35%
ICAPM-Petkova	7.46%	10.48%

trades 22 days earlier and see whether the returns improve sufficiently.

The results displayed in table 22 clearly illustrate that the poor performance of the trades using standard prices is not solely due to the non-convergent trades closing late, but it is more due to the poor fit of the models caused by the imprecise parameter estimates of the regressions using standard prices.

6. Conclusion

It is a well-known fact that asset prices are plagued with a variety of noises. Filtration of the noise undoubtedly should result in better analysis and improved financial gain. In this paper, we suggest using the wavelet transform for de-noising stock price series and examining its benefits in two commonly utilized methods of pairs trading. For the application, we select pairs from the constituents of S&P 500 without any sectoral restriction. We also utilize a large set of pairs to have a better view of how well the wavelet transformation improves the results and see how consistent the improvement is across seven successive trading periods.

Using *sym*22 wavelet filters on the paired asset prices, we observe significant improvement over the pairs' returns obtained under the standard spread approach. We show that the underlying factor of the improvement in the returns comes from a more precise estimation of the parameters. Without filtration of the prices, the parameter estimates obtained in the training period are of little use in the trading period. However, better parameter estimates after the filtration lead to higher numbers of more profitable trades. The primary effect is observed in more conservative and faster mean-reverting spreads, implying that the training period parameter estimates also sustain their relevance in the trading period. Such a pattern not only increases the number of successfully closed trades but also leads to an improvement in the gains of the ones that are force-closed at the end of the trading period. We also show that wavelet-filtering prices considerably reduces the downside risk of the trades.

Table 22. Profits from standard pairs trading when the trades are forced closed 22 days ago.

Cointegration			Minimum Distance		
Training	Trading	Profit	Training	Trading	Profit
2010–2011	2011–2012	0.21%	2010–2011	2011–2012	1.80%
2011–2012	2012–2013	− 0.96%	2011–2012	2012–2013	− 0.72%
2012–2013	2014–2015	− 1.64%	2012–2013	2014–2015	1.32%
2013–2014	2014–2015	5.32%	2013–2014	2014–2015	− 0.49%
2014–2015	2015–2016	0.54%	2014–2015	2015–2016	0.53%
2015–2016	2016–2017	− 2.12%	2015–2016	2016–2017	1.50%
2016–2017	2017–2018	− 2.48%	2016–2017	2017–2018	− 3.07%
Average		− 0.16%	Average		0.13%

We intentionally focus on the filtration after the pairs are determined to keep the message of the analysis clear. However, more precise parameters have additional benefits in the pair selection as well. A closer analysis of this fact and a more thorough inspection of what sort of information/noise is filtered out is left for further research.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Supplemental data

Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/14697688.2023.2230249.

ORCID

Burak Alparslan Eroğlu http://orcid.org/0000-0001-6814-747X
Haluk Yener http://orcid.org/0000-0003-2654-5810

References

Avellaneda, M. and Lee, J.-H., Statistical arbitrage in the us equities market. *Quant. Finance*, 2010, **10**(7), 761–782.
Barberis, N., Shleifer, A. and Wurgler, J., Comovement. *J. Financ. Econ.*, 2005, **75**(2), 283–317.
Bertram, W.K., Analytic solutions for optimal statistical arbitrage trading. *Phys. A: Stat. Mech. Appl.*, 2010, **389**(11), 2234–2243.
Campbell, J.Y., Understanding risk and return. *J. Pol. Econ.*, 1996, **104**(2), 298–345.
Carhart, M.M., On persistence in mutual fund performance. *J. Finance*, 1997, **52**(1), 57–82.
Clegg, M. and Krauss, C., Pairs trading with partial cointegration. *Quant. Finance*, 2018, **18**(1), 121–138.
Cohen, L. and Frazzini, A., Economic links and predictable returns. *J. Finance*, 2008, **63**(4), 1977–2011.
Conlon, T., Cotter, J. and Gençay, R., Long-run wavelet-based correlation for financial time series. *Eur. J. Oper. Res.*, 2018, **271**(2), 676–696.
Cont, R., Empirical properties of asset returns: stylized facts and statistical issues. *Quant. Finance*, 2001, **1**(2), 223–236.
Do, B. and Faff, R., Does simple pairs trading still work? *Financ. Anal. J.*, 2010, **66**(4), 83–95.
Do, B. and Faff, R., Are pairs trading profits robust to trading costs? *J. Financ. Res.*, 2012, **35**(2), 261–287.
Endres, S. and Stübinger, J., Optimal trading strategies for lévy-driven ornstein–uhlenbeck processes. *Appl. Econ.*, 2019, **51**(29), 3153–3169.
Engle, R.F. and Granger, C.W., Co-integration and error correction: representation, estimation, and testing. *Econometrica: J. Econom. Soc.*, 1987, **55**(2), 251–276.
Fama, E.F. and French, K.R., Common risk factors in the returns on stocks and bonds. *J. Financ. Econ.*, 1993, **33**(1), 3–56.
Fama, E.F. and French, K.R., A five-factor asset pricing model. *J. Financ. Econ.*, 2015, **116**(1), 1–22.
Fama, E.F. and MacBeth, J.D., Risk, return, and equilibrium: Empirical tests. *J. Pol. Econ.*, 1973, **81**(3), 607–636.
Fan, Y. and Gençay, R., Unit root and cointegration tests with wavelets. In *Canadian Econometric Study Group Meeting*, September. Citeseer, pp. 29–30, 2006 (McGill University: Montreal).
Flori, A. and Regoli, D., Revealing pairs-trading opportunities with long short-term memory networks. *Eur. J. Oper. Res.*, 2021, **295**(2), 772–791.
Gatev, E., Goetzmann, W.N. and Rouwenhorst, K.G., Pairs trading: Performance of a relative-value arbitrage rule. *Rev. Financ. Stud.*, 2006, **19**(3), 797–827.
Gatev, E.G., Goetzmann, W.N. and Rouwenhorst, K.G., Pairs trading: performance of a relative average arbitrage rule. NBER Working Paper 7032, 1999.
Gençay, R., Selçuk, F. and Whitcher, B.J., *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*, 2001 (Elsevier: San Diego, CA).
Hou, K., Xue, C. and Zhang, L., Digesting anomalies: An investment approach. *Rev. Financ. Stud.*, 2015, **28**(3), 650–705.
Ingersoll, J.E., *Theory of Financial Decision Making*, Vol. 3, 1987 (Rowman & Littlefield: Amsterdam).
Johansen, S., Statistical analysis of cointegration vectors. *J. Econ. Dyn. Control*, 1988, **12**(2–3), 231–254.
Johansen, S., *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, 1995 (Oxford University Press on Demand: New York).
MacKinnon, J.G., Numerical distribution functions for unit root and cointegration tests. *J. Appl. Econom.*, 1996, **11**(6), 601–618.
Mallat, S., A theory for multiresolution signal decomposition: The wavelet representation. pattern analysis and machine intelligence. *IEEE Trans.*, 1989, **11**(7), 674–693.
Menzly, L. and Ozbas, O., Market segmentation and cross-predictability of returns. *J. Finance*, 2010, **65**(4), 1555–1580.
Merton, R.C., An intertemporal capital asset pricing model. *Econometrica: J. Econom. Soc.*, 1973, **41**(5), 867–887.
Newey, W.K. and West, K.D., A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 1987, **55**(3), 703–708.
Petkova, R., Do the fama–french factors proxy for innovations in predictive variables? *J. Finance*, 2006, **61**(2), 581–612.
Rad, H., Low, R.K.Y. and Faff, R., The profitability of pairs trading strategies: distance, cointegration and copula methods. *Quant. Finance*, 2016, **16**(10), 1541–1558.
Strang, G. and Nguyen, T., *Wavelets and Filter Banks*, 1996 (Cambridge Press: Wellesley).
Stübinger, J., Mangold, B. and Krauss, C., Statistical arbitrage with vine copulas. *Quant. Finance*, 2018, **18**(11), 1831–1849.
Sun, E.W. and Meinl, T., A new wavelet-based denoising algorithm for high-frequency financial data mining. *Eur. J. Oper. Res.*, 2012, **217**(3), 589–599.
Uddin, G.S., Gençay, R., Bekiros, S. and Sahamkhadam, M., Enhancing the predictability of crude oil markets with hybrid wavelet approaches. *Econ. Lett.*, 2019, **182**, 50–54.
Vidyamurthy, G., *Pairs Trading: Quantitative Methods and Analysis*, Vol. 217, 2004 (John Wiley & Sons: Hoboken).
Vijh, A.M., S&P 500 trading strategies and stock betas. *Rev. Financ. Stud.*, 1994, **7**(1), 215–251.

Appendix

A.1. Trading performance evaluation & transaction costs

We have $n = 1, \ldots, 7$ trading periods. To compute the return of a single trade of a pair in trading period n, let $t_{o,m,n}$ be the opening time and $t_{c,m,n}$ be the closing time of the mth trade in trading period n. Assuming that we have M number of trades in a single trading period, the single period trading return of a pair from the mth trade in period n is

$$R^{m,n}_{i,j,t} = \begin{cases} \text{sign}\left(\epsilon_{w,t_{0,m,n}}\right) \left(-\frac{S_{i,t}-S_{i,t-1}}{S_{i,t_{o,m,n}}} + |\beta|\frac{S_{j,t}-S_{j,t-1}}{S_{j,t_{o,m,n}}}\right), & \text{if } t_{o,m,n} < t \le t_{c,m,n}, \\ & \text{for each } m = 1, \dots, M; \\ 0, & \text{otherwise,} \end{cases} \quad \text{(A1)}$$

where $\epsilon_{w,t_{0,m,n}}$ is the spread at the opening of the mth trade in trading period n, and sign$(\cdot)$ is the sign function specified as

$$\text{sign}\left(\epsilon_{w,t_{0,m,n}}\right) = \begin{cases} 1, & \text{if } \epsilon_{w,t_{0,m,n}} > 0; \\ -1, & \text{if } \epsilon_{w,t_{0,m,n}} < 0, \end{cases}$$

and the value zero in the second line of (A1) implies no active trading. In the above, M clearly changes based on a pair. Here, M includes the number of full-turn trades and all open trades that had to be closed at the end of a trading period n. The subscript w is used to denote the wavelet based spread. Obviously, the computation of the spread under the standard application is done similarly.

Next, let N^n be the total number of the considered pairs including all active and non-active pairs in the trading period n. It follows that the average return from all pairs at time t in the trading period n is:

$$R^n_t = \frac{\sum_{i \ne j}^{N^n} R^{m,n}_{i,j,t}}{N^n}. \quad \text{(A2)}$$

The above specification is the calculation of profits from committed capital. We omit the calculation of the employed capital due to the length and structure of the analysis conducted in the study. While its addition would be informative, its omission will not change the essence of our results. In effect, the above would usually yield a lower return than that would be obtained under the employed capital formula because the denominator of the employed capital formula is the sum of active pairs, which may be less than N^n. As a result, the return value calculated from the committed capital will be less.

Next, we proceed to the formulation of the total average return of all pairs considered in the trading period n. First, we write the total return of a pair from the mth trade in the nth trading period as $R^{m,n}_{i,j} = \sum_{s=t_{o,m,n}+1}^{t_{c,m,n}} R^{m,n}_{i,j,s}$. This is the return from one full-turn trade of a pair. Assuming M full-turn trades in the $n$the period, the total return of a pair in the nth trading period is $R^n_{i,j} = \sum_{m=1}^{M} R^{m,n}_{i,j}$. The average return of all pairs considered in the sample for the nth trading period is then given by,

$$\bar{R}^n = \frac{\sum_{i \ne j}^{N^n} R^n_{i,j}}{N^n}. \quad \text{(A3)}$$

Equations (A2) and (A3) are the computations for the returns of concern in our study. While we mostly rely on the latter when reporting the results in the first analysis, the former return is also utilized for the time series analysis of returns.

A.1.1. Transaction costs.

For the computation of the transaction costs, we let $\theta^{m,n}$ be the proportional transaction cost parameter at the mth trade in the trading period n. Furthermore, at the opening time $t_{o,m,n}$, if we short (long) $S_{i,t_{o,m,n}}$ by \$1, we long (short) $S_{j,t_{o,m,n}}$ by \$$|\beta|$. Then, at the mth trade, if $\epsilon_{w,t_{o,m,n}} > 0$† the total transaction cost for opening mth trade is $(1 + |\beta|) \times \theta^{m,n}$. To compute the transaction cost at the closing of a trade:

- First consider the short position of $1/S_{i,t_{o,m,n}}$ shares of stock i, and long position of $|\beta|/S_{j,t_{o,m,n}}$ shares of stock j at the trade opening;
- When the trade is closed at time $t_{c,m,n}$, we sell stock j, which is worth $S_{j,t_{c,m,n}}|\beta|/S_{j,t_{o,m,n}}$ and buy stock i, which costs $S_{i,t_{c,m,n}}/S_{i,t_{o,m,n}}$;
- Accordingly, the trader should pay $(S_{i,t_{c,m,n}}/S_{i,t_{o,m,n}} + S_{j,t_{c,m,n}}|\beta|/S_{j,t_{o,m,n}}) \times \theta^{m,n}$ at the closing of a trade.

Then, the total cost of the mth trade in the trading period n is

$$\begin{aligned} C^{m,n}_{t_o,t_c} &= \left[(1 + |\beta|) + \left(\frac{S_{i,t_{c,m,n}}}{S_{i,t_{o,m,n}}} + |\beta|\frac{S_{j,t_{c,m,n}}}{S_{j,t_{o,m,n}}}\right)\right] \times \theta^{m,n} \\ &= \left[2\,(1 + |\beta|) + \left(R^{m,n}_{i,t_o,t_c} + |\beta|R^{m,n}_{j,t_o,t_c}\right)\right] \times \theta^{m,n}, \end{aligned} \quad \text{(A4)}$$

where $R^{m,n}_{i,t_o,t_c}, R^{m,n}_{j,t_o,t_c}$ are returns from ith and jth stocks through the lifetime of the mth trade in the trading period n. Continuing with the second line in (A4), we observe that $\theta^{m,n}$ proportion from $(1 + |\beta|)$ is paid twice due to opening and closing a trade, in addition to the profit $(R^{m,n}_{i,t_o,t_c} + |\beta|R^{m,n}_{j,t_o,t_c})$ in the mth trade. With loss, transaction costs are reduced. Nonetheless, we always have $C^{m,n}_{t_o,t_c} > 0$ in the above specification, because the condition $R^{m,n}_{i,t_o,t_c}, R^{m,n}_{j,t_o,t_c} \ge -1$ must hold as stocks may not lose more than the entirety of their values. Observe that when $R^{m,n}_{i,t_o,t_c} = R^{m,n}_{j,t_o,t_c} = -1$, the investor only pays the proportion of $(1 + |\beta|)$ as expected.

A.2. A note on the application in practice

We conduct the analysis on MATLAB and the codes are available upon request. Even though our work heavily emphasizes the theoretical aspect of the problem at hand, the codes can easily be used on real trading situations. Our routine is placed in a single function, which automatically detects trade opening and closing times (dates) and computes the trade gains or losses. We employ this routine recursively as the new data is available. The following scheme summarizes the procedures for one single full turn trading.

Some useful notations for the scheme are as follows: (a) $S_t = [S_{1,t}, S_{2,t}]$ is the price vector. (b) $O_t = 1$ if trade is open and $O_t = 0$ otherwise. (c) t_o is the trade opening time. Let $S_{\cdot,0}$ be $T_0 \times 2$ matrix of training sample and $O_0 = 0$.

(1) Compute the trade signal threshold $(2\sigma > 0)$, spread coefficients and other relevant statistics using wavelet transformation of $S_{\cdot,0}$;
(2) Set $S = S_{\cdot,0}$; For $t = \{1, \dots, T\}$, where T is the trading length,
(3) Receive the new price observations at time t. Append this price observations on the price matrix S. Say the new observation is given by the matrix S_t, where $S_t = [S_{1,t}\ S_{2,t}]$ contains price values until and at time $t \in [0, T]$. Obviously, we update the matrix upon the new price information arrival;
(4) Compute the wavelet coefficients with matrix S;
(5) Using the transformed series, compute the spread $(\epsilon_{w,t})$;
(6) Check the trading condition (Open, close, do nothing). Follow the below priority list:
    (a) If $O_{t-1} = 0$ and $t = T$, terminate the procedure;
    (b) If $O_{t-1} = 0$ and $(\epsilon_{w,t} > 2\sigma$ or $\epsilon_{w,t} < -2\sigma)$, open trade and set $O_t = 1$. Let $t_o = t$. Set $t = t + 1$ and move to Step 3;
    (c) If $O_{t-1} = 0$ and $-2\sigma < \epsilon_{w,t} < 2\sigma$, set $t = t + 1$ and move to Step 3;
    (d) If $O_{t-1} = 1$ and sign$(\epsilon_{w,t}) = sign(\epsilon_{w,t_o})$ and $t < T$, set $t = t + 1$ and move to Step 3;
    (e) If $O_{t-1} = 1$ and (sign$(\epsilon_{w,t}) \ne$ sign(ϵ_{w,t_o}) or $t = T$), close the trade, Compute the trade gain or loss using the original price. Terminate the recursion to look for a new trading opportunity.

For one pair and one new data point with 252 observations in S_0 and sym22 wavelet filter, it takes on average 0.7849 s to complete the routine. Accordingly, our procedure also fits to the high frequency trading (1 m,5 m,… etc). For example, if one wishes to run 1000 pairs at the same time, it will only take 13 minutes. Hence, any trading frequency higher than 13 min is viable for our routine. If the practitioners may run the procedure for many pairs simultaneously, the code can easily be parallelized to reduce the computation time.

† We note that the case $\epsilon_{t_{o,m,n},w} < 0$ leads to the same specification, since trades involved are the same. Therefore, we only report the specification when $\epsilon_{t_{0,m},w} > 0$.,

Furthermore, we may adopt slight variations of the above routine. For instance, in our empirical exercises, we fix the training sample and do not update until the end of the trading period. One can relax this assumption and update the $S_{:0}$ matrix with the most recent T_0 observations after each full turn trade.

A.3. Noisy estimation

Let the DGP be:

$$y_t = \beta x_t + v_t$$
$$\Delta x_t = u_t - \theta u_{t-1}; \quad 0 < \theta < 1.$$

For simplicity we assume that $v_t \perp u_t$. Define $S_{u,t} = u_t/(1-L) = \sum_{i=0}^{t-1} u_{t-i}$.

Then,

$$\Delta x_t = (1-\theta)u_t + \theta \Delta u_t \longleftrightarrow x_t = (1-\theta)S_{u,t} + \theta u_t$$

where, $\mu_{xt} = (1-\theta)S_{u,t}$ is the permanent (low frequency component of x_t and

$$y_t = \beta(1-\theta)S_{u,t} + \theta u_t) + v_t = \beta(1-\theta)S_{u,t} + v_t + \beta\theta u_t,$$
$$\Delta y_t = \beta(1-\theta)u_t + \Delta[v_t + \beta\theta u_t].$$

Moreover we can decompose $y_t = \mu_{yt} + \epsilon_{yt}$, where $\mu_{yt} = \beta(1-\theta)S_{u,t}$ and $\epsilon_{yt} = v_t + \beta\theta u_t$.

We consider two regression models:

$$M1(\textit{with original data}): \quad y_t = \beta x_t + v_t,$$
$$M2\ (\textit{with filtered data}): \quad \mu_{yt} = \gamma \mu_{xt} + \epsilon_t.$$

Then, denoting standardized Brownian motion as B,

$$\hat{\beta} = \beta + \frac{\sum x_t v_t}{\sum x_t^2} \Leftrightarrow T(\hat{\beta} - \beta) \longmapsto \frac{\sigma_v \sigma_u \int (1-\theta) B_u \, \mathrm{d}B_v}{\sigma_u^2 \int (1-\theta) B_u^2} > 0,$$
$$\hat{\gamma} = \beta + \frac{\sum \mu_{xt}\epsilon_t}{\sum \mu_{xt}^2} \Leftrightarrow T(\hat{\gamma} - \beta) \longmapsto \frac{\sigma_\epsilon \sigma_u \int (1-\theta) B_u \, \mathrm{dB}_\epsilon}{\sigma_u^2 \int (1-\theta) B_u^2} \hookrightarrow 0,$$

since $\sigma_\epsilon = 0$. Thus, $M2$ yields more accurate estimates. The intuition is the same as when one regresses a dependent variable with measurement error, $y_t = y_t^* + e_t$, on x_t. That is: $y_t = \beta x_t + \epsilon_t + e_t$. Then, $V(\hat{\beta}) = (\sigma_\epsilon^2 + \sigma_e^2)/\sum x_t^2$ where as filtering is like obtaining the true series $\{y_t^*\}$ yielding $V(\hat{\beta}) = \sigma_\epsilon^2/\sum x_t^2$, which is smaller.