Discussion Papers in Economics

No. 07/2018

A flexible regime switching model with pairs trading application to the S&P 500 high-frequency stock returns

Sylvia Endres
University of Erlangen-Nürnberg

Johannes Stübinger
University of Erlangen-Nürnberg

ISSN 1867-6707

Friedrich-Alexander-Universität Erlangen-Nürnberg
Institute for Economics
https://www.iwf.rw.fau.de/research/iwf-discussion-paper-series/

A flexible regime switching model with pairs trading application to the S&P 500 high-frequency stock returns

Sylvia Endres[a,1], Johannes Stübinger[a,1]

[a] *University of Erlangen–Nürnberg, Department of Statistics and Econometrics, Lange Gasse 20, 90403 Nürnberg, Germany*

Wednesday 16th May, 2018

Abstract

This paper develops the regime classification algorithm and applies it within a fully-fledged pairs trading framework on minute–by–minute data of the S&P 500 constituents from 1998 to 2015. Specifically, the highly flexible algorithm automatically determines the number of regimes for any stochastic process and provides a complete set of parameter estimations. We demonstrate its performance in a simulation study — the algorithm achieves promising results for the general class of Lévy-driven Ornstein–Uhlenbeck processes with regime switches. In our empirical back-testing study, we apply our regime classification algorithm to propose a high-frequency pair selection and trading strategy. The results show statistically and economically significant returns with an annualized Sharpe ratio of 3.92 after transaction costs — results remain stable even in recent years. We compare our strategy with existing quantitative trading frameworks and find its results to be superior in terms of risk and return characteristics. The algorithm takes full advantage of its flexibility and identifies various regime patterns over time that are well-documented in the literature.

Keywords: Finance; Pairs trading; Statistical arbitrage; Markov regime switching; Lévy-driven Ornstein–Uhlenbeck process; High-frequency data

Email addresses: sylvia.endres@fau.de (Sylvia Endres), johannes.stuebinger@fau.de (Johannes Stübinger)

[1]The authors have benefited from many helpful discussions with Ingo Klein and Julian Knoll.

1. Introduction

Interest in statistical arbitrage opportunities in the equity world grows steadily. One of the earliest strategies in this context is pairs trading, dating back to 1980s (Vidyamurthy, 2004). The objective of pairs trading is to find pairs of stocks that are historically closely related. Then, in the event of temporary imbalances, profit opportunities are exploited by short selling the relatively overvalued share and by buying the relatively undervalued share. If the historical balance is restored, the trades are reversed and a positive benefit is gained.

Since the first acamedic research study of Gatev et al. (1999, 2006) various quantitative methods are used to select and trade stocks pairs. The most important studies are given by Elliott et al. (2005), Gatev et al. (2006), Avellaneda and Lee (2010), Do and Faff (2010), Cummins and Bucca (2012), Göncü and Akyildirim (2016b), Rad et al. (2016), Liu et al. (2017), and Stübinger and Endres (2018). Recently, regime switching models become increasingly important because they allow structural changes to be taken into account (Bock and Mestel, 2009). Specifically, Markov regime switching models are used to distinguish temporary from permanent spread deviations.

Surprisingly, only a few academic research studies on continuous-time pairs trading apply regime switching models to incorporate different states of spreads.[2] The research is limited to Yang et al. (2016), Altay et al. (2017) and Bai and Wu (2018). Yang et al. (2016) combine the Markov regime switching model and the Ornstein–Uhlenbeck (OU) process assuming that the spread is always divided into two states. Altay et al. (2017) model the spread by a Gaussian mean-reverting process whose drift rate is modulated by an unobservable continuous-time finite state Markov chain. The authors provide a numerical analysis for a two-state Markov chain. Bai and Wu (2018) introduce a regime switching OU model and give a closed-form expression for the pairs trading value function. Their numerical analysis is executed for one-state and two-state regime switching models. In summary, there exist two major research gaps in the area of continuous-time pairs trading based on regime switching models. First, there is no research study that is able to allow a flexible number of regimes. Second, there is no research study which takes into account a model that incorporates fat

[2] In the following, the terms regime and state are used synonymously.

tails and jumps, e.g., a Lévy-driven OU model within the individual regimes.

Our manuscript complements existing research in several respects. First, we develop the regime classification algorithm that automatically determines the number of regimes, for an arbitrary process in each state. Its performance is shown by a simulation study for the general class of Lévy-driven OU processes that switch between different regimes. Second, we propose a high-frequency statistical arbitrage strategy based on a regime switching model accounting for similarities and differences between volatility regimes in different periods of time — this links well with empirically observed volatility patterns in financial markets (Cai 1994, Hamilton and Susmel 1994, Dahlquist and Gray 2000, Andersen et al. 2001, Bouchaud et al. 2001, Ang and Bekaert 2002, Nath and Dalvi 2004, Bollerslev et al. 2006, Göncü et al. 2016, Chang 2009, Chevallier and Goutte 2017, Liu et al. 2018). More precisely, we use the introduced algorithm to identify the optimal number of regimes. In each regime, a Lévy-driven OU process is applied, reflecting leptokurtosis and discontinuities — both empirical features of return series (Bertram 2009, Aït-Sahalia and Jacod 2014, Göncü and Akyildirim 2016b, Endres and Stübinger 2017, Kou et al. 2017). The optimal pairs for the trading period are selected based on mean-reversion speed, volatility, and jump behavior. Third, we conduct a large-scale empirical study of the introduced statistical arbitrage framework on the S&P 500 stocks based on minute–by–minute data from January 1998 to December 2015. The value-add of our strategy is demonstrated by comparing it with established quantitative trading strategies. We find that our strategy based on the flexible regime switching model achieves an annualized return of 93.73 percent and a Sharpe ratio of 3.92 after transaction costs. Hereby, returns are statistically significant and withstand various robustness checks. The results are clearly superior to the benchmark strategies ranging between 3.93 percent for a naive buy–and–hold strategy of the S&P 500 index and 47.24 percent for a strategy based on the regime switching model with a classic OU process in each regime. In stark contrast to the benchmarks, our strategy achieves positive returns after transaction costs in recent years. Fourth, we analyze the number of regimes discovered by our algorithm. We vindicate the value-add of the algorithm's flexibility compared to approaches where the regime number is determined in advance. The algorithm is able to detect regime shifts in volatility that are well-documented in literature.

The rest of the paper is organized as follows. Section 2 describes the underlying theoretical framework. In section 3, we develop the regime classification algorithm and study its performance by means of a Monte Carlo simulation. Section 4 provides our high-frequency back-testing study. In section 5, the results and key findings are discussed. Finally, section 6 summarizes our paper and proposes directions for further research.

2. Methodology

Pairs trading strategies have the objective of finding pairs of stocks possessing a long-term equilibrium relationship — this characteristic is directly associated with a mean-reverting spread (see Elliott et al. 2005, Do et al. 2006, Baronyan et al. 2010, and Bertram 2010). In this study, $\{S_A(t)\}_{t\geq 0}$ and $\{S_B(t)\}_{t\geq 0}$ define the price series of the stocks A and B. The corresponding spread X_t at time t is specified as

$$X_t = \ln(S_A(t)/S_A(0)) - \ln(S_B(t)/S_B(0)),\ t \geq 0$$

and switches between r different regimes ($r \in \mathbb{N}$) in the considered market. This regime switching behavior is described by a continuous-time Markov chain $\{Z_t\}_{t\geq 0}$, where the random variable Z_t denotes the state of the process at time t. In each regime i ($i \in \{1, \ldots, r\}$), the corresponding part of the spread follows unique dynamics necessitating r stochastic differential equations for each $\{X_t\}_{t\geq 0}$. To be more specific, we model the spread characteristics in each regime i by the general class of Lévy-driven OU processes. Mathematically, we specify $\{X_t\}_{t\geq 0}$ by the following stochastic differential equations:

$$dX_t = \begin{cases} \theta_1 \left(\mu_1(t) - X_t\right) dt + dL_{1,t}, & \text{for } Z_t = 1 \\ \vdots & \vdots \\ \theta_r \left(\mu_r(t) - X_t\right) dt + dL_{r,t}, & \text{for } Z_t = r \end{cases} \tag{1}$$

with $X_0 = x$. For each regime $Z_t = i$, the process' parameters are the mean-reversion speed $\theta_i \in \mathbb{R}$ and the time-dependent mean-reversion level $\mu_i(t) \in \mathbb{R}$. The general Lévy process $\{L_{i,t}\}_{t\geq 0}$ for regime $Z_t = i$ is specified by the Lévy–Khintchine characteristics (b_i, σ_i^2, ν_i) (see Mai 2012). The triplet implies a decomposition of $\{L_{i,t}\}_{t\geq 0}$ into a linear drift term with slope b_i, a Brownian motion process $\{W_t\}_{t\geq 0}$ with variance $\text{Var}[W_t] = \sigma_i^2 t$ and diffusion

parameter $\sigma_i \in \mathbb{R}^+$, and a jump component described by the Lévy measure ν_i. By the Lévy–Itô decomposition and in case of finite jump activity, the driving Lévy process of $\{X_t\}_{t\geq 0}$ can be written as

$$L_{Z_t,t} = \begin{cases} \sigma_1 W_t + \sum_{j=1}^{N_{1,t}} \xi_{1,j}, & \text{for } Z_t = 1 \\ \vdots & \vdots \\ \sigma_r W_t + \sum_{j=1}^{N_{r,t}} \xi_{r,j}, & \text{for } Z_t = r \end{cases} \tag{2}$$

for standard Brownian motion $\{W_t\}_{t\geq 0}$, Poisson process $\{N_{i,t}\}_{t\geq 0}$ with rate $\lambda_i \in \mathbb{R}_0^+$, and jump sizes $\{\xi_{1,1}, \ldots, \xi_{1,N_{1,t}}, \ldots, \xi_{r,1}, \ldots, \xi_{r,N_{r,t}}\}$. The stochastic processes $\{W_t\}_{t\geq 0}$, $\{N_{i,t}\}_{t\geq 0}$, and the random variables $\{\xi_{1,1}, \ldots, \xi_{1,N_{1,t}}, \ldots, \xi_{r,1}, \ldots, \xi_{r,N_{r,t}}\}$ are independent.

In each regime, the process follows the dynamics of an OU process driven by a general Lévy process without switching. Asymptotically efficient estimators for the mean-reversion rate of these processes are constructed from a discretization of the time-continuous maximum likelihood estimators according to Mai (2012, 2014)[3]. We recover the continuous part of $\{X_t\}_{t\geq 0}$ in the high-frequency limit via jump filtering. Based on discrete variables $X_{t_1}, ..., X_{t_n}$, the following estimator is obtained:

$$\hat{\theta}_n = \frac{\sum_{i=0}^{n-1} (\mu_{t_i} - X_{t_i}) \Delta_i X \mathbb{1}_{\{|\Delta_i X| \leq \nu_n\}}}{\sum_{i=0}^{n-1} (\mu_{t_i} - X_{t_i})^2 (t_{i+1} - t_i)}, \tag{3}$$

where $\Delta_i X = X_{t_{i+1}} - X_{t_i}$ and $\Delta_n = \max_{1\leq i\leq n-1} \{|\Delta_i t|\}$, $\Delta_i t = t_{i+1} - t_i$. The continuous part and the jump part are distinguished because they exhibit a different order of magnitude on a small time scale. Increments larger than the threshold $\nu_n = \Delta_n^\beta$, $\beta \in (0, 1/2)$ are neglected.

3. Regime classification algorithm

This section introduces a new approach, called the "regime classification algorithm", to estimate a full regime switching framework (see subsection 3.1). Unlike existing literature, our flexible algorithm is able to automatically determine the number of regimes in a time series. Roughly speaking, we gradually increase the number of regimes r as long as the relative quality of statistical models improves. Finally, subsection 3.2 validates the regime classification algorithm by means of Monte Carlo simulation.

[3] In Mai (2014) the mean-reversion rate is denoted as drift parameter.

3.1. Theoretical concept

The algorithm starts with the basic case $r = 1$, i.e., we presume one regime. For a given time series $\boldsymbol{x} \in \mathbb{R}^n$, we estimate all model parameters with the function $fit.models$ and save them in vector $\boldsymbol{p}$. The underlying estimation procedure in $fit.models$ is chosen in accordance with the stylized facts of the time series. In our research study, we apply a Lévy-driven OU process as outlined in section 2 for each regime. Then, the conditional least squares error

$$CLS_r = \sum_{i=1}^{n-1} (X_{i+1} - E_{\boldsymbol{p}}[X_{i+1}|X_i, ..., X_1])^2$$

is calculated based on the one-step ahead prediction (Pinson et al. 2008). Based on CLS_r, the Bayesian information criterion is computed as

$$BIC_r = n \ln\left(\frac{CLS_r}{n}\right) + m \ln(n), \tag{4}$$

where m is the number of estimated model parameters. In the sense of Mota and Esquivel (2016), we use the Bayesian information criterion as a tool for model selection in our regime switching framework. The criterion manages the trade-off between the goodness-of-fit and the complexity of the model. The smaller BIC_r is, the better the model fits the data. For $r = 1$ we save BIC_r as BIC_1^*.

Next, we increase the number of regimes r step–by–step as long as the Bayesian information criterion decreases. For each considered r, the switching between the r regimes depends on the crossing of the thresholds (Esquivel and Mota 2014, Mota and Esquivel 2014, Bai and Wu 2018) by the rolling spread's volatility trajectories $\boldsymbol{v}$ — in this way, different volatility regimes are identified (Bee and Gatti 2015). We determine the most suitable thresholds $\boldsymbol{c} = (c_1, c_2, \ldots, c_{r-1})$ based on the Bayesian information criterion. Without loss of generality, we assume $c_1 < c_2 < \cdots < c_{r-1}$. Following Mota and Esquivel (2014), we require that at least each regime includes 15 percent of the total observations. A complete grid-search across all possible threshold combinations is avoided by implementing a smart procedure. The functions $start.grid$ and $smart.grid$ control the threshold combinations that should be tested. For each combination $\boldsymbol{c}$ recommended by the smart grid-search, the data is classified into r subsets $\boldsymbol{S_1} = \boldsymbol{x}[\boldsymbol{v} < c_1], \boldsymbol{S_2} = \boldsymbol{x}[\boldsymbol{v} < c_2 \ \& \ \boldsymbol{v} \geq c_1], ..., \boldsymbol{S_r} = \boldsymbol{x}[\boldsymbol{v} \geq c_{r-1}]$ by the function $classify.data$. Using the observations in each of the regimes, conditional estimators

$\boldsymbol{p} = \{\boldsymbol{p_1}, ..., \boldsymbol{p_r}\}$ are computed by $fit.models$, applying the estimators for the parameters in the simple process context. Then, we calculate the corresponding Bayesian information criterion as in equation (4) and store it as $BIC.local$. The smart procedure determines the thresholds to be considered next, always optimizing one threshold of the previous combination while keeping the others fixed. Generally, all discrete data points of $\boldsymbol{x} \in \mathbb{R}^n$ are considered in the pool of possible thresholds. This procedure is carried out for different threshold combinations until an optimum with lowest $BIC.local$ is found, stored as BIC_r^*. The thresholds $\boldsymbol{c}_{best}$ corresponding to BIC_r^* determine the best setting for r regimes.

If BIC_r^* is smaller than BIC_{r-1}^*, then r is increased by one. Otherwise, $r-1$ is considered as the optimal number of regimes. In the end, the algorithm provides an optimal number of regimes and the corresponding parameter estimations.

Algorithm 1 Regime classification algorithm (1/2)

Input: Data series $\boldsymbol{x} \in \mathbb{R}^n$

Output: Classification of $\boldsymbol{x}$ into $r \in \mathbb{N}$ regimes with regime-thresholds $\boldsymbol{c} = (c_1, \ldots, c_{r-1})$ and set of estimated model parameter sets $\boldsymbol{p} = \{\boldsymbol{p_1}, ..., \boldsymbol{p_r}\}$.

Functions:

$fit.models(\boldsymbol{S_1}, ..., \boldsymbol{S_r})$: Function returning estimated model parameter set $\boldsymbol{p} = \{\boldsymbol{p_1}, ..., \boldsymbol{p_r}\}$ based on data sets $\boldsymbol{S_1}, ..., \boldsymbol{S_r}$.

$calc.BIC(\boldsymbol{S_1}, ..., \boldsymbol{S_r}, \boldsymbol{p})$: Function returning BIC for data sets $\boldsymbol{S_1}, ..., \boldsymbol{S_r}$, and corresponding model parameter set $\boldsymbol{p} = \{\boldsymbol{p_1}, ..., \boldsymbol{p_r}\}$.

$start.grid(\boldsymbol{x}, r)$: Function returning set $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{z \times (r-1)}$ of z possible starting vectors, each including $r-1$ thresholds, given the data series $\boldsymbol{x}$.

$num.start(\boldsymbol{x}, r)$: Function returning the number l of possible threshold combination sets $\boldsymbol{\mathcal{I}}$ for r regimes and data series $\boldsymbol{x}$.

$classify.data(\boldsymbol{x}, r, \boldsymbol{c})$: Function returning r subsets $\boldsymbol{S} = \{\boldsymbol{S_1}, ..., \boldsymbol{S_r}\}$ of data series $\boldsymbol{x}$, separated by thresholds $\boldsymbol{c} \in \mathbb{R}^{r-1}$.

$smart.grid(\boldsymbol{x}, \boldsymbol{\mathcal{I}}, BIC_{\boldsymbol{\mathcal{I}}})$: Function returning next threshold set $\boldsymbol{\mathcal{I}}_{next} \in \mathbb{R}^{z \times (r-1)}$ for previous set $\boldsymbol{\mathcal{I}}$ with corresponding $BIC_{\boldsymbol{\mathcal{I}}}$.

Algorithm 1 Regime classification algorithm (2/2)

Algorithm:

$\boldsymbol{p} \leftarrow fit.models(\boldsymbol{x})$; $BIC_1^* \leftarrow calc.BIC(\boldsymbol{x}, \boldsymbol{p})$;

$r \leftarrow 2$;

loop

$\boldsymbol{\mathcal{I}} \leftarrow start.grid(\boldsymbol{x}, r)$; $l \leftarrow num.start(\boldsymbol{x}, r)$;

$a \leftarrow 1$; $BIC[a] \leftarrow +\infty$; $\boldsymbol{c}_b \leftarrow \boldsymbol{c}_{best} \leftarrow (0, ..., 0) \in \mathbb{R}^{r-1}$;

loop

$i \leftarrow 1$;

loop

$\boldsymbol{c} \leftarrow \boldsymbol{\mathcal{I}}[i]$;

$\boldsymbol{S_1}, ..., \boldsymbol{S_r} \leftarrow classify.data(\boldsymbol{x}, r, \boldsymbol{c})$;

$\boldsymbol{p} \leftarrow fit.models(\boldsymbol{S_1}, ..., \boldsymbol{S_r})$;

$\boldsymbol{BIC.local}[i] \leftarrow calc.BIC(\boldsymbol{S_1}, ..., \boldsymbol{S_r}, \boldsymbol{p})$;

if $i = z$ **then break**;

$i \leftarrow i + 1$;

end loop

$a \leftarrow a + 1$; $\boldsymbol{c}_{b.prev} \leftarrow \boldsymbol{c}_b$;

$\boldsymbol{BIC}[a] \leftarrow \min(\boldsymbol{BIC.local})$; $\boldsymbol{c}_b \leftarrow \boldsymbol{\mathcal{I}}[argmin(\boldsymbol{BIC.local})]$;

if $\boldsymbol{BIC}[a-1] < \boldsymbol{BIC}[a]$ **or** $a - 1 \leftarrow l$ **then break**;

$\boldsymbol{\mathcal{I}} \leftarrow smart.grid(\boldsymbol{x}, \boldsymbol{\mathcal{I}}, \boldsymbol{BIC.local})$;

end loop

$\boldsymbol{c}_{best.prev} \leftarrow \boldsymbol{c}_{best}$; $\boldsymbol{c}_{best} \leftarrow \boldsymbol{c}_{b.prev}$;

$BIC_r^* \leftarrow \boldsymbol{BIC}[a-1]$;

if $BIC_{r-1}^* < BIC_r^*$ **then break**;

$r \leftarrow r + 1$;

end loop

$\boldsymbol{S_1}, ..., \boldsymbol{S_{r-1}} \leftarrow classify.data(\boldsymbol{x}, r-1, \boldsymbol{c}_{best.prev})$;

$\boldsymbol{p} \leftarrow fit.models(\boldsymbol{S_1}, ..., \boldsymbol{S_{r-1}})$;

$\boldsymbol{BIC} \leftarrow calc.BIC(\boldsymbol{S_1}, ..., \boldsymbol{S_{r-1}}, \boldsymbol{p})$;

3.2. Simulation study

In this section, we study the performance of our regime classification algorithm by means of a Monte Carlo simulation. The finding of this study is twofold. On the one hand, we show that the algorithm is applicable for Lévy-driven OU processes with regime switches and produces favorable results. On the other hand, we demonstrate the outperformance of our jump-based estimation (section 2) compared to estimation of a classic OU process.

In our simulation study, we opt for processes $\{X_t\}_{t\geq 0}$ with r regimes (see section 2). In regime i ($i \in \{1, ..., r\}$), the dynamics evolve according to a Lévy-driven OU processes with parameter set $\Theta_i = (\mu_i, \theta_i, \sigma_i)$, finite jump intensity $\lambda \in \mathbb{R}_0^+$, and i.i.d. jump heights following the $\mathcal{N}(0, \sigma_J^2)$-distribution (see Mai 2012). In N Monte Carlo iterations, we generate the series $X_{t_1}, ..., X_{t_n}$ and estimate a full regime switching framework according to algorithm 1.

Table 1 contains simulation results for $N = 1000$ Monte Carlo iterations and sample size $n = 10000$ according to Masuda (2010) and Uehara (2017). For $r \in \{1, 2, 3\}$ the parameter sets Θ_1, Θ_2 and Θ_3 reflect the dynamics in each of the regimes. We choose $\Theta_1 = (3, 1, 2)$ as proposed in Iacus (2008) as well as $\Theta_2 = (3, 2, 1)$ and $\Theta_3 = (2, 3, 1)$, which include the elements of Θ_1 in permuted order. The intensity $\lambda = 0.004$ and the jump size volatility $\sigma_J \in \{2\sigma_i, 2.5\sigma_i, 3\sigma_i\}$ as a multiple of the volatility of the regular innovations are selected in accordance with Mai (2012), Jondeau et al. (2015) and Fischer et al. (2018). Intensity $\lambda = 0$ depicts a process without any jumps. We use a simulation time step of $\Delta_n = 1$ and a jump threshold of Δ_n^β with $\beta = 0.3$ (according to Mai 2012). For varying $r \in \{1, 2, 3\}$, $\lambda \in \{0, 0.004\}$ and $\sigma_J \in \{2\sigma_i, 2.5\sigma_i, 3\sigma_i\}$, algorithm 1 returns an estimated number of regimes $\hat{r}$. In each regime, we estimate either a Lévy-driven OU process according to section 2 (left side) or a classic OU process neglecting any jumps (right side). In table 1 hit rates and error rates of both variants are presented — boxes with gray background display cases where the true number of regimes r coincides with the estimated number $\hat{r}$. All scenarios based on estimation of a Lévy-driven OU process produce hit rates, i.e., shares of correctly detected regime numbers, that are greater than 50 percent. As expected, best results are achieved for $r = 1$ and $\lambda = 0$, i.e., zero jumps — the estimated number of regimes $\hat{r}$ matches the true number r in 94 percent of the cases. With increasing jump activity as well as increasing number of regimes r the hit rate worsens in most cases. But still, the hit rate of 57 percent

is satisfactory even for $r = 3$ and jump size variance of $3\sigma_i^2$. Next, we compare the results based on estimation of a Lévy-driven OU process with those of a classic OU process to investigate the influence of jumps and their non-consideration. Hit rates of a classic OU process range between 39 and 76 percent in case of jump activity. Hereby, the hit rates of a Lévy-driven OU process outperform the classic OU process in all cases because disregarding of jumps leads to blurring of regime switches on the on hand and distortion on the other. We may conclude that the regime classification algorithm works properly for a Lévy-driven OU process, even for higher number of regimes and higher jump activity.

		Lévy-driven OU process				Classic OU process			
	$r \setminus \hat{r}$	1	2	3	4	1	2	3	4
$\lambda = 0.000, \sigma_J = 0.0\sigma_i$	1	0.94	0.06	0.00	0.00	0.94	0.06	0.00	0.00
	2	0.13	0.87	0.00	0.00	0.12	0.88	0.00	0.00
	3	0.00	0.18	0.82	0.00	0.00	0.09	0.85	0.06
$\lambda = 0.004, \sigma_J = 2.0\sigma_i$	1	0.91	0.09	0.00	0.00	0.76	0.24	0.00	0.00
	2	0.13	0.85	0.02	0.00	0.24	0.72	0.04	0.00
	3	0.00	0.32	0.66	0.02	0.00	0.14	0.48	0.38
$\lambda = 0.004, \sigma_J = 2.5\sigma_i$	1	0.82	0.18	0.00	0.00	0.60	0.40	0.00	0.00
	2	0.00	0.79	0.20	0.01	0.04	0.56	0.40	0.00
	3	0.00	0.33	0.63	0.04	0.00	0.16	0.46	0.38
$\lambda = 0.004, \sigma_J = 3.0\sigma_i$	1	0.62	0.38	0.00	0.00	0.39	0.61	0.00	0.00
	2	0.00	0.67	0.32	0.01	0.00	0.57	0.43	0.00
	3	0.00	0.38	0.57	0.05	0.00	0.19	0.50	0.31

Table 1: Hit and error rates of estimation of Lévy-driven OU process (left) and classic OU process (right) given a Lévy-driven OU process with r regimes, jump intensity λ and jump size variance σ_J^2. The number of regimes estimated by algorithm 1 is denoted by $\hat{r}$.

Table 2 varies the path length n for the case of medium jump activity and evaluates the effects on the mean square error (MSE). Specifically, the MSEs for the estimations of μ, θ and σ for each regime are depicted for small ($n = 100$), medium ($n = 1000$) and large ($n = 10000$) samples. Only those scenarios where algorithm 1 detects the same number of regimes for $n = 100, 1000, 10000$ are included to make the results comparable. We observe that the MSE decreases for increasing n — a desirable property for any statistician.

Concluding, the regime classification algorithm including the smart procedure that avoids

a full grid-search shows strong performance for Lévy-driven OU models in light of robustness and feasibility.

n	r	λ	σ_J	$\text{MSE}_\mu^{R_1}$	$\text{MSE}_\theta^{R_1}$	$\text{MSE}_\sigma^{R_1}$	$\text{MSE}_\mu^{R_2}$	$\text{MSE}_\theta^{R_2}$	$\text{MSE}_\sigma^{R_2}$	$\text{MSE}_\mu^{R_3}$	$\text{MSE}_\theta^{R_3}$	$\text{MSE}_\sigma^{R_3}$
100	1	0.004	2.5 σ_i	0.0191	0.1071	0.1009	-	-	-	-	-	-
1000	1	0.004	2.5 σ_i	0.0052	0.0352	0.0242	-	-	-	-	-	-
10000	1	0.004	2.5 σ_i	0.0005	0.0181	0.0064	-	-	-	-	-	-
100	2	0.004	2.5 σ_i	0.0279	11.5920	0.9940	0.1549	10.9816	3.7600	-	-	-
1000	2	0.004	2.5 σ_i	0.0021	0.9286	0.8174	0.0145	0.7631	1.2274	-	-	-
10000	2	0.004	2.5 σ_i	0.0001	0.3835	0.7080	0.0016	0.7131	1.1651	-	-	-
100	3	0.004	2.5 σ_i	0.9279	31.7789	1.6923	0.7755	22.6386	1.4892	0.9613	17.5025	1.3634
1000	3	0.004	2.5 σ_i	0.8173	8.1719	1.6207	0.3003	3.6506	0.3470	0.9561	3.9425	1.0300
10000	3	0.004	2.5 σ_i	0.7761	0.4534	1.5280	0.2623	2.0942	0.1084	0.8301	3.7474	0.5251

Table 2: Mean-square error for $r \in \{1, 2, 3\}$, $n \in \{100, 1000, 10000\}$ and a Lévy-driven OU process in each regime with intensity λ and jump size variance σ_J^2.

4. Study design

In accordance with the high-frequency research studies of Liu et al. (2017) and Stübinger and Endres (2018), our pairs trading strategy consists of a 30-day formation and a 5-day trading period. To reduce the influence of choosing starting points in the formation period and the influence of pairs with excessively high or low returns in the trading period, overlapping intervals are considered (Yang et al. 2016). A moving window of 35 days length is always shifted by one day until the entire data set is covered (see figure 1).

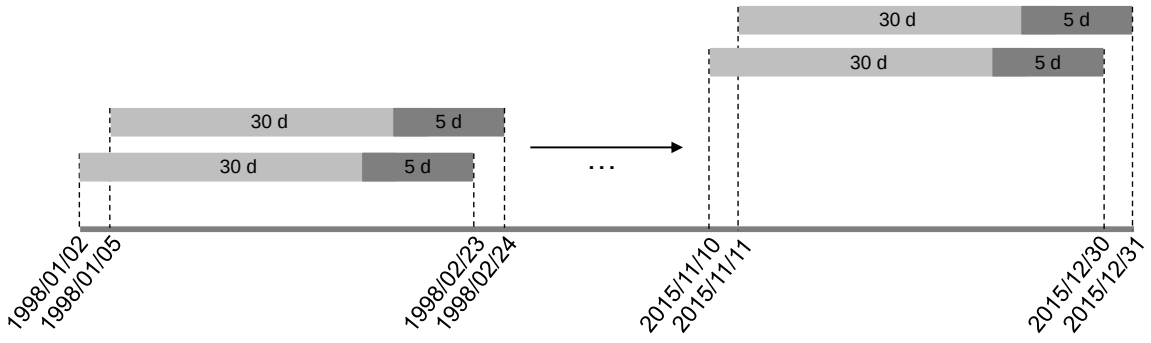

Figure 1: The period of 35 days length, consisting of 30 days formation and 5 days trading, is always shifted by one day. There are 4494 overlapping intervals of 35 days length from 1998/01/02 until 2015/12/31.

4.1. Data

For our back-testing framework, we choose the constituents of the S&P 500 index, which is widely regarded as the best single gauge of large-cap U.S. equities. It includes the top 500 companies of the U.S. economy and captures approximately 80 percent coverage of available market capitalization (S&P Dow Jones Indices 2015). Thus, the S&P 500 index serves as a crucial test for any potential stock market inefficiency.

We download S&P 500 minute–by–minute prices for a period of 18 years from January 1998 until December 2015 from QuantQuote (QuantQuote 2016). Our data base covers 4529 trading days and for each trading day, quotes from 9.30 am until 4.00 pm are available, i.e., we have 391 data points per stock per day. The data is adjusted for stock splits, corporate events, and dividend payments.

For eliminating survivor bias from our data, we follow Krauss and Stübinger (2017) and proceed as follows. First, we obtain daily constituent lists for all S&P 500 stocks from January 1998 to December 2015. These lists are aggregated into a binary matrix, indicating for each day and each stock whether it is an index constituent in the respective day or not. Second, we gather the minute–by–minute prices for all these index constituents during the corresponding time frames. In this way, we are able to replicate the S&P 500 index at any point in time between January 1998 and December 2015.

All relevant analyses are executed in the statistical programming language `R` (R Core Team, 2017). Computationally intensive tasks are implemented in `C++` and connected to `R`.

4.2. Formation period

In the formation period, which is set to 30 days[4], we conduct an in-sample training of all stock pairs and a selection procedure to identify the most suitable pairs for the out-of-sample trading period. For each pair, algorithm 1 is applied to the corresponding spread time series with a Lévy-driven OU model (LDM) in each regime. The estimator of Mai (2014) is used for each regime and the corresponding data subset (see section 2). Outputs of the algorithm are the optimal number of regimes and the corresponding parameter estimations.

[4]This setting is well in line with the length of our simulated time series in section 3.2.

In our application, to determine the threshold $\nu_n = \Delta_n^\beta$ we choose β as the upper limit of $\beta \in (0, 1/2)$ in line with Mancini (2009), Cont and Mancini (2011) and Endres and Stübinger (2017). The time interval is $\Delta_n = \frac{1}{250 \cdot 391}$ (Cont and Mancini 2011, Liu et al. 2017) for our minute–by–minute data. According to Liu et al. (2017), we assume that the spread process reverts back to a level $\mu(t)$ calculated by the mean of last two daily opening and closing values. In accordance with Cartea and Figueroa (2005), the volatility σ is estimated by the sample standard deviation.

Following Liu et al. (2017), we form a portfolio of most suitable pairs for trading by creating a ranking on estimated model parameters. At this we require that two stocks always belong to the same industry and we weight parameters by corresponding regime size. Pairs are sorted in descending order by their mean-reversion speed, volatility, and number and size of detected jumps. The top p pairs ($p \in \mathbb{N}$) with smallest ranking are transferred to the trading period. The selection procedure identifies pairs with desirable properties — fast mean-reversion combined with high trading activity (Zeng and Lee 2014, Liu et al. 2017, Yeo and Papanicolaou 2017).

4.3. Trading period

In the trading period, which is set to 5 days, we conduct an out-of-sample trading of the top p pairs. To capture the mean-reversion of spreads in our 5-day trading period, the trading strategy takes advantage of their behavior — stock pairs are opened when they reach an extreme value and closed when they revert back.

More specific, the trading signals are determined by Bollinger bands (Bollinger 1992, Bollinger 2001). Upper and lower band $\mu(t) \pm k\sigma(t)$ ($k \in \mathbb{R}^+$) are constructed by adding (subtracting) k-times the running standard deviation $\sigma(t)$ to (from) the mean-reversion level $\mu(t)$. We develop the trade rules as follows: If the spread X_t exceeds the upper Bollinger band, i.e.,

$$X_t > \mu(t) + k\sigma(t)$$

we go short in stock A and long in stock B. If the spread X_t falls below the lower Bollinger

band, i.e.,

$$X_t < \mu(t) - k\sigma(t)$$

we go long in stock A and short in stock B. We apply symmetric trading bands around the equilibrium level (see Bertram 2010 and Zeng and Lee 2014), closing positions at the opposite side of their opening value. If positions do not clear until the end of the trading period, we reverse them at 4.00 pm on the last trading day.

For our empirical application, we follow the existing literature and opt for a portfolio consisting of the top $p = 10$ pairs (Miao 2014, Stübinger and Endres 2018). For constructing Bollinger bands, we choose $k = 0.5$. This is in accordance with the "$1-\sigma$ rule" of asymmetric trading strategies — a common practice taking positions at one standard deviation and clearing them at the mean (see Avellaneda and Lee (2010), Zeng and Lee (2014), Göncü and Akyildirim 2016b, and Stübinger et al. 2018). The last 1955 minutes, i.e., 5 trading days, are used to calculate the running standard deviation (see Stübinger and Endres 2018).

The daily performance returns during the trading period are computed according to Gatev et al. (2006) as follows:

$$r_{P,t} = \frac{\sum_{i=1}^{p} w_{i,t} r_{i,t}}{\sum_{i=1}^{p} w_{i,t}}, \quad t \in T$$

$$w_{i,t} = (1 + r_{i,1}) \cdot (1 + r_{i,2}) \cdot \ldots \cdot (1 + r_{i,t-1}), \quad t \in T$$

where $r_{i,t}$ defines the return of stock i at day t, $w_{i,t}$ the corresponding weight, and p the number of pairs in portfolio P. We consider the return on committed capital and the fully-invested return. The former is the return on the number of pairs that are selected for trading, the latter is the return on actual employed capital.

To assess the added value of our LDM-based strategy, we compare it with the following established pairs trading variants: (i) classic correlation model (CCM), (ii) Bollinger Bands model (BBM), (iii) Ornstein–Uhlenbeck model (OUM) and an (iv) S&P 500 buy–and–hold strategy (MKT). To ensure a fair performance analysis, we set data and framework conditions identical to the LDM. The following lines explain the most important facts of the four benchmarks mentioned.

Classic correlation model (CCM)

For the simplest pairs trading strategy, we follow Chen et al. (2017)'s approach and select the top pairs based on the highest Pearson's correlation coefficients. In the following out-of-sample trading period, we open pairs when the spread deviates by more than two standard deviations. The trade is closed on return to the historical equilibrium level. For further details on this approach, we refer to Gatev et al. (2006) and Chen et al. (2017).

Bollinger Bands model (BBM)

The second benchmark uses the time-variable trading thresholds of Bollinger (2001) to improve CCM. Similar to before, we select the top pairs using correlation. However, the rigid trading thresholds are replaced by time-variable input and output signals, i.e., we determine the upper (lower) Bollinger band by adding (subtracting) twice the running standard deviation to (from) the running average. To be consistent with LDM, the last 1955 minutes are used to calculate the data points. The application of Bollinger bands allows us to capture drifts and volatility clusters which are typical characteristics of financial market data (see Cont 2007).

Ornstein–Uhlenbeck model (OUM)

We follow the research studies of Elliott et al. (2005), Avellaneda and Lee (2010) and Göncü and Akyildirim (2016a) and describe the spread dynamics in each regime using an OU model, which is one of the most popular mean-reverting models in continuous-time finance. Again, we apply the regime classification algorithm and select the pairs based on the highest mean-reversion speed and the highest variance. The trading thresholds are identical to those of LDM. In summary, OUM is a reduced version of LDM with the strong limitation of not being able to model jumps.

S&P 500 buy–and–hold strategy (MKT)

Finally, we compare our LDM with a naive S&P 500 buy–and–hold strategy. For this benchmark, we buy the S&P 500 index in 1998 and hold it throughout the trading period. This passive investment strategy runs independently of any market conditions.

5. Results

We follow Stübinger et al. (2018) and run a fully-fledged performance analysis for the top 10 pairs of LDM from February 1998 to December 2015, compared to the benchmarks CCM, BBM, OUM, and MKT. For this purpose, we evaluate risk–return characteristics and trading statistics (subsection 5.1), analyze the performance over time (subsection 5.2), and conduct a regime analysis (subsection 5.3). Then, we perform further analyses including common risk factors and robustness checks (subsection 5.4).

Following the high-frequency research studies of Liu et al. (2017) and Stübinger and Endres (2018), we depict transaction costs of 5 bps per share per half-turn, i.e., in total transaction costs of $c = 20$ bps. This widely accepted assumption is feasible given our high turnover strategy in a highly liquid stock universe based on minute–by–minute data. We show returns that are calculated based on committed capital, which is in line with the vast majority of the literature.

5.1. Strategy performance

Table 3 reports daily return characteristics and risk measures before and after transaction costs for the top 10 pairs from February 1998 to December 2015. We observe that all strategies achieve positive daily returns after transaction costs, ranging between 0.01 percent for CCM and 0.27 percent for LDM. This statement is confirmed by the Newey–West t-Statistic — a value of 11.41 for LDM indicates statistically significant returns after transaction costs, compared to 1.40 for a naive buy–and–hold strategy of the S&P 500 index. The median, a robust averaging method for large time series, of CCM, BBM, LDM, and MKT is at a similar level as the mean return. In stark contrast, the median of OUM (0.06 percent) shows a strong discrepancy pointing out that the returns of this strategy seem to be driven by outliers. The range as well as the standard deviation of the asymmetric trading strategies are less pronounced than of the symmetric trading strategies, e.g., the standard deviation amounts 0.30 percent for CCM, 0.39 percent for BBM, 1.45 percent for LDM, and 1.58 percent for OUM — this fact is not surprising since the symmetric trading strategies identify more situations of market inefficiency. Contrary to the general market, the return distributions of all strategies possess right skewness representing a favorable property for

investors. Measuring the amount, duration, and frequency of losses corroborates the results — the maximum drawdown for OUM (79.00 percent) is strongly elevated, compared to BBM (25.09 percent), LDM (29.70 percent), and CCM (41.33 percent). Finally, the hit rate of LDM, i.e., the share of returns greater than zero, clearly outperforms the other strategies with approximately 62 percent after transaction costs. Concluding, LDM generates promising risk–return characteristics, even after transaction costs. Our responsibility is to check the robustness and to investigate loadings on any common sources of systematic risk.

	Before transaction costs				After transaction costs				
	CCM	BBM	OUM	LDM	CCM	BBM	OUM	LDM	MKT
Mean return	0.0004	0.0006	0.0039	0.0055	0.0001	0.0002	0.0017	0.0027	0.0002
Standard error (NW)	0.0001	0.0001	0.0003	0.0003	0.0001	0.0001	0.0003	0.0002	0.0002
t-Statistic (NW)	5.7398	7.0307	11.1893	21.1395	1.3915	2.5207	5.1920	11.4135	1.3973
Minimum	-0.0187	-0.0332	-0.1384	-0.1105	-0.0192	-0.0334	-0.1390	-0.1112	-0.0903
Quartile 1	-0.0008	-0.0012	-0.0021	-0.0015	-0.0011	-0.0016	-0.0039	-0.0036	-0.0056
Median	0.0001	0.0004	0.0026	0.0056	-0.0001	0.0000	0.0006	0.0029	0.0006
Quartile 3	0.0012	0.0020	0.0074	0.0128	0.0010	0.0016	0.0050	0.0096	0.0062
Maximum	0.0805	0.0718	0.3488	0.2405	0.0785	0.0679	0.3014	0.2345	0.1158
Standard deviation	0.0031	0.0040	0.0167	0.0151	0.0030	0.0039	0.0158	0.0145	0.0126
Skewness	5.9282	2.7547	5.2885	1.1223	5.5453	2.6073	4.6573	1.2201	-0.0181
Kurtosis	132.9813	45.4779	85.1442	22.6084	123.6459	43.1416	74.5525	25.6670	7.6714
Historical VaR 1%	-0.0066	-0.0087	-0.0299	-0.0364	-0.0069	-0.0090	-0.0316	-0.0378	-0.0344
Historical CVaR 1%	-0.0098	-0.0130	-0.0502	-0.0507	-0.0101	-0.0133	-0.0517	-0.0519	-0.0492
Historical VaR 5%	-0.0032	-0.0044	-0.0125	-0.0172	-0.0036	-0.0047	-0.0144	-0.0192	-0.0194
Historical CVaR 5%	-0.0054	-0.0073	-0.0244	-0.0295	-0.0057	-0.0076	-0.0261	-0.0310	-0.0297
Maximum drawdown	0.1532	0.1848	0.2734	0.2679	0.4133	0.2509	0.7900	0.2970	0.5678
Share with return > 0	0.5415	0.5662	0.6504	0.7023	0.4723	0.5001	0.5336	0.6231	0.5307

Table 3: Daily return characteristics and risk measures before and after transaction costs for the top 10 pairs of CCM, BBM, OUM, LDM, compared to an S&P 500 long-only benchmark (MKT) from February 1998 until December 2015. NW denotes Newey–West standard errors with five-lag correction and CVaR the Conditional Value at Risk.

Table 4 summarizes trading statistics per 5-day trading period. We find that the number of actually traded pairs is quite different for CCM (5.79), compared to BBM (9.70), OUM (10.00), and LDM (10.00) — the difference between static and time-varying trading thresholds is clearly pointed out by this characteristic factor. As expected, the average number of

round-trip trades per pair is strongly reduced for the strategies CCM and BBM. A value of around 1.40 is explained by the well-known research result that high correlation does not necessarily implicate a cointegration relationship (Alexander, 2001). LDM achieves the highest trade number (6.60) — this fact is not surprising since we select pairs based on high volatility as well as agile jump behavior. The reduced trade duration for OUM and LDM (around 1.10) confirms the selection procedure presented in section 4. High mean-reversion speed results in a fast convergence to the long-term mean level minimizing the risk of financial losses.

	CCM	BBM	OUM	LDM
Average number of pairs traded per 5-day period	5.7924	9.6984	9.9993	10.0000
Average number of round-trip trades per pair	1.3507	1.4140	5.2913	6.5988
Standard deviation of number of round-trip trades per pair	5.1518	1.6800	3.4571	4.0925
Average time pairs are open in days	2.7006	2.7112	1.2412	1.0358
Standard deviation of time open, per pair, in days	1.9967	1.8875	0.8504	0.7344

Table 4: Trading statistics for the top 10 pairs of CCM, BBM, OUM, and LDM per 5-day trading period.

In table 5, we depict annualized risk–return measures. Annualized mean returns after transaction costs vary between 2.15 percent for CCM and 93.73 percent for LDM, compared to 3.93 percent for the general market MKT. The standard deviation of both the symmetric trading strategies and the general market amounts approximately 20 percent. Across all strategies, downside deviation is at a very low level indicating that volatility is caused by upside deviations. Notably, the Sharpe ratio, i.e., the excess return per unit of standard deviation, of LDM clearly outperforms the benchmarks with a value of 3.92 after transaction costs. As anticipated from table 4, returns based on committed capital and employed capital lead to (almost) the same results for BBM, OUM, and LDM, since the top pairs open in almost all cases.

	Before transaction costs				After transaction costs				
	CCM	BBM	OUM	LDM	CCM	BBM	OUM	LDM	MKT
Mean return	0.0952	0.1677	1.5874	2.8309	0.0215	0.0552	0.4724	0.9373	0.0393
Mean excess return	0.0659	0.1443	1.5360	2.7549	-0.0061	0.0332	0.4431	0.8987	0.0185
Standard deviation	0.0486	0.0640	0.2658	0.2402	0.0483	0.0627	0.2503	0.2295	0.2003
Downside deviation	0.0249	0.0347	0.1162	0.1295	0.0274	0.0372	0.1263	0.1398	0.1416
Sharpe ratio	1.4125	2.1787	5.7799	11.4690	-0.0676	0.4717	1.7706	3.9161	0.0924
Sortino ratio	3.8170	4.8313	13.6611	21.8672	0.7873	1.4850	3.7400	6.7044	0.2772
Mean return on employed capital	0.1846	0.1707	1.5876	2.8309	0.0493	0.0547	0.4725	0.9373	0.0393
Sharpe ratio on employed capital	2.1707	2.2440	5.7804	11.4690	0.3810	0.5226	1.7707	3.9161	0.0924

Table 5: Annualized risk–return measures before and after transaction costs for the top 10 pairs of CCM, BBM, OUM, and LDM, compared to an S&P 500 long-only benchmark (MKT) from February 1998 until December 2015.

5.2. Performance over time

Motivated by the time-varying performance results of Krauss and Stübinger (2017) and Stübinger and Bredthauer (2017), we ascertain the stability and potential drawdowns of the strategies over time. Therefor, figure 2 presents the development of an investment of 1 USD after transaction costs for CCM, BBM, OUM, LDM (first row) and the general market MKT (second row) over three sub-periods.

The first sub-period ranges from 1998 to 2006 and is characterized by the internet bubble bursting and the subsequent bull market. We clearly see the difference between symmetric and asymmetric trading strategies: annualized returns of LDM (154.04 percent) and OUM (137.85 percent) are well above those of BBM (8.97 percent) and CCM (1.82 percent). This strong outperformance is caused by the still unknown methods and algorithms during that time.

The second sub-period ranges from 2007 to 2009 and specifies the global financial crisis and its aftereffects. The general market shows strong swings and large drawdowns in the wake of the subprime mortgage crisis. As a typical feature of pairs trading, the strategies are robust against bear markets, e.g., LDM generates an annualized return of 105.14 percent, compared to 20.09 percent for OUM, 8.33 percent for BBM, and 9.11 percent for CCM.

The third sub-period ranges from 2010 to 2015 and describes the time span of comebacks and recommencements. LDM clearly outperforms the benchmarks and the general market

with annualized returns of 26.20 percent after transaction costs suggesting that our quantitative strategy exploits persistent capital market anomalies even in recent times. The trading algorithm OUM is adversely affected by the high amount of round-trip trades leading to a negative performance — there is at least full cost recovery by the symmetric strategies CCM and BBM.

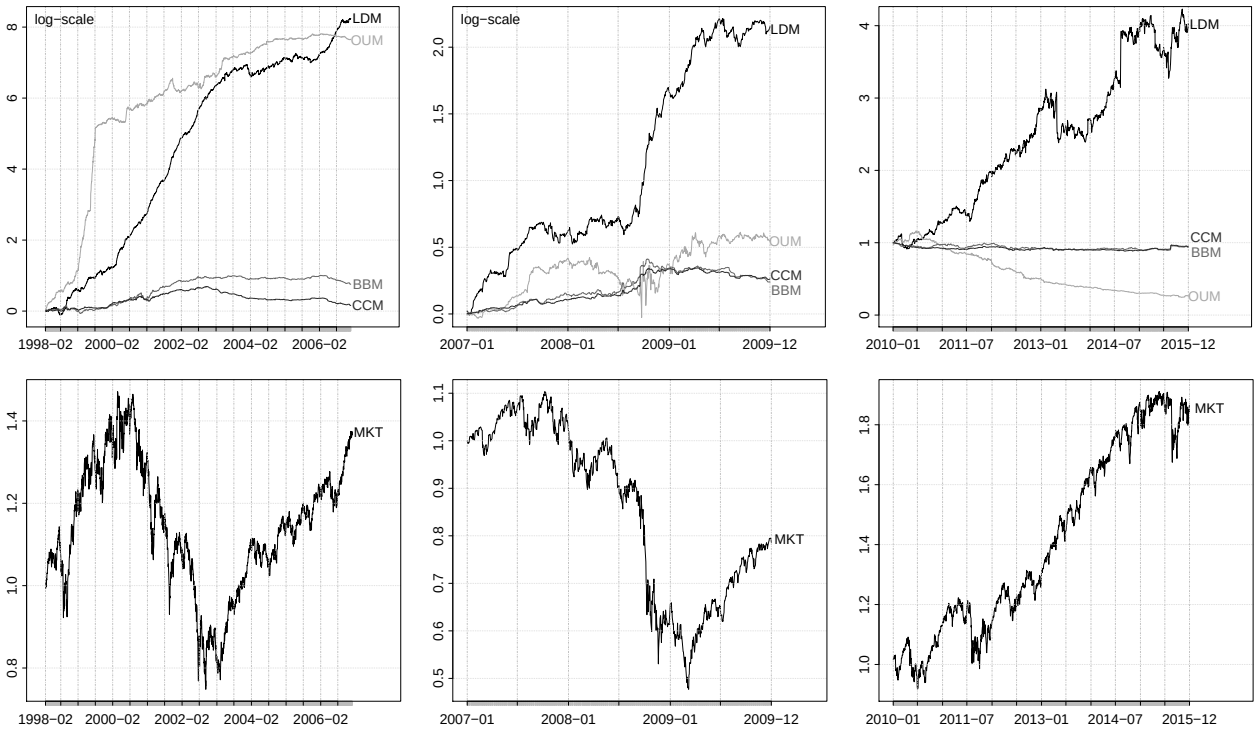

Figure 2: Development of an investment of 1 USD after transaction cots for the top 10 pairs of CCM, BBM, OUM, and LDM in the first row, compared to the S&P 500 index (MKT) in the second row. The time span from 1998 until 2015 is split into three sub-periods (1998–2006, 2007–2009, 2010–2015).

5.3. Regime analysis

The regime classification algorithm of section 3 generalizes approaches with a fixed number of regimes by choosing the states based on the underlying time series. This section vindicates the value-add of this flexibility (Hamilton 2010) and analyzes whether the well-documented regime shifts in the volatility process of financial return series (see Andersen et al. 2001, Hardy 2001, Nath and Dalvi 2004, Chen 2009, Liu et al. 2011, Li and Nolte 2016) are detected successfully in our models.

In figure 3, the percentage of identified regimes from 1998 until 2015 is depicted for LDM

and OUM. We observe that the share of pairs where 1 regime is found in the data amounts 53 percent for LDM and 66 percent for OUM, the share of 2 regimes 39 percent and 31 percent respectively. Thus, we confirm the assertion of Ang and Bekaert (2002), Li and Nolte (2016), and Elliott and Bradrania (2018), that the number of regimes in practice is rather small. But still, each of pairs with 1, 2, and 3 regimes all contribute a substantial share. Consequently, the flexible number of regimes is fully exploited by the algorithm. Forcing all pairs to have the same number of regimes would lead to misspecification of the model. Furthermore, for LDM the percentage of pairs exhibiting more than 1 regime is clearly higher than for OUM. It seems that disregarding of jumps leads to blurring of regime switches in the volatility that can not be detected any more.

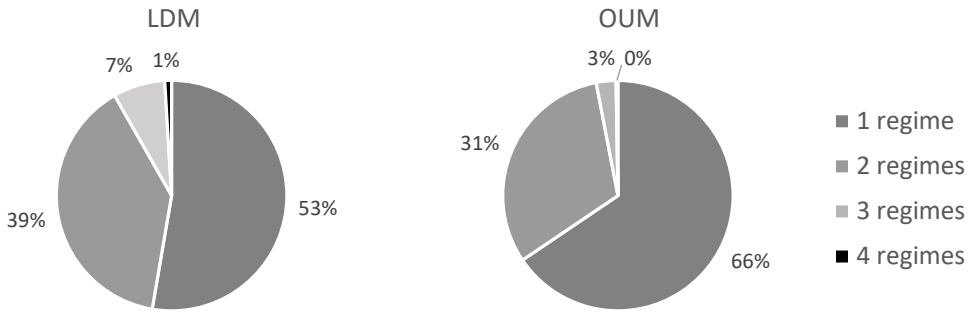

Figure 3: Number of detected regimes of all pairs of LDM (left) and OUM (right) in the period from 1998 to 2015 across all sectors.

In figure 4, we investigate the development of the number of detected regimes for both LDM (left) and OUM (right) over time. In accordance with figure 3, LDM tends to detect more regimes than OUM. The disregarded jumps of OUM seem to distort the detection of regime shifts in the volatility process. Especially in the financial crisis (2007–2009) we observe this effect — for LDM we find that the number of identified regimes increases clearly compared to OUM. Consequently, the possibility of time-varying regime patterns is fully exploited by LDM.

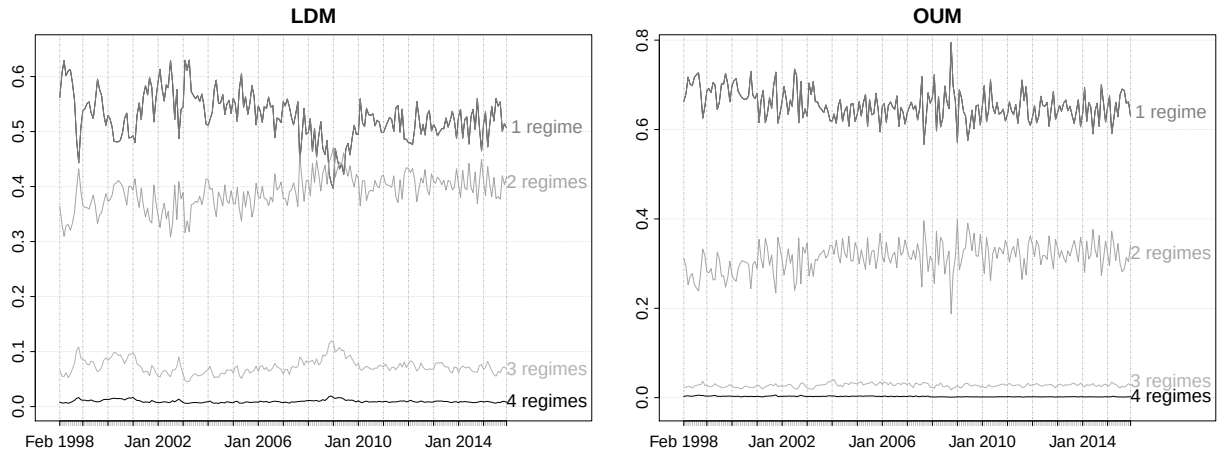

Figure 4: Percentage of number of detected regimes over time for all sectors for LDM (left) and OUM (right).

5.4. Further analyses

5.4.1. Common risk factors

Next, we analyze whether the LDM strategy is exposed to common systematic risk sources. In this context, three types of regressions are performed. The well-known Fama–French 3-factor model (FF3) of Fama and French (1996) covers exposure to the overall market, small minus large capitalization stocks (SMB) and high minus low book–to–market stocks (HML). The Fama–French 3+2-factor model (FF3+2) outlined in Gatev et al. (2006) adds a momentum factor and a short-term reversal factor to FF3. The Fama–French 5-factor model (FF5) (Fama and French, 2015) appends two further factors to FF3, namely equity portfolios with robust minus weak profitability (RMW) and conservative minus aggressive (CMA) investment behavior. All data needed for the analysis come from the website of Kenneth R. French.[5]

Table 6 analyzes the returns after transaction costs for the top 10 pairs of LDM. Regardless of the applied Fama–French model, we notice statistically and economically significant alphas of 0.26 and 0.27 percent per day. In addition, returns have no loading on the market — this is not surprising, as our strategy is dollar-neutral. We observe a statistically significant positive impact on the reversal factor, suggesting that LDM buys short-term losers and sells short-term winners. As expected, all other variables SMB, HML, MOM, SMB5, HML5,

[5]We thank Kenneth R. French for providing all the data needed for analysis.

RMW5, and CMA5 are small and not significant. In summary, LDM generates statistically and economically significant returns and does not depict any systematic sources of risk.

	FF3	FF3+2	FF5
(Intercept)	0.0027***	0.0026***	0.0027***
	(0.0002)	(0.0002)	(0.0002)
Market	−0.0270	−0.0350	−0.0259
	(0.0170)	(0.0189)	(0.0198)
SMB	−0.0170	−0.0211	
	(0.0343)	(0.0345)	
HML	0.0320	0.0576	
	(0.0323)	(0.0346)	
Momentum		0.0350	
		(0.0241)	
Reversal		0.0587*	
		(0.0243)	
SMB5			−0.0005
			(0.0370)
HML5			0.0514
			(0.0367)
RMW5			0.0520
			(0.0478)
CMA5			−0.0686
			(0.0588)
R^2	0.0009	0.0026	0.0014
Adj. R^2	0.0002	0.0014	0.0003
Num. obs.	4494	4494	4494
RMSE	0.0144	0.0144	0.0144

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 6: Exposure to systematic sources of risk for the daily returns of the top 10 pairs of LDM after transaction costs from February 1998 until December 2015. Standard errors are depicted in parentheses.

5.4.2. Robustness checks

As mentioned in section 4, we follow the existing literature and set the number of top pairs ($p = 10$), the trading threshold ($k = 0.5$), and the transaction costs ($c = 20$). Since data snooping is a frequently discussed topic in this context, we investigate the robustness of our results with regard to deviations from these parameters. In table 7, we vary p, k, and c in two directions and present the corresponding annualized return and Sharpe ratio for the top 10 pairs of LDM. Most importantly, we see that our results withstand parameter changes.

A lower number of top pairs (p) leads to higher annualized returns, suggesting that our pair selection process outlined in section 4 is meaningful — the corresponding Sharpe ratios are rising due to a reduced portfolio standard deviation. Higher results can generally be found at lower k-values, i.e., the higher transactions costs are exceeded by the higher profits as a result of increasing trading frequencies. Following Liu et al. (2017) and Stübinger and Endres (2018), we vary the transaction costs since investors are exposed to short selling constraints and bid–ask bounces. As expected, rising transaction costs lead to lower profitability — the break-even point for our standard parameter setting accounts for approximately $c = 40$.

p	$c \setminus k$	Return			Sharpe ratio		
		0.3	0.5	0.7	0.3	0.5	0.7
Top 5	0	7.0447	2.9009	1.6228	22.7765	10.1143	6.0602
	10	3.6054	1.7725	1.0520	12.0994	6.2985	3.9702
	20	1.6388	0.9706	0.6053	5.6776	3.4954	2.2900
	30	0.5132	0.4007	0.2557	1.7942	1.4307	0.9369
	40	-0.0440	-0.0044	-0.0179	-0.3074	-0.0952	-0.1548
Top 10	0	6.7593	2.8309	1.5233	24.6251	11.4690	6.4923
	10	3.4674	1.7239	0.9783	13.0721	7.1138	4.2073
	20	1.5749	0.9373	0.5509	6.1099	3.9161	2.3686
	30	0.4856	0.3782	0.2158	1.8922	1.5619	0.8866
	40	-0.1420	-0.0194	-0.0469	-0.6817	-0.1772	-0.3099
Top 20	0	6.5001	2.6540	1.4606	25.8072	11.5886	6.8689
	10	3.3455	1.6125	0.9328	13.7059	7.1567	4.4206
	20	1.5208	0.8683	0.5182	6.3930	3.8916	2.4494
	30	0.4640	0.3363	0.1925	1.9511	1.4799	0.8599
	40	-0.1488	-0.0440	-0.0633	-0.7665	-0.3074	-0.4241

Table 7: Yearly returns and Sharpe ratios for a varying number of top pairs (p), the k-times of the standard deviation, and the transaction costs (c) of LDM from February 1998 until December 2015.

To obtain another check on robustness, we conduct a bootstrap in the spirit of Gatev et al. (2006). The aim is to compare the returns of the true pairs of LDM with those of bootstrapped pairs and to contrast their performance. The following procedure is conducted 200 times. We combine the original entry and exit signals of the top pairs of LDM over the period from 1998 until 2015 with two randomly chosen securities at that time. The random stocks are drawn from all stocks of the S&P 500 except the two original ones. Then, we compare the average daily returns before transaction costs of the random trading with those we achieved with LDM. For the random trading, returns are close to zero at -0.008 percent per day — this is well in line with Gatev et al. (2006), who find slightly negative returns.

Opposite to this, even after transaction costs the LDM model produces daily returns of 0.27 percent. This provides a strong indication that our model exploits temporal effects that cannot be achieved by simple random trading.

6. Conclusion

In this paper, we develop the regime classification algorithm and apply it to high-frequency data of the S&P 500 constituents from January 1998 to December 2015. In that respect, our manuscript complements existing research in four aspects.

Our first contribution bears on the developed regime classification algorithm that estimates a full regime switching framework. The number of regimes is determined automatically for arbitrary processes in each regime based on statistical methods. Results from simulations demonstrate the algorithm's performance for Lévy-driven OU processes that switch between different regimes.

Our second contribution relies on the statistical arbitrage strategy in a high-frequency context — regime shifts in high-frequency volatility are detected and data is separated accordingly. Hereby, the introduced algorithm identifies the optimal number of regimes regarding the Bayesian information criterion. In each regime, a flexible Lévy-driven OU process is used to model spread time series, taking into account jumps and fat tails. The optimal pairs for the trading period are selected based on mean-reversion speed, volatility, and jump behavior.

The third contribution is based on our large-scale empirical study on high-frequency S&P 500 constituents from January 1998 to December 2015. Results from back-testing generate promising risk–return characteristics. With an annualized return of 93.73 percent and a Sharpe ratio of 3.92 after transaction costs we clearly outperform benchmark strategies from literature — among them a regime switching approach with a classic OU process in each state and an S&P 500 long-only strategy. In stark contrast to the benchmarks, our strategy achieves positive returns after transaction costs in recent years. Bootstrapping results show the profitability of the proposed model and we see that our results are robust to parameter changes.

Our fourth contribution consists of findings regarding the number of detected regimes across models and time. Our algorithm fully exploits the flexibility regarding the number of regimes compared to classic approaches where the regime number is determined in advance and identifies various regime patterns over time.

There are several possible directions for future research. First, optimal thresholds should be introduced in the trading strategy of our model. Then, entry and exit signals of the strategy are determined by thresholds that maximize an objective function, e.g., return or Sharpe ratio per unit time. Second, transition probabilities between the different regimes should be used in the selection of pairs for trading. Third, a multivariate model that accounts for common interactions between several stocks has to be estimated.

Bibliography

Aït-Sahalia, Y., Jacod, J., 2014. High-frequency financial econometrics. Princeton University Press, New Jersey, USA.

Alexander, C., 2001. Market models: A guide to financial data analysis. John Wiley & Sons, Chichester, England.

Altay, S., Colaneri, K., Eksi, Z., 2017. Pairs trading under drift uncertainty and risk penalization. Working Paper, Vienna University of Technology.

Andersen, T. G., Bollerslev, T., Das, A., 2001. Variance-ratio statistics and high-frequency data: Testing for changes in intraday volatility patterns. The Journal of Finance 56 (1), 305–327.

Ang, A., Bekaert, G., 2002. International asset allocation with regime shifts. The Review of Financial Studies 15 (4), 1137–1187.

Avellaneda, M., Lee, J.-H., 2010. Statistical arbitrage in the US equities market. Quantitative Finance 10 (7), 761–782.

Bai, Y., Wu, L., 2018. Analytic value function for optimal regime-switching pairs trading rules. Quantitative Finance 18 (4), 637–654.

Baronyan, S. R., Boduroğlu, İ. İ., Şener, E., 2010. Investigation of stochastic pairs trading strategies under different volatility regimes. The Manchester School 78 (1), 114–134.

Bee, M., Gatti, G., 2015. An improved pairs trading strategy based on switching regime volatility. Working Paper, University of Trento.

Bertram, W. K., 2009. Optimal trading strategies for Itô diffusion processes. Physica A: Statistical Mechanics and its Applications 388 (14), 2865–2873.

Bertram, W. K., 2010. Analytic solutions for optimal statistical arbitrage trading. Physica A: Statistical Mechanics and its Applications 389 (11), 2234–2243.

Bock, M., Mestel, R., 2009. A regime-switching relative value arbitrage rule. In: Fleischmann, B., Borgwardt, K.-H., Klein, R., Tuma, A. (Eds.), Operations Research Proceedings 2008. Springer, Berlin, Germany and Heidelberg, Germany, pp. 9–14.

Bollerslev, T., Litvinova, J., Tauchen, G., 2006. Leverage and volatility feedback effects in high-frequency data. Journal of Financial Econometrics 4 (3), 353–384.

Bollinger, J., 1992. Using Bollinger bands. Stocks & Commodities 10 (2), 47–51.

Bollinger, J., 2001. Bollinger on Bollinger bands. McGraw-Hill, New York, NY, USA.

Bouchaud, J.-P., Matacz, A., Potters, M., 2001. Leverage effect in financial markets: The retarded volatility model. Physical Review Letters 87 (22), 1–4.

Cai, J., 1994. A Markov model of switching-regime ARCH. Journal of Business & Economic Statistics 12 (3), 309–316.

Cartea, Á., Figueroa, M. G., 2005. Pricing in electricity markets: A mean reverting jump diffusion model with seasonality. Applied Mathematical Finance 12 (4), 313–335.

Chang, K.-L., 2009. Do macroeconomic variables have regime-dependent effects on stock return dynamics? Evidence from the markov regime switching model. Economic Modelling 26 (6), 1283–1299.

Chen, H., Chen, S., Chen, Z., Li, F., 2017. Empirical investigation of an equity pairs trading strategy. Management Science, Forthcoming.

Chen, S.-S., 2009. Predicting the bear stock market: Macroeconomic variables as leading indicators. Journal of Banking & Finance 33 (2), 211–223.

Chevallier, J., Goutte, S., 2017. On the estimation of regime-switching Lévy models. Studies in Nonlinear Dynamics & Econometrics 21 (1), 3–29.

Cont, R., 2007. Volatility clustering in financial markets: Empirical facts and agent-based models. In: Teyssière, G., Kirman, A. P. (Eds.), Long Memory in Economics. Springer, Berlin, Germany and Heidelberg, Germany, pp. 289–309.

Cont, R., Mancini, C., 2011. Nonparametric tests for pathwise properties of semimartingales. Bernoulli 17 (2), 781–813.

Cummins, M., Bucca, A., 2012. Quantitative spread trading on crude oil and refined products markets. Quantitative Finance 12 (12), 1857–1875.

Dahlquist, M., Gray, S. F., 2000. Regime-switching and interest rates in the European monetary system. Journal of International Economics 50 (2), 399–419.

Do, B., Faff, R., 2010. Does simple pairs trading still work? Financial Analysts Journal 66 (4), 83–95.

Do, B., Faff, R., Hamza, K., 2006. A new approach to modeling and estimation for pairs trading. In: Proceedings of 2006 Financial Management Association European Conference.

Elliott, R. J., Bradrania, R., 2018. Estimating a regime switching pairs trading model. Quantitative Finance 18 (5), 877–883.

Elliott, R. J., van der Hoek, J., Malcolm, W. P., 2005. Pairs trading. Quantitative Finance 5 (3), 271–276.

Endres, S., 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).

Esquivel, M. L., Mota, P. P., 2014. On some auto-induced regime switching double-threshold glued diffusions. Journal of Statistical Theory and Practice 8 (4), 760–771.

Fama, E. F., French, K. R., 1996. Multifactor explanations of asset pricing anomalies. The Journal of Finance 51 (1), 55–84.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116 (1), 1–22.

Fischer, T., Krauss, C., Treichel, A., 2018. Machine learning for time series forecasting-a simulation study. FAU Discussion Papers in Economics, Friedrich-Alexander University Erlangen-Nuremberg, Institute for Economics (2).

Gatev, E., Goetzmann, W. N., Rouwenhorst, K. G., 1999. Pairs trading: Performance of a relative value arbitrage rule. Working paper, Yale School of Management's International Center for Finance.

Gatev, E., Goetzmann, W. N., Rouwenhorst, K. G., 2006. Pairs trading: Performance of a relative-value arbitrage rule. Review of Financial Studies 19 (3), 797–827.

Göncü, A., Akyildirim, E., 2016a. Statistical arbitrage with pairs trading. International Review of Finance 16 (2), 307–319.

Göncü, A., Akyildirim, E., 2016b. A stochastic model for commodity pairs trading. Quantitative Finance 16 (12), 1843–1857.

Göncü, A., Karahan, M. O., Kuzubaş, T. U., 2016. A comparative goodness-of-fit analysis of distributions of some Lévy processes and Heston model to stock index returns. The North American Journal of Economics and Finance 36, 69–83.

Hamilton, J. D., 2010. Regime switching models. In: Macroeconometrics and time series analysis. Palgrave Macmillan, London, pp. 202–209.

Hamilton, J. D., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. Journal of Econometrics 64 (1), 307–333.

Hardy, M. R., 2001. A regime-switching model of long-term stock returns. North American Actuarial Journal 5 (2), 41–53.

Iacus, S. M., 2008. Simulation and inference for stochastic differential equations. Springer Series in Statistics, New York, USA.

Jondeau, E., Lahaye, J., Rockinger, M., 2015. Estimating the price impact of trades in a high-frequency microstructure model with jumps. Journal of Banking & Finance 61 (2), 205–224.

Kou, S., Yu, C., Zhong, H., 2017. Jumps in equity index returns before and during the recent financial crisis: A Bayesian analysis. Management Science 63 (4), 988–1010.

Krauss, C., Stübinger, J., 2017. Non-linear dependence modelling with bivariate copulas: Statistical arbitrage pairs trading on the S&P 100. Applied Economics 49 (52), 5352–5369.

Li, Y., Nolte, I., 2016. High-frequency volatility modelling: A markov-switching autoregressive conditional intensity model. Working Paper, Lancaster University.

Liu, B., Chang, L.-B., Geman, H., 2017. Intraday pairs trading strategies on high frequency data: The case of oil companies. Quantitative Finance 17 (1), 87–100.

Liu, F., Pantelous, A. A., von Mettenheim, H.-J., 2018. Forecasting and trading high frequency volatility on large indices. Quantitative Finance 18 (5), 737–748.

Liu, Z., Waggoner, D. F., Zha, T., 2011. Sources of macroeconomic fluctuations: A regime-switching DSGE approach. Quantitative Economics 2 (2), 251–301.

Mai, H., 2012. Drift estimation for jump diffusions: Time-continuous and high-frequency observations. Ph.D. thesis, Humboldt-Universität zu Berlin.

Mai, H., 2014. Efficient maximum likelihood estimation for Lévy-driven Ornstein-Uhlenbeck processes. Bernoulli 20 (2), 919–957.

Mancini, C., 2009. Non–parametric threshold estimation for models with stochastic diffusion coefficient and jumps. Scandinavian Journal of Statistics 36 (2), 270–296.

Masuda, H., 2010. Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes. Electronic Journal of Statistics 4, 525–565.

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 6 (3), 96–110.

Mota, P. P., Esquivel, M. L., 2014. On a continuous time stock price model with regime switching, delay, and threshold. Quantitative Finance 14 (8), 1479–1488.

Mota, P. P., Esquivel, M. L., 2016. Model selection for stock prices data. Journal of Applied Statistics 43 (16), 2977–2987.

Nath, G. C., Dalvi, M., 2004. Day of the week effect and market efficiency - Evidence from Indian equity market using high frequency data of national stock exchange. Working paper, Lansdale School of Business.

Pinson, P., Christensen, L. E., Madsen, H., Sørensen, P. E., Donovan, M. H., Jensen, L. E., 2008. Regime-switching modelling of the fluctuations of offshore wind generation. Journal of Wind Engineering and Industrial Aerodynamics 96 (12), 2327–2347.

QuantQuote, 2016. QuantQuote market data and software. https://quantquote.com.

R Core Team, 2017. stats: A language and environment for statistical computing. R package.

Rad, H., Low, R. K. Y., Faff, R., 2016. The profitability of pairs trading strategies: Distance, cointegration and copula methods. Quantitative Finance 16 (10), 1541–1558.

S&P Dow Jones Indices, 2015. S&P Global – Equity S&P 500 index. https://us.spindices.com/indices/equity/sp-500.

Stübinger, J., Bredthauer, J., 2017. Statistical arbitrage pairs trading with high-frequency data. International Journal of Economics and Financial Issues 7 (4), 650–662.

Stübinger, J., Endres, S., 2018. Pairs trading with a mean-reverting jump-diffusion model on high-frequency data. Quantitative Finance, Forthcoming.

Stübinger, J., Mangold, B., Krauss, C., 2018. Statistical arbitrage with vine copulas. Quantitative Finance, Forthcoming.

Uehara, Y., 2017. Statistical inference for misspecified ergodic Lévy driven stochastic differential equation models. Working Paper, Institute of Statistical Mathematics, Tokyo.

Vidyamurthy, G., 2004. Pairs trading: Quantitative methods and analysis. John Wiley & Sons, Hoboken, NJ, USA.

Yang, J.-W., Tsai, S.-Y., Shyu, S.-D., Chang, C.-C., 2016. Pairs trading: The performance of a stochastic spread model with regime switching-evidence from the S&P 500. International Review of Economics & Finance 43, 139–150.

Yeo, J., Papanicolaou, G., 2017. Risk control of mean-reversion time in statistical arbitrage. Risk and Decision Analysis 6 (4), 263–290.

Zeng, Z., Lee, C.-G., 2014. Pairs trading: Optimal thresholds and profitability. Quantitative Finance 14 (11), 1881–1893.