Jonas Rende*

Pairs trading with the persistence-based decomposition model

1. Introduction

It is challenging to accurately model the relationship between a set of securities, yet it is the nucleus of every trading strategy involving more than one security. Prominent examples are trading strategies based on statistical arbitrage. An extensively studied form of statistical arbitrage is pairs trading, which has been introduced to the scientific community by the seminal paper of Gatev et al. (2006). The underlying idea is to identify two stocks with a historically strong co-movement in prices. Next, the price difference between the two stocks, referred to as the spread, is tracked. If the spread diverges, i.e., crosses a pre-defined threshold level, the winner is sold short and the loser is bought. If the historical relationship holds, convergence occurs and results in a profit. The key requirement for the strategy to be successful is the presence of exploitable mean-reverting patterns in the spread. A variety of methods have been proposed to model the potential mean-reversion embedded in the spread, ranging from model-free distance methods, to cointegration techniques to complex time series approaches[1]. A sample of important contributions are Vidyamurthy (2004), Do and Faff (2012), Caldeira and Moura (2013), Rad et al. (2016), Liu et al. (2017) and Clegg and Krauss (2018). Krauss (2017) provides an overview about the pairs trading streams

---

* University of Erlangen–Nürnberg, Department of Statistics and Econometrics, Lange Gasse 20, 90403 Nürnberg, Germany, e-mail: jonas.rende@fau.de.

[1] The distance-approach utilizes simple distance metrics to assess the degree of co-movement of stock pairs. The cointegration approach assumes that all shocks to the spread are solely transient and model the spread as a pure mean-reverting process (Engle and Granger 1987). Most of the studies applying the time series approach uses stochastic differential equations underlying different mean-reverting processes to model the spread.

in the literature – mostly utilizing daily data. In the last years, intraday statistical arbitrage attracts scientific interest on an increasing level.

Next, we briefly review selected high-frequency pairs trading applications. Using U.K. FTSE 100 data for 2007, sampled on 60 minutes, Bowen et al. (2010) report significant excess returns of up to 19.80 percent per annum applying a distance-based pairs selection rule. The authors point out that returns are very sensitive to transaction costs and a waiting rule of one period to account for bid-ask spread. Miao (2014) uses a distance metric and cointegration tests to detect pairs suitable for a pairs trading application on 5 minute bins for 177 selected U.S. stocks between 2012 and 2013. The author finds annual returns of up to 56.85 percent. Focusing on 5 minute bins for U.S. oil stocks (2008, 2013–2015), Liu et al. (2017) achieve annual returns of 187.80 percent applying stochastic differential equations to model the spread's mean-reversion. Recently, Stübinger and Endres (2018) apply a mean- reverting jump diffusion model to S&P oil stocks using minute-by-minute data from 1998 to 2015, reporting annual returns of 60.61 percent. In addition, the authors find supporting evidence for the existence of jumps in the high-frequency data. Using the same data set as Stübinger and Endres (2018), but without restricting themselves to oil companies, Stübinger and Bredthauer (2017) achieve returns of up to 50.50 percent per year with simple distance metrics. Mikkelsen and Kjærland (2018) apply the distance and cointegration approach to 100 stocks of the commodity dominated Oslo Stock Exhange from January 2012 to March 2016. For the cointegration approach the authors report annualized returns of 25.1 percent, wheareas for the distance approach the best specification yields annualized excess returns of 17.8 percent. Other noticeable studies are Dunis et al. (2010), Kim (2011), Kishore (2012), Bowen and Hutchinson (2016) and Endres and Stübinger (2017).

Recently, Rende et al. (2019) propose the persistence-based decomposition model (PBD), as one that adapts well to noisy high-frequency data. The PBD model decomposes the spread of two time series into three components, capturing different levels of shock persistence, namely infinite, finite and no persistence. Thereby, the infinite persistence component is a random walk, the finite persistence component is modeled as a stationary AR(1) process and the no persistence component is pure noise. To evaluate the model in terms of goodness of fit and predictive power, Rende et al. (2019) apply the model to S&P 500 minute-by-minute data, starting from January 1998 and ending in November 2016. Compared to suitable benchmark models the PBD model provides the best fit for the spread between two stocks in 35.38 percent according to the BIC (Schwarz 1978). The author find that the model is superior in light of directional accuracy and RMSE compared to the partial cointegration (PCI) model (Clegg and Krauss 2018), an AR(1) model and a naive approach. The predictive edge of the PBD model is

based on the model's ability to exploit mean-reverting patterns in the presence of noise – the same mechanism underlying a successful pairs trading strategy. Thus, from a trading perspective the finite persistent stationary $AR(1)$ component reflects potentially tradable mean-reversion, whereas the non-persistent component captures non-tradable mean- reversion. With the exception of Bowen et al. (2010), Bowen and Hutchinson (2016) and Liu et al. (2017) the baseline in the current high-frequency pairs trading literature is to execute trades without an execution gap. Leaving out an execution gap leads to *delay-zero alpha* (Kakushadze 2016) returns which cannot be realized in practice: As high-frequency data are subject to microstructure noise. Aït-Sahalia and Yu (2009), backtesting without waiting rules is exposed to trading the bid-ask bounce. Second, due to signal processing it is not possible to generate the signal and execute the order simultaneously.

In total, we make three contributions to the current high-frequency pairs trading literature: First, we put the PBD model to test in a large scale pairs trading application on the same data set as in Rende et al. (2019), using a rigorous backtesting framework. As benchmarks, we choose the PCI model and a buy-and-hold strategy. Second, to gain a better understanding of the model mechanics, we analyze the performance over time, the exposure to common risk factors and compare parameter and industry profiles for all pairs and the most profitable pairs. Third, we quantify the impact of execution limitations on high-frequency pairs trading returns by relaxing the limitations of our backtesting framework.

Our findings are as follows: First, backtesting the PBD model on S&P 500 minute-by-minute data yields a statistically significant and economically meaningful average annual return after costs of 9.16 percent compared to −14.41 percent for the PCI model and 4.38 percent for a buy-and-hold strategy. Thereby, the backtesting engine relies on the following trading costs and execution limitations: (i) We assume trading costs of 0.20 percent per full turn per pair (Avellaneda and Lee 2010). (ii) We implement an execution gap of one period (entry and exit) to account for bid-ask spread (Gatev et al. 2006). (iii) An annual short-selling fee of 1 percent p.a. is charged (Do and Faff 2012). (iv) If for one of the stocks forming a pair, the volume at the time of execution is zero (open and closing trigger), we delay the execution of the order to the next period where volume is larger than zero for both stocks. The PBD model is superior in terms of distributional characteristics, risk measures as well as risk and return measures. Second, an evaluation of the strategy's development over time shows that it performs exceptionally well in bear markets, such as the financial crisis, due to the good performance of the short-leg. To the best of our knowledge, this manuscript is the first pairs trading study documenting the latter fact for the U.S. market on minute-by-minute data. The strategy is mostly robust to common sources of systematic risk. For the most profitable pairs the share of industries typically associated with a similar business

model is high. Thus, industry affiliation is an important return driver. Furthermore, the most profitable pairs exhibit strong pronounced mean-reverting patterns and simultaneously less pronounced permanent effects. Third, relaxing execution limitations drastically increases returns and affects all risk and returns measures in a favorable way. If order execution is not restricted, we find annual returns after costs of 138.6 percent for the PBD model, indicating that delay-zero alpha (high-frequency) returns are upward biased with respect to practical feasibility.

The paper is organized as follows. In section 2, we describe data and software. In section 3, we outline the PCI and the PBD model. In section 4, we provide an in-depth explanation of our backtesting procedure. In section 5, backtesting results and the impact of execution limitations on returns are presented. Finally, section 6 provides some concluding remarks.

2. Data and software

2.1. Data and pre-pocessing

Following Krauss et al. (2017), backtesting is performed on S&P 500 index constituents. Due to its market efficiency, high liquidity[2] and intensive academic, analyst and investor coverage the S&P 500 serves as the litmus test for every trading strategy. In order to eliminate survivor bias we proceed in line with Krauss et al. (2017) and download month end constituents lists for the S&P 500 from Thomson Reuters Datastream, ranging from January 1998 until November 2016. First, the monthly constituents list are transferred into a matrix, covering in a binary fashion if a stock is an index member or not. Second, for every stock having ever been a S&P 500 constituent, minute-by-minute close price information as well as trading volume data are downloaded from QuantQuote starting in 1998 and ending in November 2016 (QuantQuote 2016). Prices are adjusted for splits, dividends and special corporate events, e.g., symbol changes, mergers and acquisitions (QuantQuote 2012). To deal with missing prices, we forward fill close prices between the first and the last available price for every stock. Missing volume information is filled with zeros. In accordance with U.S. trading hours, we solely consider price and volume information between 9.30 AM and 16.00 PM.

In addition, the industry classification according to the Global Industry Classification Standard (GICS) is downloaded for every constituent during the considered time frame. Every stock is matched with its industry.

---

[2] The S&P 500 covers around 80 percent of the available U.S. market capitalization (S&P 500 Dow Jones Indices, 2015)

2.2. Software

Our high-frequency (HF) trading backtester is written in R (R Development Core Team 2018). Risk and return metrics, such as the Sharpe ratio (Sharpe 1994), are calculated using the implementations in the R package *PerformanceAnalytics*. Those metrics need proper time series objects as provided by the package *xts*. To store and access the data in *HDF5* containers and access them we use the package *rhdf5* (The HDF Group 2010). Data manipulation such as slicing and merging has been done in the *data.table* framework. We use R's base implementation *parallel* for parallel processing. A full overview of the R packages, their authors and the purpose is provided in Table 1.

Table 1

R packages involved in the backtester

Name	Author	Purpose
xts	Ryan and Ulrich (2018)	Converting time series objects
Performance Analytics	Peterson and Carl (2018)	Risk and return metrics
rhdf5	Fischer and Pau (2017)	Reading and writing HDF5 container
partialCI	Clegg et al. (2018)	Fitting a PCI model
parallel	R Development Core Team (2018)	Parallel processing
data.table	Dowle and Srinivasan (2018)	Data handing
FKF	Luethi et al. (2018)	Fast Kalman filter estimation
tictoc	Izrailev (2014)	Runtime calculation and benchmarking
compiler	R Development Core Team (2018)	Pre-compile functions to improve run time

3. The models

Next, we will briefly describe the models our pairs trading strategies build up on: First, our baseline is the partial cointegration model of Clegg and Krauss (2018). Second, the persistence-based decomposition model contributed by Rende et al. (2019).

3.1. The partial cointegration model

Compared to classic cointegration (Engle and Granger 1987) partial cointegration accounts for transient and permanent effects. Especially in financial applications, the assumption that the spread between two securities is solely driven by short-term deviations from a long-run equilibrium is too restrictive. To overcome this limitation, the model incorporates idiosyncratic shocks to the spread time series. Examples for such shocks can be a successful management, technological advantages as well as analyst upgrades (Clegg and Krauss 2018). In the partial cointegration framework two securities are connected by a partially autoregressive (PAR – see Poterba and Summers (1988) and Clegg (2015)) process consisting of the sum of a random walk and a stationary $AR(1)$ process, i.e., the absolute value of the $AR(1)$-coefficient $\rho^{P\,CI}$ has to be smaller than one. The model equations are given as,

$$
\begin{aligned}
Y_t &= \beta^{PCI} X_t + Z_t^{PCI} \\
Z_t^{PCI} &= M_t^{PCI} + R_t^{PCI} \\
M_t^{PCI} &= \rho^{PCI} M_{t-1}^{PCI} + \varepsilon_{M,t}^{PCI} \\
R_t^{PCI} &= R_{t-1}^{PCI} + \varepsilon_{R,t}^{PCI}
\end{aligned}
\tag{1}
$$

The error terms of the two components R_t^{PCI} and M_t^{PCI} are assumed to be mutually independent normally distributed zero-mean white-noise processes. Model parameters are the cointegrating coefficient β^{PCI}, the variances of the error terms of the transient component (σ_M^{PCI}) as well as the random walk component (σ_R^{PCI}) and ρ^{PCI}. From a trader's perspective, the mean-reversion, captured by the M component, reflects potential trading opportunities and is therefore of special interest. Due to orthogonality, the model allows to easily calculate the proportion of variance attributable to mean-reversion (R_M^2), allowing to assess the degree of mean-reversion embedded in a time series. As neither the spread nor the distinct components are observable, estimation is solely possible in state space. The authors derive the state space representation and show that Maximum likelihood estimates of the associated Kalman filter are consistent and that the model is identified.

As an initial showcase, the PCI model is applied to S&P 500 daily data starting from 1990 and ending in 2015. The aim of the authors is to construct a relative-value arbitrage strategy relying on the mean-reverting component of the price spread. The distance approach, as proposed by Gatev et al. (2006), and the cointegration approach (Caldeira and Moura 2013) serve as benchmark methods. With monthly returns after costs of 1 percent the PCI model outperforms the benchmark returns of the distance approach (0.1 percent) and cointegration (0.1 percent and 0.15 percent depending on the specification) by far. Therefore, the PCI model serves as the baseline for our high-frequency pairs trading application.

3.2. The persistence-based decomposition model

As pointed out by Rende et al. (2019), the relationship between securities, paradigmatically represented by the spread, might be subject to shocks exhibiting different levels of persistence, namely infinite-, finite-, and non-persistent shocks. Examples for finite shocks can be uninformed trading as well as the short-term price impact of trades, while a prominent examples for shocks with no persistence is microstructure noise due to micro-level market frictions (Rende et al. 2019). Within the PBD time series model the three orthogonal components are modeled by a random walk (infinite persistence), a stationary $AR(1)$ process with coefficient ρ^{PBD} (finite persistence) and pure noise (no persistence). The structural model equations are as follows:

$$Y_t = \beta^{PBD} X_t + Z_t^{PBD}$$
$$Z_t^{PBD} = M_t^{PBD} + R_t^{PBD} + W_t^{PBD}$$
$$M_t^{PBD} = \rho^{PBD} M_{t-1}^{PBD} + \varepsilon_{M,t}^{PBD} \tag{2}$$
$$R_t^{PBD} = R_{t-1}^{PBD} + \varepsilon_{R,t}^{PBD}$$
$$W_t^{PBD} = \varepsilon_{W,t}^{PBD}$$

The full PBD model consists of five parameters. The spread coefficient β^{PBD}, the variance of the stationary $AR(1)$ process (σ_M^{PBD}), the variance of the random walk (σ_R^{PCD}), the variance of the pure noise (σ_W^{PBD}) and ρ^{PBD}. The authors define multiple components to quantify a spread's mean-reversion, each attributable to a different source: First, R_M^2, representing the proportion of variance attributable to finite-persistent mean-reversion. Second, R_W^2 the mean-version attributable to non-persistent mean-reversion. Third, $R_M^2 W$ the total mean-reversion which is the sum of R_W^2 and R_M^2. Given that we cannot observe the components, the model has to be transferred into state space. Rende et al. (2019) derive the likelihood of the associated Kalman filter and elaborate on statistical properties such as identification and consistency of the (quasi) Maximum likelihood estimators.

To gain a better understanding of the PBD model, Rende et al. (2019) investigate if, compared to suitable alternatives, the PBD model fits well to minute-by-minute price spreads for S&P 500 constituents from 1998 to 2016. According to the Bayesian information criterion (Schwarz 1978) and a likelihood ratio test routine for the PBD model in 35.38 percent of the 2 333 202 analyzed pairs the PBD model helps to accurately model the spread. In addition, the out-of-sample prediction power is investigated. In terms of directional accuracy and root-mean-square error the PBD model is superior compared to the PCI model, a stationary $AR(1)$ model and a naive approach. The latter findings are stable over time.

Based on the findings of Rende et al. (2019), we hypothesize that in a high-
-frequency pairs trading application the PBD model adapts better to the noisy
data than the PCI model. To evaluate our hypothesis, we develop a relative-value
arbitrage strategy in the light of Clegg and Krauss (2018) building on the *M* component of the PBD model extracted from the spread.

4. The backtester

4.1. Building blocks, computational challenges, and execution limitations

We implement a backtester to compare the trading results of PCI and
PBD. First, we provide an overview about the different building blocks, computational challenges, and execution limitations. Then, we provide a detailed
description of the different building blocks. Table 2 provides a summary of the
main parameters and selection criteria.

Table 2

Summary of the restricted backtesting framework

	PCI	PBD
Number of non-overlapping study periods	949	949
Formation period, in days	5	5
Trading period, in days	5	5
Eligibility criterion 1	Same industry	
Eligibility criterion 2	Top 5 correlated stocks	
Eligibility criterion 3	LRS in bottom 5%	
Eligibility criterion 4	$0.5 < \rho^{PCI} < 1$	$0.5 < \rho^{PBD} < 1$
Eligibility criterion 5	$R^2_{MR} > 0.5$	$R^2_M > 0.5$
Portfolio size, in pairs	10	
Portfolio selection	In-sample Sharpe ratio	
Opening signal τ_o	2.0	
Closing signal τ_c	−2.0	
Stop loss, in percent	10	
Transaction costs, in percent	0.05	
Execution gap, in periods (minutes)	1	
Volume constraint (open)	$(v_o^Y > 0 \ \& \ v_o^X > 0)$	
Volume constraint (close)	$(v_c^Y > 0 \ \& \ v_c^X > 0)$	

Study periods: We split our data set into 949 non-overlapping study periods (Papadakis and Wysocki 2007). Each study period consists of a five days formation period (FP) and a five day trading period (TP)[3]. Thereby, five days correspond to one trading week. In the formation period the PCI and PBD models are calibrated and the most suitable according to a variety of eligibility criteria (see Table 2 eligibility criteria 1 to 5) are traded in the subsequent trading period.

Return computation: We calculate our returns in line with Gatev et al. (2006). For all pairs, we scale the sum of the payoffs at the end of a trading day with the total amount of the invested capital at the end of the previous trading day. This approach results in a daily return time series. Regarding the invested capital we solely consider committed capital, i.e., 1 U.S. dollar (USD) is spend on every pair at the beginning of the trading period if a position is opened or not. Consequently, we calculate the return on committed capital.

Execution limitations and transactions costs: Our rigorous backtesting framework relies on the following execution limitations and transactions costs:

1. Execution gap: In line with Gatev et al. (2006), we account for bid-ask spread and time for signal processing by shifting an incoming trading signal (entry and exit) at the end of minute t to the end of minute $t + 1$.

2. Volume constraint for opening / closing a position: We only open or close a position if at time $t + 1$ there is positive trading volume for both stocks Y and X, i.e., $(v_o^Y > 0$ & $v_o^X > 0)$ and $(v_c^Y > 0$ & $v_c^X > 0)$, respectively. Thereby, v_o^i (v_c^i) denotes the volume of stock i at opening (o) (closing (c)), where $i = \{Y, X\}$. If the volume conditions are violated we postpone the execution of the order to the next minute where the condition is fulfilled. If during the trading period of interest the condition is never met, the order is canceled.

3. Transaction costs: Following Avellaneda and Lee (2010) we assume transaction costs of 0.05 percent (5 basis points) per half turn per stock, i.e., 0.20 percent per full turn per pair.

4. Short-selling fees: The short leg of a pair is charged an additional fee for the days $\Delta^{d.o}$ a pair is open. We follow Do and Faff (2012) and assume a conservative fee of 1 percent per annum payable for the open days $\Delta^{d.o}$ scaled by 360: $0.01 \cdot \dfrac{\Delta^{d.o.}}{360}$. If $0 < \Delta^{d.o} \leq 1$ we charge one day, i.e, we charge a minimum fee of $0.01 \cdot \dfrac{1}{360}$.

Computational challenges: In our large scale HF application, the size of 65 GB for the raw data imposes a real challenge with respect to RAM handling in the

---

[3] In a trading context, the term day refers to trading days.

context of parallel processing. To deal with RAM constraints, we cut the data into nineteen yearly time-slices, each consisting of approximately 50 batches, where each batch corresponds to a study period, and store them in a HDF5 container. For further analysis, those batches are distributed among five machines consisting of 80 cores (2.4 GHZ to 3.6 GHZ) as well as 206 GB RAM. The jobs are processed in parallel. In this setting, it takes 14 days to fit one specification of the PBD model and 8 days to fit one specification of the PCI model to all 2333202 pairs. Trading takes an additional two days for each model and each specification.

4.2. Formation

Available stocks and pairs: Stocks are treated as *available* if (i) the stocks are a S&P 500 constituent at the last day of the formation period and (ii) no price information are missing within the formation period. A pair is available if the two stocks forming that pair are individually available.

Overnight jumps: Abnormal changes occurring overnight are a well-known phenomenon in high-frequency stock prices (Stübinger and Endres 2018). We replace every price time series by a cumulated return time series, where overnight changes are set to zero (Rende et al. 2019). This approach guarantees that our maximum likelihood estimators are not harmed by overnight jumps (Liu et al. 2017). Consequently, the fitting procedures are carried out on the adjusted time series.

Eligible pairs: Within the 949 study periods we fit a PCI and a PBD model to 2 333 202 pairs fulfilling the eligbility criteria 1 and 2 (Tab. 2), i.e., 2458.59 pairs on average per formation period. Those available pairs are divided into eligible and non-eligible pairs. A pair is considered as *eligible* if the subsequent five eligibility criteria hold simultaneously – see Table 2. First, the stocks have to be in the same industry according to the GICS industry classification. The latter is a common requirement in pairs trading applications (Gatev et al. 2006). Second, a higher correlation between stocks is associated with a higher chance of co-movement (Gatev et al. 2006; Chen et al. 2017). Therefore, we only consider for stock *i* the five industry partner stocks exhibiting the highest correlation (Rende et al. 2019). In addition, limiting the number of available pairs renders our high-frequency pairs trading study computationally feasible. Third, both models rely on maximum likelihood estimation, i.e., it is possible to calculate a likelihood ratio score (LRS) between reasonable nested alternatives and the full PBD or full PCI model, respectively. Turning to the PCI model a reasonable alternative is the pure random walk model ($\sigma_M^{PCI} = \rho^{PCI} = 0$), while for the PBD model the random walk with noise model ($\sigma_M^{PBD} = \rho^{PBD} = 0$) is a suitable alternative. Because lower LRS are associated with a higher chance of following the full model rather than

the alternative we only consider pairs with a LRS located in the five percent quantile. Fourth, we demand $0.5 < \rho^m < 1$, where $m \in \{PCI, PBD\}$. The rationale is that a larger ρ^m is associated with a higher half-life or mean-reversion and, thus higher trade duration potentially reducing bid-ask bounce effects. Fifth, potentially exploitable mean-reversion – as represented by the variance attributable to the stationary $AR(1)$ process mean-reversion – should be pronounced in the spread time series. Specifically, we aim for $R^2_M > 0.5$ for the PBD model and $R^2_{MR} > 0.5$ for the PCI model. The eligibility criteria three to five are inspired by Clegg and Krauss (2018).

Top pairs: We let all eligible pairs participate in an in-sample trading and calculate the respective Sharpe ratio – see Bertram (2010), Dunis et al. (2010), Caldeira and Moura (2013), Clegg and Krauss (2018). Next, for each stock *i* its eligible partner stocks are ranked in descending order according to their in-sample Sharpe ratio. We refer to stock *i*'s partner with the highest in-sample Sharpe ratio as its *trading partner* (Clegg and Krauss 2018). We extract a list solely containing stocks and their trading partners. A portfolio consisting of the top 10 pairs (Miao 2014, Stübinger et al. 2018) in the sense of the largest in-sample Sharpe ratio is then traded in the the following trading period.

4.3. Trading

As a relative-value arbitrage strategy relies on mean-reverting patterns, we monitor a transformation of both models *M* component (see – equation (3)) during the trading period to detect trading signals. We do not monitor the spread itself as typical in the distance approach (Gatev et al. 2006) or the cointegration approach (Caldeira and Moura 2013) because the spread is non-stationary by construction in our case. To be specific, we calculate the standard score for both models for every minute *t* in the five days trading period given as

$$S_t^{m,TP} = \frac{M_t^{m,TP}}{\sigma_M^{m,FP}} \tag{3}$$

where $m = \{PCI, PBD\}$. It is common to assume that the estimated in-sample spread relationship between two securities carry over to the out-of-sample trading period. Hence,

the parameters needed to calculate the standard scores in the trading period are estimated in the formation period using the model equations (1) and (2), respectively. Loosely speaking, a trading signal is a *sufficiently large* deviation of the *M* component from its mean value. In mathematical terms, sufficiently large

means $|S^{m,TP}| > \tau_o$, i.e., the standard score crosses the opening threshold τ_o. Following Gatev et al. (2006), Huck (2015), and Rad et al. (2016) we opt for $\tau_o = 2$. A position is opened by going long the undervalued security and shorting the overvalued stock. If $|\tau_o|$ is crossed while a position is already open we do not take an action. An investment of 1 USD in the long position is offset by selling short the price ratio of the two securities forming the pair weighted with the in-sample spread coefficient β^m. In general, the latter will differ from 1 USD. The portfolio value might change minute-by-minute depending on whether we open a position as reaction to a trading signal. When a trade is entered the maximum amount of capital is invested. The lower bound of the investment is a stop-loss rule of 10 percent, i.e., if the portfolio value drops below 90 percent of its initial value the pair is excluded from any further trading activities (Nath 2003, Caldeira and Moura 2013). An upper bound is not implemented. If no position is open the portfolio is held in cash, i.e., we assume that there are no interest earnings. Open positions are closed if $|S^{m,TP}| < \tau_c$, where τ_c denotes the closing threshold. We choose symmetric thresholds and set $\tau_c = -2$ (Bertram 2010). In addition, a closing action is forced if a stock is not listed anymore or if the end of trading period is reached and a position is still open. The last available price is then used to calculate profits. The PBD model, as well as the PCI model, can notch up losses from converging trades, due to the behavior of the random walk component. By contrast, the distance and the cointegration approach always earn profits from converging trades.

5. Results

In this section, we evaluate the performance of the PBD, the PCI model and the general market (MKT) from a financial perspective. The latter is a simple S&P 500 buy-and-hold strategy. We proceed as follows: First, we analyze the performance in the light of our backtesting framework outlined in section 4. Second, we relaxing the execution limitations step-by-step to gain a better understanding of how the restrictions affect the trading results.

5.1. Backtesting results

5.1.1. Risk and return characteristics

First, we pin down the performance of the different approaches in terms of a variety of daily and annualized risk as well as return measures. Table 3 summarizes the results.

Panel A – daily return metrics: With a value of 0.04 percent per day, the PBD model yields the highest mean return after transaction costs compared to PCI (−0.06 percent) and the general market (0.02 percent). For the PBD and the PCI model the return values are highly significant. The large number of trades explains the fairly low standard deviations of 0.0076 for the PBD model, 0.0103 for the PCI model and 0.0124 for MKT, respectively. Moreover, the PBD model exhibits the highest hit rate (Share > 0) with a value of 54.56 percent compared to 47.28 percent for the PCI model and 53.01 percent for MKT. While the PCI approach and the MKT are skewed to the left, the daily return time series of the PBD model is skewed to the right. The latter indicates that the right tail is more pronounce which is beneficial from an investors perspective (Cont 2001). Kurtosis values of 10.46 for the PBD model, 7.54 for the PCI model and 7.75 for the MKT are associated with a notable risk in the tails of the distribution. This leptokurtic behavior is well in line with the stylized facts of financial time series (Cont 2001).

Table 3

Daily and annualized risk-return metrics – restricted backtesting. Panel A depicts daily return characteristics, panel B depicts risk and panel C annualized risk-return metrics for the the persistence-based decomposition model (PBD), the partial cointegration model (PCI) and the general market (MKT)

	PBD	PCI	MKT
A			
Mean return	0.0004	−0.0006	0.0002
Standard Error (Newey-West)	0.0001	0.0002	0.0002
t-Statistic	3.0886	−2.9522	1.5479
Minimum	−0.0583	−0.0881	−0.0903
25% Quantile	−0.0031	−0.0051	−0.0055
Median	0.0001	−0.0003	0.0005
75% Quantile	0.0035	0.0043	0.0061
Maximum	0.0859	0.0794	0.1158
Share > 0	0.5456	0.4728	0.5301
Standard dev.	0.0076	0.0103	0.0124
Skewness	0.6063	−0.3100	−0.0272
Kurtosis	10.4654	7.5358	7.7488

Table 3 cont.

	PBD	PCI	MKT
B			
Hist. VaR 1%	−0.0313	−0.0447	−0.0515
Hist. CVaR 1%	−0.0282	−0.0432	−0.0483
Hist. VaR 5%	−0.0091	−0.0169	−0.0184
Hist. CVaR 5%	−0.0162	−0.0260	−0.0293
Maximum drawdown	0.4637	0.9813	0.5678
C			
Return p.a.	0.0916	−0.1441	0.0438
Standard dev. p.a.	0.1200	0.1639	0.1974
Downside dev p.a.	0.0776	0.1229	0.1396
Sharpe ratio p.a.	0.7630	−0.8792	0.2217
Sortino ratio p.a.	1.1793	−1.1721	0.3135

Panel B – daily risk metrics: In terms of all risk metrics the PBD model is favorable to the benchmark approaches. The values for the 1 and 5 percent value at risk (VaR) and the corresponding conditional versions (CVaR) indicate that the PBD approach exhibits a lower tail risk. An an example compare the 1 percent VaR of −0.0313 for the PBD model to −0.0447 for the PCI model and −0.0515 for the MKT. In addition, the exposure to drawdown risk, as measured by the maximum drawdown is found to be lower for the PBD model.

Panel C – annualized risk metrics: With an average annualized return of 9.16 percent after transaction costs the PBD model outperforms the general market (4.38 percent) and the PCI model (−14.41 percent) by far. The higher annualized returns are achieved with a significantly smaller standard deviation which translates to the highest annualized Sharpe ratio of 0.7630 for the PBD model. For the general market the Sharpe ratio is 0.2217 and the PCI model exhibits a negative Sharpe ratio of −0.8792. If we solely focus on downside risk as reflected by the downside deviation and the Sortino ratio, the advantage of the PBD model compared to the PCI model and the MKT is even more pronounced. This is in line with the positive skewness and a higher rate as described in panel A.

5.1.2. Trading statistics

Next, we compare the PBD and the PCI approach in light of trading statistics – see Table 4). The average number of pairs traded per 5-days trading period for the

PBD model is 9.824, i.e, nearly all of the 10 pairs are traded. Turning to the PCI model, only about half of the 10 pairs are traded on average. A deep-dive analysis on the latter point yields that the method is not able to identify a sufficiently large number of eligible pairs. Panel B shows descriptive statistics of the number of eligible pairs identified across all 949 study periods. Note that due to eligibility criterion 3 (Tab. 2) the maximum number of eligible pairs is 5 percent of the pairs fulfilling criteria 1 and 2, i.e., on average $2458.59 \cdot 0.05 \approx 123$ pairs. Out of the approximately 123 pairs solely 7.55 pairs are eligible on average for the PCI approach compared to 30.26 for the PBD approach. This low mean value combined with a median value of 5 for the PCI method explains the difference in the average number of pairs traded per 5-days period. Interestingly, the trading frequency, quantified by the average number of round-trip trades per pair, is more than twice as high for the PCI model (14.78) compared to the PBD approach (6.52). Consequently, the average time pairs are open per round-trip is significantly smaller for PCI. On average pairs are open 0.25 days for the PCI approach, while this value is 0.63 for the PBD model. The share of trades where the volume constraints are violated is about 6 percentage points larger for the PBD model and average waiting time is longer (4.9 minutes versus 3.8 minutes). The latter findings provide a first indication that volume constraints might affect returns. With a value of 94.82 percent the share of converging trades is very high for the PCI model. For the PBD model the share of converging trades is about 8 percentage points lower.

Table 4

Trading statistics and eligible pairs – restricted backtesting. Panel A depicts summary statistics of the number, the duration of trades, the volume constraints and the distribution of converging and diverging trades, and panel B shows descriptive statistics for the number of eligible pairs across the 949 study periods for the the persistence-based decomposition model (PBD) and the partial cointegration model (PCI)

	PBD	PCI
A		
Average number of pairs traded per 5-days period	9.8240	5.0474
Average number of round-trip trades per pair	6.5168	14.7762
Standard deviation of number of round-trip trades per pair	6.1182	10.6481
Average time pairs are open per round-trip, in days	0.6310	0.2477
Standard deviation of time pairs are open per round-trip, in days	1.0037	0.8648
Share of trades with zero volume	0.2204	0.1460

Table 4 cont.

	PBD	PCI
Average time a trade is delayed due to zero volume, in minutes	4.9233	3.8097
Share of converged trades	0.8659	0.9482
Share of trades where closing is forced, stop-loss	0.0044	0.0022
Share of trades where closing is forced, end of data	0.1297	0.0496
B		
Minimum	0.00	0.00
25% Quantile	20.00	1.00
Median	30.00	5.00
Mean	30.26	7.55
75% Quantile	40.00	11.00
Maximum	87.00	56.00
Share study periods with zero eligible pairs [%]	2.00	15.00

The analysis of trading statistics at least partly explains why PCI returns are not robust: First, the low number of eligible pairs is associated with a lower chance of detecting profitable pairs. Second, the duration of round-trip trades for the PCI model might be not long enough for an exploitable relationship to establish. Third, the high number of converging trades along with the low average return could be an indication that the random walk component develops in an unfavorable way harming profits.

5.1.3. Performance over time

Pairs trading profits are known to decline over time – see Do and Faff (2010), Bowen and Hutchinson (2016) and Stübinger and Endres (2018). Figure 1 tracks the cumulative profit (committed capital) of a 1 USD investment in the PBD approach (black), the PCI approach (blue) and the general market (red) from 1998 to 2016. With exception of the early years the PBD approach is superior to the performance of PCI and the general market. For the PBD approach we observe a *M*-pattern, with peak periods in bear market phases. The first peak period starts 2001 and ends 2003. This period is associated with decimalization, the emergence of the dot-com bubble, the September 11 attacks and the beginning of the Iraq war. The second peak corresponds with the financial crisis. The good performance

of pairs trading strategies in market phases exhibiting bearish conditions is well-documented for low- and high-frequency data – see Do and Faff (2010), Bowen and Hutchinson (2016), Liu et al. (2017), Clegg and Krauss (2018) and Stübinger and Endres (2018). The reason is that, by construction, the strategy is liquidity providing (Do and Faff 2010). In the period ranging from 2001 to 2003 the PCI model also achieves positive profits but at a significantly lower level than the PBD model. In addition, the PCI model cannot capitalize on the financial crisis. The general market shows a weak positive trend.

After 2010, profits for the PBD and the PCI model start to decline, while the general market shows a positive trend. As pointed out by Clegg and Krauss (2018), declining pairs trading profitability is most probably due to an increase in market efficiency due to the wide-spread adoption of quantitative strategies, such as pairs trading. At the end of period covered in our data, the PBD model yields a cumulative profit of 4.05 USD, whereas the PCI model yields a cumulative profit of –0.91 USD. For the general market this value is 0.45 USD.

The evaluation of Figure 1 shows that the PBD model performs very well in bear markets. We would expect that in bear markets especially the short leg contributes to the profitability of the strategy. For lower frequencies, supporting evidence is provided by Bowen and Hutchinson (2016). By contrast, for high-frequency data only little is known about the contribution of the long and short leg at different market phases. At the time of writing, there is no high-frequency study providing empirical evidence for the U.S. market[4]. To evaluate the hypothesis, we separately plot the financial performance (Fig. 2) of the long and short leg (black) as well as of the long (red) and the short leg (blue). Before 2001, with the exception of the 2000s, both legs roughly provide the same contribution to the total profit. Starting from the end of 2001 to the beginning of 2016, the short leg yields higher profits than the long leg. During this time span, the USD development of the short leg has two peaks, which coincide with the two peaks of the USD development of both legs together, i.e., the two prominent bear market periods. The latter supports the hypothesis that within bear market phases the short leg contributes more to the overall profit than the long leg for high-frequency data. Note that from 2010 to November 2016, short-selling profits decrease, while profits due to the long position show a positive trend. The underlying rationale is that after the financial crisis the U.S. economy started to recover and the general market increased (see Fig. 1).

---

[4] Using the distance and cointegration approach, Mikkelsen and Kjærland (2018) find for 100 Oslo Stock Exchange constituents from January 2012 to March 2016 that on average the long position contributes more to the overall profit than the short leg.
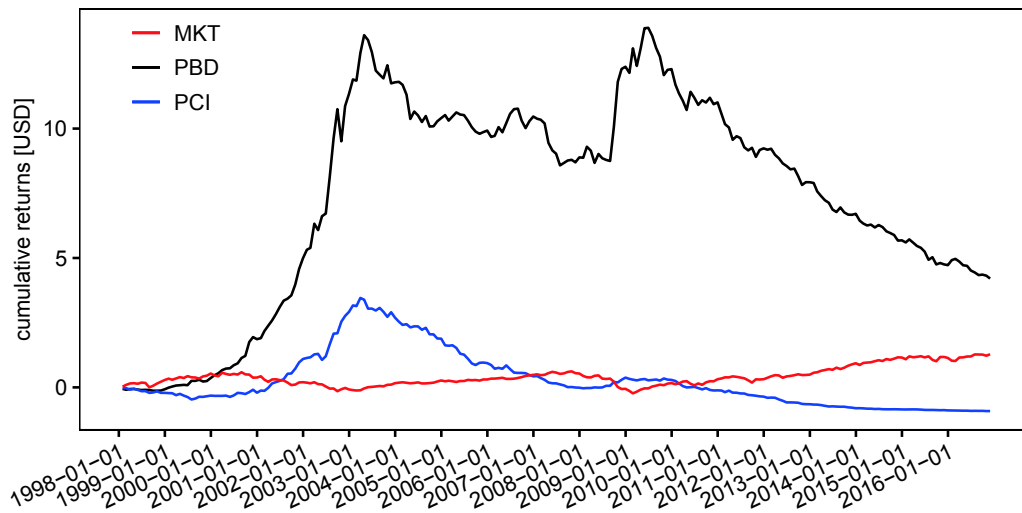
Figure 1. Financial performance over time. We plot the development of the financial performance (committed capital) over time for the PBD model, the PCI model and the general market
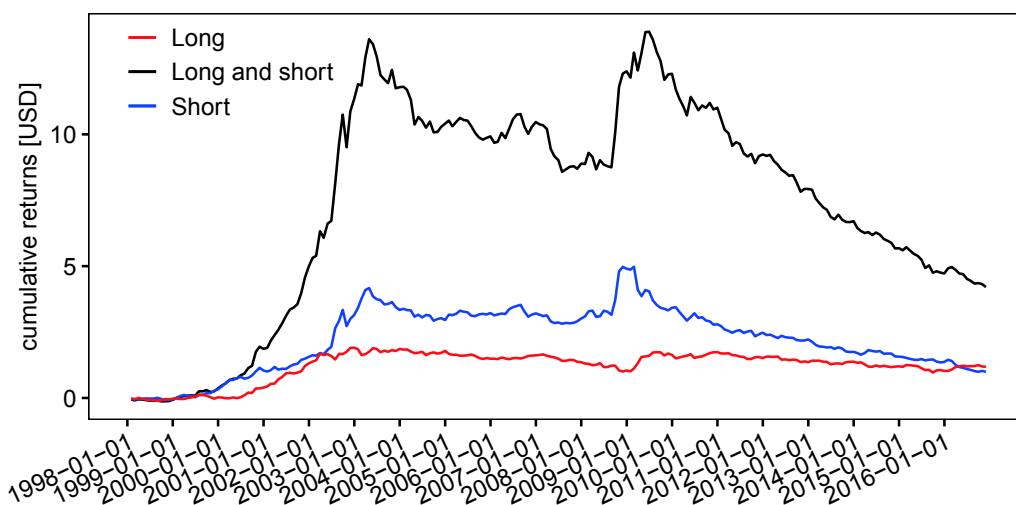

Figure 2. Financial performance over time of the long leg and the short leg

5.1.4. Exposure to common risk factors

To investigate exposure to systematic risk factors, we run three Fama-French type regressions: First, a classic Fama-French 3-factor model (FF3) (Fama and French 1996). The model quantifies if a strategy is subject to general market risk (Market), small minus big capitalization stocks (SMB), and high minus low book-to-market stocks (HML). Second, a variant of the Fama-French 3 Factor model

extended by two additional factors, namely a short-term reversal factor (Reversal) and a momentum factor (Momentum) (Gatev et al. 2006). In line with Clegg and Krauss (2018) we refer to this model as FF3+2. Third, the Fama-French 5-factor model as outlined by Fama and French (2015) (FF5). It augments FF3 by a factor accounting for robust minus weak profitability (RMW) and a second factor capturing the difference in investment behavior between conservative and aggressive stocks (CMA)[5]. Table 5 summarizes the regressions results. A "5" in the label of a factor indicates that this factor belongs to FF5. Standard errors are reported below the corresponding coefficient in parentheses. From our Fama-French analysis we draw the following conclusions: First, for all variants, we find statistically and economically significant excess returns of approximately 0.04 percent per day after transaction costs. Second, the market factor is very close to zero and insignificant for all variants. The latter does not come as a surprise, as the PBD pairs trading strategies follow a long-short design. In addition, the SMB, CMA, and RMW factors are insignificant. Fourth, The HML factor and Reversal factor exhibit a positive loading and are both significant, indicating that the PBD results are partially driven by reversal patterns and by selecting pairs of stocks exhibiting a high book-to-market ratio. The highly significant alphas provide evidence that our PBD based pairs trading results are different from a simple reversal strategy. As expected, the loading of the momentum factor is negative. Fifth, with a value of 0.0010 the FF3+2 has the largest adjusted R^2. The majority of the increase seems to be driven by the Reversal factor.

5.1.5. Return drivers: parameter and industry profiles

In this section, we deep-dive into the average parameter and industry profile difference between all pairs, the 10 percent most profitable and the 1 percent most profitable pairs, to identify influential factors. Table 6 summarizes the results. The most profitable pairs tend to share the following characteristics. First, they exhibit an average level of σ_M of 0.0030, compared to 0.0009 for average pairs. By contrast, the average level of σ_R is at 0.0015 for average pairs, while it is at 0.0004 for the most profitable pairs. Hence, we can cautiously infer that profitable pairs show stronger mean-reversion that can potentially be exploited. Thereby, the ratio $\frac{\sigma_M}{\sigma_W}$ is almost constant at a value of four. On the other hand, we observe that the ratio $\frac{\sigma_M}{\sigma_R}$ is increasing with the profitability of the pairs. Second, in terms of profitability β^{PBD}, R_M^2 and ρ^{PBD} are of minor importance. Third, the industry is an

---

important driver of profitability. A large proportion of the 1 percent most profitable pairs belongs to industries associated with a similar business model such as utilities (35.05 percent) or financials (14.43 percent). With the exception of the oil & gas industry (1.03 percent), these finding coincide with the idea that the probability for the existence of a relationship between companies is higher when they run similar business models (Gatev et al. 2006). Using the same data set, Rende et al. (2019) report similar findings evaluating the industry composition for those pairs where the PBD model yields the best fit.

Table 5

Systematic sources of risk – restricted backtesting. Shows the exposure to common risk factors for the PBD strategy. *** := $p < 0.001$, ** := $p < 0.01$, * := $p < 0.05$

	FF3	FF3+2	FF5
Intercept	0.0004***	0.0004***	0.0004***
	(0.0001)	(0.0001)	(0.0001)
Market	0.0000	0.0000	0.0000
	(0.0001)	(0.0001)	(0.0001)
SMB	0.0001	0.0001	–
	(0.0002)	(0.0002)	–
HML	0.0004*	0.0005**	–
	(0.0002)	(0.0002)	–
Momentum	–	−0.0001	–
	–	(0.0001)	–
Reversal	–	0.0003*	–
	–	(0.0001)	–
SMB5	–	–	0.0000
	–	–	(0.0002)
HML5	–	–	0.0002
	–	–	(0.0002)
RMW5	–	–	−0.0002
	–	–	(0.0003)
CMA5	–	–	0.0004
	–	–	(0.0003)
R^2	0.0010	0.0020	0.0018

Table 5 cont.

Adjusted R^2	0.0004	0.0010	0.0007
Observations	4745.00	4745.00	4745.00
RMSE	0.0076	0.0076	0.0076

Table 6

Parameter and industry profiles. Shows parameter and industry profiles across all pairs, the 10 percent most profitable pairs and the 1 percent most profitable pairs. Thereby, σ_M^{PBD}, σ_W^{PBD} and σ_R^{PBD} denote the variances of the stationary AR(1) process, the noise component and the random walk, respectively. In addition, ρ^{PBD} is the AR(1) coefficient, R_2^M is the proportion of mean-reversion attributable to the stationary AR(1) process and β^{PBD} is the spread coefficient

	All pairs	10% most protable	1% most protable
β^{PBD}	0.2371	0.2591	0.2725
σ_M^{PBD}	0.0009	0.0016	0.0030
σ_W^{PBD}	0.0002	0.0004	0.0007
σ_R^{PBD}	0.0015	0.0008	0.0004
$\dfrac{\sigma_M}{\sigma_W}$	4.5000	4.0000	4.2857
$\dfrac{\sigma_M}{\sigma_R}$	0.6000	2.0000	7.5000
ρ^{PBD}	0.8917	0.8165	0.8050
R_2^M	0.7187	0.7005	0.7162
Basic Materials [%]	13.1813	9.5337	2.0619
Consumer Goods [%]	12.8497	10.3627	7.2165
Consumer Services [%]	12.5078	14.5078	8.2474
Financials [%]	14.2383	12.4352	14.4330
Health Care [%]	7.4611	6.0104	3.0928
Industrials [%]	16.5699	14.3005	8.2474
Oil & Gas [%]	5.7306	5.2850	1.0309
Technology [%]	9.4197	11.1917	18.5567
Telecommunications [%]	0.8290	0.9326	2.0619
Utilities [%]	7.2124	15.4404	35.0515

5.2. Beyond returns: evaluating backtest assumptions

Within this section, we aim to quantify the impact of the execution limitations underlying our backtester, as immediate order execution is often the baseline in the high-frequency pairs trading literature – among others, see Miao (2014) and Stübinger and Bredthauer (2017). For this purpose, we relax our backtesting frictions step-by-step and evaluate the performance. In total, we consider six new strategy variants – three per approach. Table 7 summarizes which execution constraints are active in the different strategies variants. The other parameters remain the same as in the fully constraint backtesting framework – see Table 2. First, we forgo on short-selling fees (PBD 1, PCI 1), i.e., we keep the execution gap and volume constraints. Second, we only account for an execution gap (PBD 2, PCI 2). Third, we do not account for any execution limitations (PBD 3, PCI 3).

Table 7

Summary of constraints in the relaxed backtesting framework. Shows the restrictions imposed in the different strategy variants for the PBD (PBD 1, PBD 2 and PBD 3) and the PCI model (PCI 1, PCI 2 and PCI 3)

Constraint	PBD 1	PBD 2	PBD 3	PCI 1	PCI 2	PCI 3
Execution gap, in periods	1	1	0	1	1	0
Volume constraint (open)	$(v_o^Y > 0$ & $v_o^X > 0)$	None	None	$(v_o^Y > 0$ & $v_o^X > 0)$	None	None
Volume constraint (close)	$(v_c^Y > 0$ & $v_c^X > 0)$	None	None	$(v_c^Y > 0$ & $v_c^X > 0)$	None	None
Short-selling fees [%, p.a.]	0					

Next, we will outline how flexibilizing execution limitations affect risk and returns metrics. Table 8 provides an overview of the performance of the different strategy variants.

Relaxing – short-selling fees: For the PBD model, not accounting for short-selling fees (PBD 1) has a minor effect. With a value of 0.04 percent per day after costs the mean return practically does not change. In general, risk metrics slightly decrease or remain on the same level and return metrics as well as risk-return metrics slightly increase. For example, the annualized Sharpe ratio increases from 0.76 to 0.84. The PCI model (PCI 1) benefits more from not charging the short-selling fee, although the absolute impact is small. As an example, consider the change in the mean return after transaction costs from –0.06 percent

to –0.05 percent per day. The reason for the latter is that the average time pairs are open per round-trip trade is smaller than 1 for both strategies and therefore often solely the minimum fee for one day is charged. Thus, for the PCI model, the impact on trading results is larger because the average number of round-trip trades per pair is much larger. Note that both the average number of round-trip trades per pair, as well as the average time pairs are open per round-trip trade, almost do not change (not shown).

Table 8

Daily and annualized risk-return metrics – relaxed backtesting. Panel A depicts daily return characteristics, panel B depicts risk and panel C annualized risk-return metrics for different strategy variants of the PBD model (PBD 1, PBD 2 and PBD 3), the PCI model (PCI 1, PCI 2 and PCI 3) and the general market (MKT)

	PBD 1	PBD 2	PBD 3	PCI 1	PCI 2	PCI 3	MKT
A							
Mean return	0.0004	0.0016	0.0035	−0.0005	0.0011	0.0027	0.0002
Standard Error (N-W)	0.0001	0.0002	0.0003	0.0002	0.0003	0.0002	0.0002
t-Statistic	3.3471	8.9109	13.2533	−2.6337	4.3191	13.4990	1.5479
Minimum	−0.0583	−0.0585	−0.0599	−0.0880	−0.0914	−0.0832	−0.0903
25% Quantile	−0.0031	−0.0027	−0.0018	−0.0051	−0.0043	−0.0020	−0.0055
Median	0.0001	0.0005	0.0014	−0.0003	0.0003	0.0017	0.0005
75% Quantile	0.0035	0.0044	0.0059	0.0043	0.0055	0.0067	0.0061
Maximum	0.0860	0.0864	0.1303	0.0795	0.0909	0.1045	0.1158
Share > 0	0.5598	0.5756	0.6105	0.4755	0.5172	0.6015	0.5301
Standard dev.	0.0076	0.0088	0.0108	0.0103	0.0119	0.0110	0.0124
Skewness	0.6117	1.5260	2.4645	−0.3129	0.8838	0.8623	−0.0272
Kurtosis	10.4667	11.2909	14.7676	7.5936	9.0918	11.9495	7.7488
B							
Hist. VaR 1%	−0.0312	−0.0245	−0.0146	−0.0449	−0.0408	−0.0434	−0.0515
Hist. CVaR 1%	−0.0281	−0.0267	−0.0238	−0.0432	−0.0420	−0.0391	−0.0483
Hist. VaR 5%	−0.0091	−0.0067	−0.0022	−0.0168	−0.0132	−0.0099	−0.0184
Hist. CVaR 5%	−0.0161	−0.0148	−0.0130	−0.0260	−0.0242	−0.0212	−0.0293
Maximum drawdown	0.4588	0.4119	0.2366	0.9778	0.9345	0.3847	0.5678

Table 8 cont.

	PBD 1	PBD 2	PBD 3	PCI 1	PCI 2	PCI 3	MKT
C							
Return p.a.	0.1004	0.4689	1.3860	−0.1309	0.2946	0.9397	0.0438
Standard dev. p.a.	0.1201	0.1391	0.1709	0.1640	0.1894	0.1739	0.1974
Downside dev p.a.	0.0774	0.0713	0.0611	0.1225	0.1139	0.0954	0.1396
Sharpe ratio p.a.	0.8362	3.3701	8.1089	−0.7984	1.5555	5.4027	0.2217
Sortino ratio p.a.	1.2973	6.5781	22.6789	−1.0686	2.5852	9.8540	0.3135

Relaxing – short-selling fees and volume constraints: Naturally, the question arises as to how to trade in the backtesting engine when there might be zero trading volume available at a point in time. The problem originates from the fact that the econometric approaches applied in trading applications demand equidistant observations – a requirement which is often not fulfilled in high-frequency stock price data sets. To address the equidistance issue prices, are usually forward filled and missing volume information are replaced with zeros. Thus, to account for the latter in the restricted backtesting framework, we impose the constraint that the trading volume for both stocks have to be greater than zero. Solely accounting for an execution gap leads to an increase in daily mean returns after transaction costs of 300 percent resulting in an annualized return of 46.89 percent for the PBD model (PBD 2). A comparison of PCI 1 and PCI 2 yields that daily returns after costs is now positive and highly significant with a value of 0.11 percent. This translates to an annualized return of 29.36 percent. We observe that for both models standard deviation (daily and annualized) and kurtosis slightly increase, but besides that all distributional characteristics as well as risk and return metrics exhibit favorable changes compared to PBD 1 and PCI 1, respectively. Thereby, the PBD model is outperforming the PCI model in every metric.

Relaxing – short-selling fees, volume constraints and execution gap: Finally, we do not consider any execution restrictions. As a consequence annual returns increase to 138.60 percent after costs for the PBD model (PBD 3) and to 93.97 percent for the PCI model (PBD 3). If we contrast PBD 3 with PBD 2 and PCI 3 with PCI 2 we observe significant improvements in terms of return, risk and, risk and return metrics with the exception of standard deviations and kurtosis which is as expected. As in the other cases PBD is superior.

From our analysis we can draw the following conclusions: First, the PBD model is superior to the benchmark model for every strategy variant. Second, short-selling fees have a minor impact on returns. Third, the timing of order execution in light of volume constraints and a general execution gap is crucial

for the success of a strategy. Thereby, the volume constraints contribute more to the relative increase in returns compared to the waiting rule in our setting. The sharp increase of returns after not accounting for an execution gap and volume constraints indicates that the driving force behind large high-frequency delay-zero alpha returns is the bid-ask bounce. The latter finding for the effect of a waiting rule is not new: Among others, Gatev et al. (2006) report decreasing returns after accounting for an execution gap on daily data. Bowen et al. (2010) and Bowen and Hutchinson (2016) report similar patterns for high-frequency data. To the best of our knowledge, no high-frequency pairs trading study quantifies the effect of volume constraints. Thus, not accounting for an execution gap and volume constraints might result in significantly upward-biased returns. As in our case, annual returns increase from about 10 percent after costs to about 138 percent for the PBD model if we do not account for any execution gap.

6. Conclusion

With this manuscript, we have made three contributions to the high-frequency pairs trading literature. First, we apply the persistence-based decomposition model introduced by Rende et al. (2019) to S&P 500 minute-by-minute starting from 1998-01-02 and ending 2016-11-18. The PBD model decomposes a spread into the sum of a random walk (infinite persistence), a stationary $AR(1)$ process (finite persistence) and noise (no persistence) to account for different levels of persistence. Thereby, our backtester accounts for (i) bid-ask bounce typically embedded in high-frequency data (Aït-Sahalia and Yu 2009); (ii) signal processing; (iii) entry and exit volume constraints, i.e., we solely trade if the available trading volume is larger than zero for both stocks, otherwise we postpone the trade until the condition is met; (iv) short-selling fees and; (v) conservative transaction costs of 0.20 percent per full turn per pair. The PBD model yields an annual return after costs of 9.16 percent compared to a value of –14.41 percent for the partial cointegration (PCI) (Clegg and Krauss 2018) and 4.38 percent for a simple buy-and-hold strategy. The PBD model is superior in terms of risk and return metrics. Second, we shed light onto the model mechanics and the development of return over time. We find that the volume constraint is violated in 22.04 percent of all trades, giving a first indication that volume constraints are not negligible. Evaluating the performance over time yields that the PBD model shows an outstanding performance in bear markets such as the financial crisis. A deep-dive analysis points out that during turbulent market phases the short-leg contributes significantly more to the profitability of the strategy than the long leg. While this finding is not new for pairs trading strategies applied

daily (Bowen and Hutchinson 2016), we are the first high-frequency pairs trading study reporting this effect for the U.S. market. Fama-French type analyses provide evidence that our returns can only partially be explained by exposure to systematic risk. Concerning parameter and industry profiles of the most profitable pairs we conclude that the most influential drivers are industry affiliation and a dominance of exploitable mean-reversion over the pronouncedness of the random walk. Specifically, we find that among the most profitable pairs, the share of industries with a similar business model is very high. The third contribution is to quantify the effect of not accounting for execution limitations. If no execution limitations are imposed during the backtesting process, the PBD model yields annual returns after costs of 138.6 percent. In addition, all risk and return metrics improve. These results indicate that intraday pairs trading returns are highly sensitive to execution limitations.

References

[1] Aït-t-Sahalia, Y., Yu, J. (2009) 'High frequency market microstructure noise estimates and liquidity measures', *The Annals of Applied Statistics*, vol. 3, pp. 422–457, doi: 10.1214/08-AOAS200.

[2] Avellaneda, M. and Lee, J.H. (2010) 'Statistical Arbitrage in the U.S. Equities Market', *Quantitative Finance*, vol. 10, pp. 761–782.

[3] Bertram, W.K. (2010) 'Analytic solutions for optimal statistical arbitrage trading', *Physica A: Statistical Mechanics and its Applications*, vol. 389, pp. 2234–2243, [Online], Available: https://www.jstor. org/stable/pdf/2953682. pdf.

[4] Bowen, D., Hutchinson, M.C. and O'Sullivan, N. (2010) 'High Frequency Equity Pairs Trading: Transaction Costs, Speed of Execution and Patterns in Returns', *Journal of Trading*, vol. 5, pp. 31–38.

[5] Bowen, D.A. and Hutchinson, M.C. (2016) 'Pairs trading in the UK equity market: risk and return', *The European Journal of Finance*, vol. 22, pp. 1363–1387.

[6] Caldeira, J.F. and Moura, G.V. (2013) 'Selection of a Portfolio of Pairs Based on Cointegration: A Statistical Arbitrage Strategy', *Brazilian Review of Finance*, pp. 49–80.

[7] Chen, H., Chen, S., Chen, Z. and Li, F. (2017) 'Empirical Investigation of an Equity Pairs Trading Strategy', *Management Science*, vol. 65.

[8] Clegg, M. (2015), 'Modeling Time Series with Both Permanent and Transient Components Using the Partially Autoregressive Model', *SSRN Electronic Journal*, doi: 10.2139/ssrn.2556957.

[9] Clegg, M. and Krauss, C. (2018) 'Pairs trading with partial cointegration', *Quantitative Finance*, vol. 18, pp. 121–138, doi:10.1080/14697688.2017.1 370122.

[10] Clegg, M., Krauss, C. and Rende, J. (2018) 'partialCI: Partial Cointegration', [Online], Available: https://CRAN.R-project.org/package=partialCI.

[11] Cont, R. (2001) 'Empirical properties of asset returns: Stylized facts and statistical issues', *Quantitative Finance*, vol. 1, 223–236.

[12] Do, B. and Faff, R. (2010) 'Does Simple Pairs Trading Still Work?', *Financial Analysts Journal*, vol. 66, pp. 83–95.

[13] Do, B. and Faff, R. (2012) 'Are pairs trading profits robust to trading costs?', *Journal of Financial Research*, vol. 35, pp. 261–287.

[14] Dowle, M. and Srinivasan, A. (2018) 'data.table: Extension of «data.frame»', [Online], Available: https://CRAN.R-project.org/package=data.table.

[15] Dunis, C.L., Giorgioni, G., Laws, J. and Rudy, J. (2010) 'Statistical Arbitrage and High-Frequency Data with an Application to Eurostoxx 50 Equities', Liverpool Business School, Working paper.

[16] Endres, S. and Stübinger, J. (2017) 'Optimal trading strategies for Lévy-driven Ornstein–Uhlenbeck processes', FAU Discussion Papers in Economics, Friedrich-Alexander University Erlangen-Nuremberg, Institute for Economics.

[17] Engle, R.F. and Granger, C.W.J. (1987) 'Co-Integration and error correction: representation, estimation, and testing', *Econometrica*, vol. 55, pp. 251–276, doi:10.2307/1913236.

[18] Fama, E.F. and French, K.R. (1996) 'Multifactor explanations of Asset Pricing Anomalies', *The Journal of Finance*, vol. 51, pp. 55–84.

[19] Fama, E.F. and French, K.R. (2015) 'A five-factor asset pricing model', *Journal of Financial Economics*, vol. 116, pp. 1–22.

[20] Fischer, B. and Pau, G. (2017) 'rhdf5: HDF5 interface to R', [Online], Available: https://www.bioconductor.org/packages/release/bioc/vignettes/rhdf5/inst/doc/rhdf5.html.

[21] Gatev, E., Goetzmann, W.N. and Rouwenhorst, K.G. (2006) 'Pairs Trading: Performance of a Relative-Value Arbitrage Rule', *Review of Financial Studies*, vol. 19, pp. 797–827.

[22] Huck, N. (2015) 'Pairs trading: does volatility timing matter?', *Applied Economics*, vol. 47, pp. 6239–6256.

[23] Izrailev, S. (2014) 'tictoc: Functions for timing R scripts, as well as implementations of Stack and List structures', [Online], Available: https://CRAN.R-project.org/package=tictoc.

[24] Kakushadze, Z. (2016) '101 Formulaic Alphas', Wilmott, vol. 2016, pp. 72–81, doi:10.1002/wilm.10525.

[25] Kim, K. (2011) 'Performance Analysis of Pairs Trading Strategy Utilizing High Frequency Data with an Application to KOSPI 100 Equities', *SSRN Electronic Journal*, doi:10.2139/ ssrn.1913707.

[26] Kishore, V. (2012) 'Optimizing Pairs Trading of US Equities in a High Frequency Setting', [Online], Available: https://quantlabs.net/blog/wp-content/uploads/2015/04/Kishore-Optimize-Pair-Trading-HFT.pdf.

[27] Krauss, C. (2017) 'Statistical arbitrage pairs trading strategies: Review and outlook', *Journal of Economic Surveys*, vol. 31, pp. 513–545.

[28] Krauss, C., Anh Do, X. and Huck, N. (2017) 'Deep neural networks, gradient-boosted trees, random forests: statistical arbitrage on the S & P 500', *European Journal of Operational Research*, vol. 259, pp. 689–702.

[29] Liu, B., Chang, L.B. and Geman, H. (2017) 'Intraday pairs trading strategies on high frequency data: the case of oil companies', *Quantitative Finance*, vol. 17, pp. 87–100, doi: 10.1080/14697688. 2016.1184304.

[30] Luethi, D., Erb, P. and Otziger, S. (2018) 'FKF: Fast Kalman Filter', [Online], Available: https://CRAN.R-project.org/package=FKF.

[31] Miao, G.J. (2014) 'High Frequency and Dynamic Pairs Trading Based on Statistical Arbitrage Using a Two-Stage Correlation and Cointegration Approach', *International Journal of Economics and Finance*, vol. 6, pp. 96–110.

[32] Mikkelsen, A., Kjærland, F. (2018) 'High-frequency Pairs Trading on a Small Stock Exchange', *International Journal of Economics and Financial Issues*, vol. 8, pp. 78–88, [Online], Available: https://brage.bibsys.no/xmlui/bitstream/11250/ 2574680/5/Mikkelsen.pdf.

[33] Nath, P. (2003) 'High frequency pairs trading with US treasury securities: Risks and rewards for hedge funds', [Online], Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=565441.

[34] Papadakis, G. and Wysocki, P. (2007) 'Pairs Trading and Accounting Information'. Boston University and MIT Working Paper.

[35] Peterson, B.G., Carl, P. (2018) 'Performance Analytics: econometric tools for performance and risk analysis', [Online], Available: https://CRAN.R-project.org/package=PerformanceAnalytics.

[36] Poterba, J.M. and Summers, L.H. (1988) 'Mean reversion in stock prices', *Journal of Financial Economics*, vol. 22, pp. 27–59, doi:10.1016/0304-405X(88)90021-9.

[37] QuantQuote (2012) 'Minute Market Data', [Online], Available: https://quantquote.com/docs/QuantQuote_Minute.pdf.

[38] QuantQuote (2016) 'QuantQuote market data and software', [Online], Available: URL: https://www.quantquote.com/.

[39] R Development Core Team (2018) R: A Language and Environment for Statistical Computing. Vienna, Austria.

[40] Rad, H., Low, R.K.Y. and Faff, R. (2016). The Profitability of Pairs Trading Strategies: Distance, Cointegration and Copula Methods, *Quantitative Finance*, vol. 16, pp. 1541–1558.

[41] Rende, J., Krauss, C. and Clegg, M. (2019) 'The persistence-based decomposition (PBD) time series model: theory and empirical application', vol. 01/2019 of FAU Discussion Papers in Economics.

[42] Ryan, J.A. and Ulrich, J.M. (2018) 'xts: Extensible Time Series', [Online], Available: https://CRAN.R-project.org/package=xts.

[43] Schwarz, G. (1978) 'Estimating the Dimension of a Model', The Annals of Statistics, vol. 6, pp. 461–464, [Online], Available: https://projecteuclid.org/ download/pdf_1/euclid.aos/1176344136, doi:10. 1214/aos/1176344136.

[44] Sharpe, W.F. (1994) 'The Sharpe Ratio', *The Journal of Portfolio Management*, vol. 21, pp. 49–58, doi:10.3905/jpm.1994.409501.

[45] S&P 500 Dow Jones Indices (2015) 'Equity S&P 500', [Online], Available: http://us.spindices.com/ indices/equity/sp-500.

[46] Stübinger, J. and Bredthauer, J. (2017) 'Statistical Arbitrage Pairs Trading with High-frequency Data', *International Journal of Economics and Financial Issues*, vol. 7, pp. 650–662, [Online], Available: https://econjournals.com/ index.php/ijefi/article/download/5127/pdf.

[47] Stübinger, J. and Endres, S. (2018) 'Pairs trading with a mean-reverting jump-diffusion model on high-frequency data', *Quantitative Finance*, vol. 18, pp. 1735–1751.

[48] Stübinger, J., Mangold, B. and Krauss, C. (2018) 'Statistical arbitrage with vine copulas', *Quantitative Finance*, vol. 18, pp. 1831–1849, doi:10.1080/ 14697688.2018.1438642.

[49] The HDF Group (2010) 'Hierarchical data format version 5', [Online], Available: http://www.hdfgroup.org/HDF5.

[50] Vidyamurthy, G. (2004) Pairs Trading: Quantitative Methods and Analysis. Hoboken: Wiley Finance. John Wiley & Sons Inc.