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

Sylvia Endres & Johannes Stübinger

Routledge
Taylor & Francis Group

Check for updates

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

SYLVIA ENDRES* and JOHANNES STÜBINGER

Department of Statistics and Econometrics, University of Erlangen–Nürnberg, D-90403, Nürnberg, Germany

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 estimates. 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.

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 academic research study of Gatev *et al.* (1999), Gatev *et al.* (2006) various quantitative methods are used to select and trade stock pairs. 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), Tourin and Yan (2013), Göncü and Akyildirim (2016), Rad *et al.* (2016), Lintilhac and Tourin (2017), Liu *et al.* (2017), Cartea *et al.* (2018) and Stübinger and

Endres (2018). Recently, regime switching models have 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.† The research is limited to Yang (2016), Altay *et al.* (2017) and Bai and Wu (2018). Yang (2016) combines 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

---

*Corresponding author. Email: sylvia.endres@fau.de

† In the following, the terms regime and state are used synonymously.

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 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 2016, Kou *et al.* 2017, Endres and Stübinger 2019). 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 % 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 % for a naive buy-and-hold strategy of the S&P 500 index and 47.24 % 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 briefly depicts data and software used in this study. Section 5 provides our high-frequency back-testing study. In section 6, the results and key findings are discussed. Finally, section 7 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, 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

$$X_t = \ln(S_A(t)/S_A(0)) - \ln(S_B(t)/S_B(0)), \quad t \geq 0 \quad (1)$$

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:

$$dX_t = \begin{cases} \theta_1 \left(\mu_1(t) - X_t\right) dt + dL_{1,t}, & \text{for } Z_t = 1 \\ \vdots & \vdots \\ \theta_r \left(\mu_r(t) - X_t\right) dt + dL_{r,t}, & \text{for } Z_t = r \end{cases} \quad (2)$$

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. Assuming finite jump activity (see Mai 2012, Larsson *et al.* 2013, Mai 2014), the driving Lévy process of $\{X_t\}_{t\geq 0}$ can be written as

$$L_{Z_t,t} = \begin{cases} \sigma_1 W_t + \sum_{j=1}^{N_{1,t}} \xi_{1,j}, & \text{for } Z_t = 1 \\ \vdots & \vdots \\ \sigma_r W_t + \sum_{j=1}^{N_{r,t}} \xi_{r,j}, & \text{for } Z_t = r \end{cases} \quad (3)$$

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)† . We recover the continuous part of $\{X_t\}_{t\geq 0}$ in the high-frequency limit via jump filtering. Based

---

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

on discrete variables $X_{t_1}, \ldots, X_{t_n}$, the following estimator is obtained:

$$\hat{\theta}_n = \frac{\sum_{i=0}^{n-1} (\mu_{t_i} - X_{t_i}) \Delta_i X \mathbb{1}_{\{|\Delta_i X| \le \nu_n\}}}{\sum_{i=0}^{n-1} (\mu_{t_i} - X_{t_i})^2 (t_{i+1} - t_i)}, \qquad (4)$$

where $\Delta_i X = X_{t_{i+1}} - X_{t_i}$ and $\Delta_n = \max_{1 \le i \le n-1} \{|\Delta_i t|\}$, $\Delta_i t = t_{i+1} - t_i$. 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. 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 (2019). 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 the last two daily opening and closing values. In accordance with Cartea and Figueroa (2005), the volatility σ is estimated by the sample standard deviation. The remaining jump parameters λ and η are calibrated by their maximum likelihood estimators. The first is the intensity estimator for the Poisson process (see equation (3)) and the second the jump size parameter, obtained by fitting the jump sizes to an exponential distribution following, e.g. Ramezani and Zeng (1998), Cai and Kou (2011), and Kou *et al.* (2017).

3. Regime classification algorithm

This section introduces a new approach, called the 'regime classification algorithm', to estimate a full regime switching framework (see section 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, section 3.2 validates the regime classification algorithm by means of Monte Carlo simulation.

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

$$CLS_r = \sum_{i=1}^{n-1} (X_{i+1} - E_{\boldsymbol{p}}[X_{i+1}|X_i, \ldots, X_1])^2 \qquad (5)$$

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

is computed as

$$BIC_r = n \ln \left(\frac{CLS_r}{n} \right) + m \ln(n), \qquad (6)$$

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 $\boldsymbol{v}$, which are calculated by the running standard deviation over the last 1955 minutes, i.e. 5 trading days:

$$\boldsymbol{v}[j - 1955]$$
$$= \sqrt{\sum_{i=j-1955}^{j-1} \left((\boldsymbol{x}[i] - \bar{\boldsymbol{x}}[j - 1955])^2 \right) / (1955 - 1)},$$
$$j = 1956, \ldots, \qquad (7)$$

where $\bar{\boldsymbol{x}}$ is the running mean over the last 1955 minutes. 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 % 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. The function *start.grid* determines at which threshold combination to start. The function *smart.grid* specifies the threshold combinations that should be tested— from the current combination, one threshold is optimized while the others are kept fixed. As such, in each step a new local optimum is obtained. For example, in the case of $r = 3$ regimes, we alternately optimize the thresholds c_1 and c_2 until no further improvement is achieved (see figure 1).

For each combination $\boldsymbol{c}$ recommended by the smart grid-search, the data is classified into r subsets $S_1 = \boldsymbol{x}[\boldsymbol{v} < c_1], S_2 = \boldsymbol{x}[\boldsymbol{v} < c_2 \, \& \, \boldsymbol{v} \ge c_1], \ldots, S_r = \boldsymbol{x}[\boldsymbol{v} \ge c_{r-1}]$ by the function *classify.data*. Using the observations in each of the regimes, conditional estimators $\boldsymbol{p} = \{\boldsymbol{p_1}, \ldots, \boldsymbol{p_r}\}$ 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 (6) 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
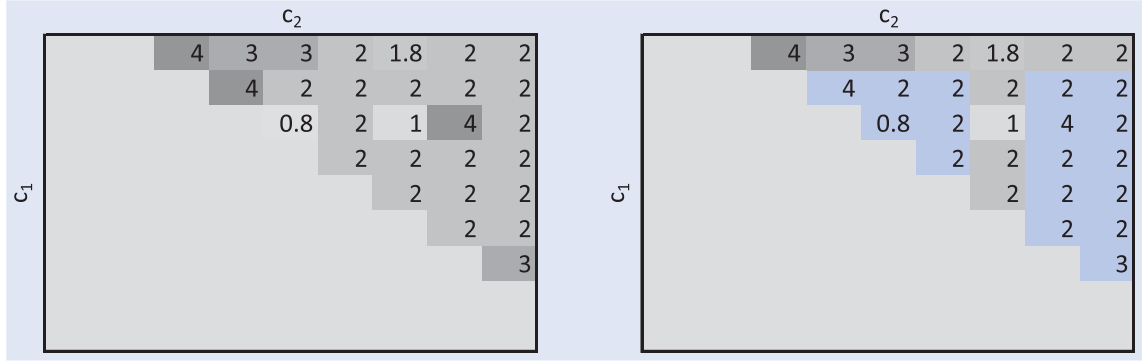
Figure 1. Full grid-search (left) compared to smart grid-search (right) controlled by *smart.grid* function for $r = 3$ regimes and 2 thresholds c_1 and c_2.

an optimum with lowest *BIC.local* is found, stored as BIC_r^*. The thresholds 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.

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

---

Algorithm 1 Regime classification algorithm (1/2)

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

Functions:
  fit.models$(S_1, ..., S_r)$:   Function   returning   estimated model parameter set $p = \{p_1, ..., p_r\}$ based on data sets $S_1, ..., S_r$.
  calc.BIC$(S_1, ..., S_r, p)$: Function returning BIC for data sets $S_1, ..., S_r$, and corresponding model parameter set $p = \{p_1, ..., p_r\}$.
  start.grid(x, r): Function returning set $\mathcal{I} \in \mathbb{R}^{z \times (r-1)}$ of z possible starting vectors, each including $r - 1$ thresholds, given the data series x.
  num.start(x, r): Function returning the number l of possible threshold combination sets $\mathcal{I}$ for r regimes and data series x.
  classify.data(x, r, c): Function returning r subsets $S = \{S_1, ..., S_r\}$ of data series x, separated by thresholds $c \in \mathbb{R}^{r-1}$.
  smart.grid$(x, \mathcal{I}, BIC_{\mathcal{I}})$: Function returning next threshold set $\mathcal{I}_{next} \in \mathbb{R}^{z \times (r-1)}$ for previous set $\mathcal{I}$ with corresponding $BIC_{\mathcal{I}}$.

Regime classification algorithm (2/2)

Algorithm:
  $p \leftarrow fit.models(x)$; $BIC_1^* \leftarrow calc.BIC(x, p)$;
  $r \leftarrow 2$;
  loop
    $\mathcal{I} \leftarrow start.grid(x, r)$; $l \leftarrow num.start(x, r)$;
    $a \leftarrow 1$; $BIC[a] \leftarrow +\infty$; $c_b \leftarrow c_{best} \leftarrow (0, ..., 0) \in \mathbb{R}^{r-1}$;
    loop
      $i \leftarrow 1$;
      loop
        $c \leftarrow \mathcal{I}[i]$;
        $S_1, ..., S_r \leftarrow classify.data(x, r, c)$;
        $p \leftarrow fit.models(S_1, ..., S_r)$;
        $BIC.local[i] \leftarrow calc.BIC(S_1, ..., S_r, p)$;
        if $i = z$ **then break**;
        $i \leftarrow i + 1$;
      end loop
      $a \leftarrow a + 1$; $c_{b.prev} \leftarrow c_b$;
      $BIC[a] \leftarrow \min(BIC.local)$;
      $c_b \leftarrow \mathcal{I}[argmin(BIC.local)]$;
      if $BIC[a - 1] < BIC[a]$ **or** $a - 1 \leftarrow l$ **then break**;
      $\mathcal{I} \leftarrow smart.grid(x, \mathcal{I}, BIC.local)$;
    end loop
    $c_{best.prev} \leftarrow c_{best}$; $c_{best} \leftarrow c_{b.prev}$;
    $BIC_r^* \leftarrow BIC[a - 1]$;
    if $BIC_{r-1}^* < BIC_r^*$ **then break**;
    $r \leftarrow r + 1$;
  end loop
  $S_1, ..., S_{r-1} \leftarrow classify.data(x, r - 1, c_{best.prev})$;
  $p \leftarrow fit.models(S_1, ..., S_{r-1})$;
  $BIC \leftarrow calc.BIC(S_1, ..., S_{r-1}, p)$;

---

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 ($i \in \{1, \ldots, r\}$), 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 $X_{t_1}, \ldots, X_{t_n}$ and estimate a full regime switching framework according to algorithm 1.

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.

	$r \setminus \hat{r}$	Lévy-driven OU process				Classic OU process			
		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

Note: The number of regimes estimated by algorithm 1 is denoted by $\hat{r}$.

Table 2. Mean-square error for $r \in \{1, 2, 3\}$, $n \in \{100, 1000, 10\,000\}$ and a Lévy-driven OU process in each regime with intensity λ and jump size variance σ_J^2.

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 1 contains simulation results for $N = 1000$ Monte Carlo iterations and sample size $n = 10\,000$ 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 %. 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 % 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 % 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 % 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.

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 ($n = 10\,000$) samples. Only those scenarios where algorithm 1 detects the same number of regimes for $n = 100, 1000, 10\,000$ 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.
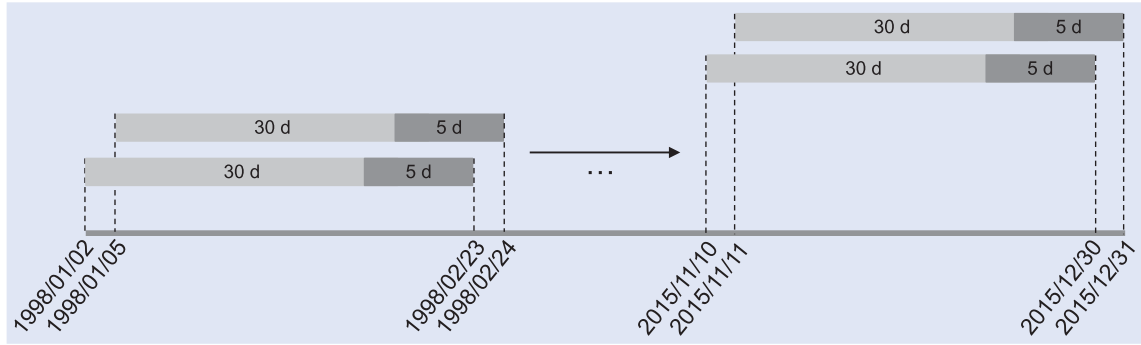
Figure 2. The period of 35 days length, consisting of 30 days formation and 5 days trading, is always shifted by one day. There are 4494 overlapping intervals of 35 days length from 1998/01/02 until 2015/12/31.

4. 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 % 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 (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.

5. 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 period (see section 5.1) and a 5-day trading period (see section 5.2). 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 2).

5.1. *Formation period*

In the formation period, which is set to 30 days†, 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.

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. The pairs with the best rankings are transferred to the trading period. Specifically, our selection procedure is conducted as follows. For each pair, we estimate mean-reversion speed θ, volatility σ, and number λ and size η of detected jumps. Based on the estimations, we calculate the corresponding rankings $r_\theta, r_\sigma, r_\lambda$, and r_η. Then, we select those pairs with smallest ranking $r_\theta + r_\sigma + r_\lambda + r_\eta$, i.e we weight all four rankings equally and choose the top pairs for trading, which exhibit highest model parameters θ, σ, λ, and η. Our selection procedure relies on two main assumptions. First, the parameter mean-reversion speed represents the tendency of the process to move back to its long-term equilibrium level (see Do *et al.* 2006, Zeng and Lee 2014, Stübinger and Endres 2018). This tendency is quantified by the half-life $\ln(2)/\theta$, i.e. the expected duration until convergence to the long-term mean level. By selecting pairs with strong mean-reversion, i.e. short half-life, we reduce the risk of spreads that might diverge during the investment horizon of the trader. Thus, the parameter θ is an essential driver for successful and fast termination of our pairs trading strategy. Second, volatility and jump behavior quantify the extent of process fluctuation and activity (see Zeng and Lee 2014, Liu *et al.* 2017, Stübinger and Endres 2018). Consequently, larger values of these parameters lead to high-frequency stock price dynamics creating major trading opportunities. The top p pairs ($p \in \mathbb{N}$) with smallest ranking are transferred to the trading period. The selection procedure identifies pairs with desirable

---

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

properties—fast mean-reversion combined with high trading activity (Zeng and Lee [2014](#), Liu *et al.* [2017](#), Yeo and Papanicolaou [2017](#)).

5.2. *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](#), [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.

$$X_t > \mu(t) + k\sigma(t) \tag{8}$$

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

$$X_t < \mu(t) - k\sigma(t) \tag{9}$$

we go long in stock A and short in stock B. We apply symmetric trading bands around the equilibrium level (see Bertram [2010](#), 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 [2016](#), 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](#)).

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.

5.2.1. 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](#)).

5.2.2. 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](#)).

5.2.3. Ornstein–Uhlenbeck model (OUM). We follow the research studies of Elliott *et al.* ([2005](#)), Avellaneda and Lee ([2010](#)) and Göncü and Akyildirim ([2016](#)) 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.

5.2.4. S&P 500 buy-and-hold strategy (MKT). Finally, we compare our LDM with a naive S&P 500 buy-and-hold strategy (see Knoll *et al.* [2018](#), Stübinger [2018](#)). 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.

5.3. *Transaction costs and return calculation*

Since our high-frequency trading strategy is based on permanent trading on the stock market, it would be naive to neglect transaction costs. Following Liu *et al.* ([2017](#)) and Stübinger and Endres ([2018](#)), we depict trading charges 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. A high percentage of available market capitalization (see section [4](#)) implies no further capacity constraints.

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

$$r_{P,t} = \frac{\sum_{i=1}^{p} w_{i,t} r_{i,t}}{\sum_{i=1}^{p} w_{i,t}}, \quad t \in T \tag{10}$$

$$w_{i,t} = (1 + r_{i,1}) \cdot (1 + r_{i,2}) \cdot \cdots \cdot (1 + r_{i,t-1}), \quad t \in T, \tag{11}$$

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. In

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.

	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

Note: NW denotes Newey–West standard errors with five-lag correction and CVaR the Conditional Value at Risk.

literature, there exist two variants—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. We show returns that are calculated based on committed capital, which is in line with the vast majority of the literature (see Gatev *et al.* (2006) for further details).

6. 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 (section 6.1), analyze the performance over time (section 6.2), and conduct a regime analysis (section 6.3). Then, we perform various robustness checks (section 6.4). All tables and figures, with exception of the second part of table 5, are computed based on committed capital.

6.1. 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 % for CCM and 0.27 % 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 %) 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 % for CCM, 0.39 % for BBM, 1.45 % for LDM, and 1.58 % 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 %) is strongly elevated, compared to BBM (25.09 %), LDM (29.70 %), and CCM (41.33 %). Finally, the hit rate of LDM, i.e. the share of returns greater than zero, clearly outperforms the other strategies with approximately 62 % after transaction costs. Concluding, LDM generates promising risk-return characteristics, even after transaction costs.

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 5. High mean-reversion speed results in a fast convergence to the long-term mean level minimizing the risk of financial losses.

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

	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 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.

	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

In table 5, we depict annualized risk-return measures. Annualized mean returns after transaction costs vary between 2.15 % for CCM and 93.73 % for LDM, compared to 3.93 % for the general market MKT. The standard deviation of both the symmetric trading strategies and the general market amounts approximately 20 %. 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.

Table 6 depicts the annualized mean return and Sharpe ratio for varying transaction cost levels. In our back-testing study, we assume transaction costs of 5 bps per share per half-turn, resulting in 20 bps per round-trip per pair (see section 5.3). The results for 0 bps and 20 bps are identical with table 5. Due to lower trading frequencies (see table 4), CCM is less affected by transaction costs than BBM, OUM, and LDM. The breakeven point for CCM is between 25 bps and 30 bps, the breakeven point for BBM and OUM between 30 bps and 35 bps. LDM shows a larger level at which cost and revenue are equal—the breakeven point is between 35 and 40 bps. Higher trading frequencies generating higher transaction costs are overcompensated by the achieved returns. Concluding, LDM generates promising results, even for investors that are exposed to different market conditions and thus higher transaction costs.

6.2. 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. Therefore, figure 3 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 %) and OUM (137.85 %) are well above those of BBM (8.97 %) and CCM (1.82 %). 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 %, compared to 20.09 % for OUM, 8.33 % for BBM, and 9.11 % 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 % 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.

6.3. 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

Table 6. Annualized mean return and Sharpe ratio for the top 10 pairs of the CCM, BBM, OUM, and LDM from February 1998 until December 2015 for different transaction costs (TC) per round-trip trade per pair in bps.

|	Annualized mean return				Annualized Sharpe ratio			
TC	CCM	BBM	OUM	LDM	CCM	BBM	OUM	LDM
0	0.0952	0.1677	1.5874	2.8309	1.4125	2.1787	5.7799	11.4690
5	0.0760	0.1384	1.2469	2.2302	1.1252	1.8183	4.5940	9.1229
10	0.0574	0.1099	0.9515	1.7239	0.7493	1.3865	3.5404	7.1138
15	0.0392	0.0822	0.6950	1.2971	0.3818	0.9617	2.6039	5.3922
20	0.0215	0.0552	0.4724	0.9373	− 0.0676	0.4717	1.7706	3.9161
25	0.0043	0.0289	0.2792	0.6339	− 0.3264	0.1331	1.0286	2.6495
30	− 0.0125	0.0033	0.1115	0.3782	− 0.6669	− 0.2708	0.3673	1.5619
35	− 0.0289	− 0.0217	− 0.0342	0.1625	− 0.9982	− 0.6676	− 0.2228	0.6271
40	− 0.0449	− 0.0460	− 0.1606	− 0.0194	− 1.3205	− 1.0574	− 0.7498	− 0.1772
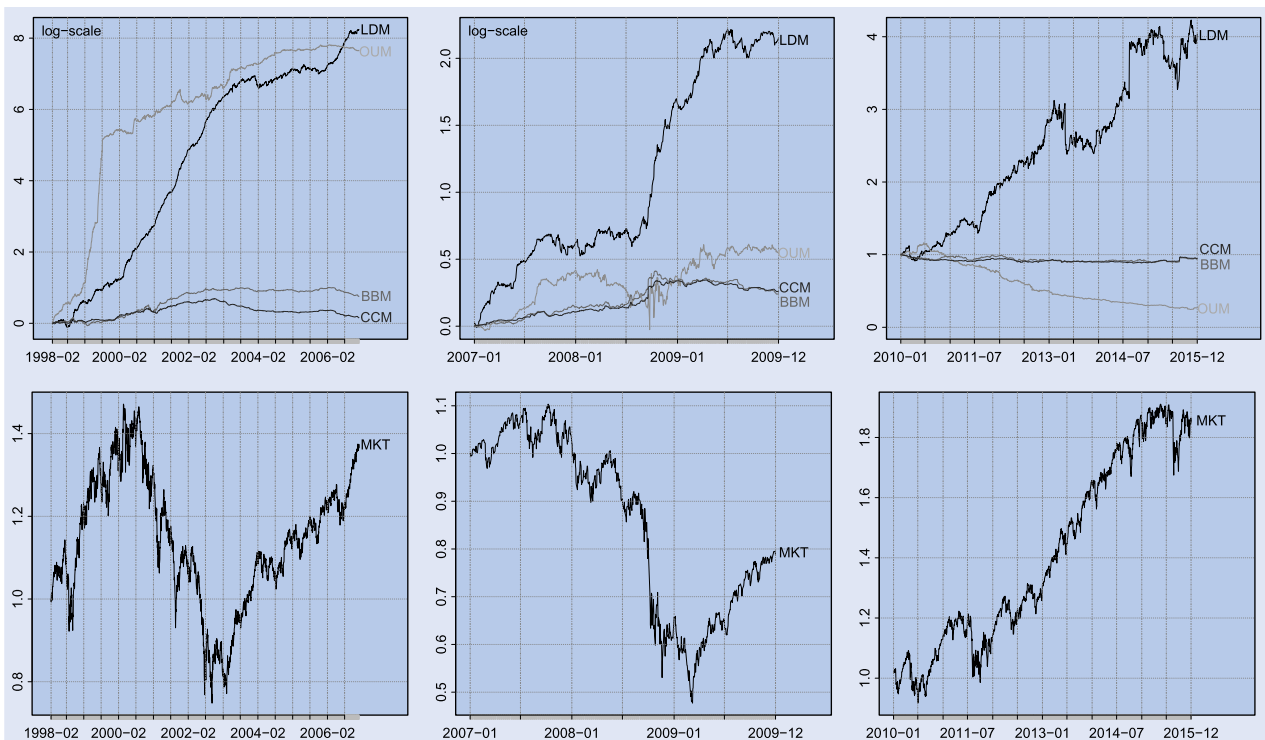
Figure 3. Development of an investment of 1 USD after transaction cots for the top 10 pairs of CCM, BBM, OUM, and LDM in the first row, compared to the S&P 500 index (MKT) in the second row. The time span from 1998 until 2015 is split into three sub-periods (1998–2006, 2007–2009, 2010–2015).

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 4, 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 % for LDM and 66 % for OUM, the share of 2 regimes 39 % and 31 %, 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.

In figure 5, we investigate the development of the number of detected regimes for both LDM (left) and OUM (right) over time. In accordance with figure 4, 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.

6.4. Robustness checks

As mentioned in section 5, we follow the existing literature and set the number of top pairs ($p = 10$), the trading
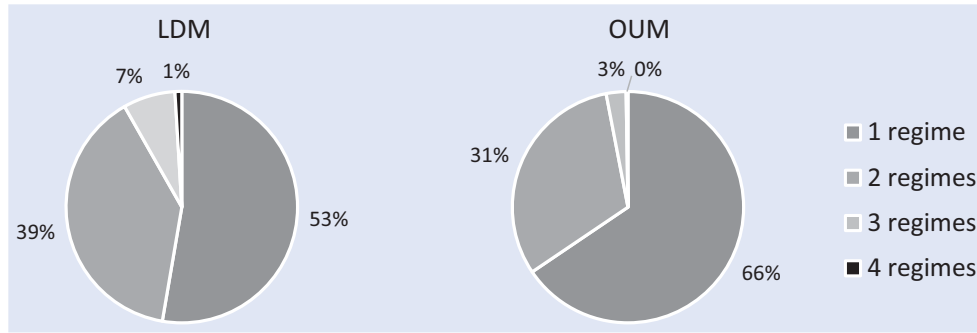
Figure 4. Number of detected regimes of all pairs of LDM (left) and OUM (right) in the period from 1998 to 2015 across all sectors.
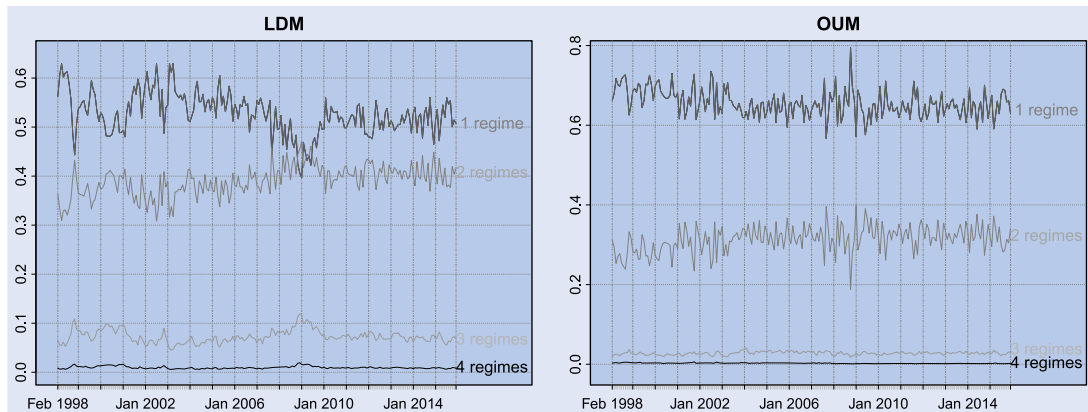

Figure 5. Percentage of number of detected regimes over time for all sectors for LDM (left) and OUM (right).

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 5 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 $c = 40$.

To obtain another check on robustness, we conduct a boot-strap 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% 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 %. This provides a strong indication that our model exploits temporal effects that cannot be achieved by simple random trading.

To further check for robustness, we follow Liu *et al.* (2017) and consider leverage in our return calculations. The margin leverage ratio (leverage) specifies the number of dollars investors are allowed to trade for every dollar they deposit in their brokerage account:

$$\text{leverage} = \frac{\text{gross market exposure}}{\text{margin deposit}}. \quad (12)$$

In our pairs trading strategies, we invest 1 USD long and 1 USD short, for 1 deposited USD, which implies a 2:1 leverage. Since large investors might get large leverages (see Liu *et al.* 2017) and results largely depend on the leverage choice, we consider different leverage ratios, i.e. a leverage of 1:1 as a conservative variant, and a leverage of 5:1 following Liu *et al.* (2017). A leverage of 1:1 requires a margin deposit of 2 USD for pairs trading, a leverage of 5:1 requires 0.4 USD. For a leverage of 1:1, we achieve annualized mean returns of 40.12 %, for a leverage of 2:1 93.73 %, and for a leverage of 5:1 373.33 %. The strategy of Liu *et al.* (2017) results in annualized returns of 1788 % in 2008 and 187.8 % during the period from 2013 until 2015. We conclude that leverage choice strongly influences trading results, but our initial leverage of 2:1 provides a moderate and realistic choice.

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.

| *p*	*c* \ *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
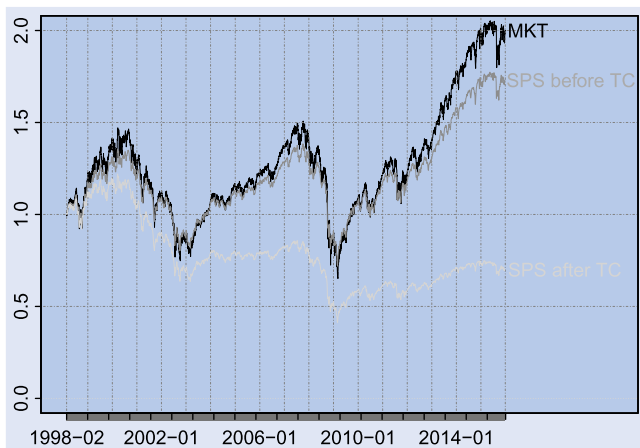
Figure 6. Development of an investment of 1 USD into S&P 500 index (MKT), S&P 500 strategy (SPS) before and after transaction costs (TC) from 1998 until 2015.

Another check on robustness is obtained following Gatev *et al.* (2006). We adopt a one-minute-waiting rule in order to deal with the bid–ask bounce, an essential issue in the context of high-frequency trading. Each time the algorithm produces an entry signal, we wait one minute until we buy the position, i.e. go long in stock A and short in stock B. Analogically, an exit signal delivered by the algorithm is not executed immediately but one minute later. Applying this rule, we achieve an annualized mean return across our sample period of about 20 %. Summarizing, the incorporation of market frictions in form of a one-minute-waiting rule negatively affects the results, but annualized returns remain above 20 %.

Finally, we take a deeper look into our S&P 500 buy–and–hold benchmark strategy (MKT). The S&P 500 index is bought in 1998 and hold throughout the full trading period. Of course the MKT is only a baseline approach for betting on the market. Therefor, we develop a more realistic strategy based on the S&P 500. In the S&P 500 strategy (SPS), we buy and sell the index at rolling 5-day intervals identical to

those of LDM. As expected, MKT and SPS show a similar development (see figure 6). MKT has an annualized return of 3.93 % (see also table 5), SPS of 3.11 %. To receive a more realistic evaluation, we consider transaction costs (SPS after TC). We observe annualized return of − 1.96 %—this fact is not surprising because it is a baseline approach without any modeling.

7. 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 % 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.

Acknowledgments

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Aït-Sahalia, Y. and Jacod, J. *High-Frequency Financial Econometrics*, 2014 (Princeton University Press: Princeton, NJ).

Alexander, C., *Market Models: A Guide to Financial Data Analysis*, 2001 (John Wiley & Sons: Chichester).

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

Andersen, T.G., Bollerslev, T. and Das, A., Variance-ratio statistics and high-frequency data: Testing for changes in intraday volatility patterns. *J. Finance*, 2001, **56**, 305–327.

Ang, A. and Bekaert, G., International asset allocation with regime shifts. *Rev. Financ. Stud.*, 2002, **15**, 1137–1187.

Avellaneda, M. and Lee, J.H., Statistical arbitrage in the US equities market. *Quant. Finance*, 2010, **10**, 761–782.

Bai, Y. and Wu, L., Analytic value function for optimal regime-switching pairs trading rules. *Quant. Finance*, 2018, **18**, 637–654.

Baronyan, S.R., Boduroğlu, İ.İ. and Şener, E., Investigation of stochastic pairs trading strategies under different volatility regimes. *Manch. Sch.*, 2010, **78**, 114–134.

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

Bertram, W.K., Analytic solutions for optimal statistical arbitrage trading. *Phys. A: Stat. Mech. Appl.*, 2010, **389**, 2234–2243.

Bertram, W.K., Optimal trading strategies for Itô diffusion processes. *Phys. A: Stat. Mech. Appl.*, 2009, **388**, 2865–2873.

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

Bollerslev, T., Litvinova, J. and Tauchen, G., Leverage and volatility feedback effects in high-frequency data. *J. Financ. Econom.*, 2006, **4**, 353–384.

Bollinger, J., Using Bollinger bands. *Stocks Commodities*, 1992, **10**, 47–51.

Bollinger, J., *Bollinger on Bollinger Bands*, 2001 (McGraw-Hill: New York, NY).

Bouchaud, J.P., Matacz, A. and Potters, M., Leverage effect in financial markets: The retarded volatility model. *Phys. Rev. Lett.*, 2001, **87**, 1–4.

Cai, J., A Markov model of switching-regime ARCH. *J. Bus. Econ. Stat.*, 1994, **12**, 309–316.

Cai, N. and Kou, S.G., Option pricing under a mixed-exponential jump diffusion model. *Manage. Sci.*, 2011, **57**, 2067–2081.

Cartea, Á. and Figueroa, M.G., Pricing in electricity markets: A mean reverting jump diffusion model with seasonality. *Appl. Math. Finance*, 2005, **12**, 313–335.

Cartea, A., Gan, L. and Jaimungal, S., Trading cointegrated assets with price impact. *Math. Finance*, forthcoming, 2018.

Chang, K.L., Do macroeconomic variables have regime-dependent effects on stock return dynamics? Evidence from the Markov regime switching model. *Econ. Model.*, 2009, **26**, 1283–1299.

Chen, H., Chen, S., Chen, Z. and Li, F., Empirical investigation of an equity pairs trading strategy. *Manage. Sci.*, 2017, **65**, 370–389.

Chen, S.S., Predicting the bear stock market: Macroeconomic variables as leading indicators. *J. Bank. Finance*, 2009, **33**, 211–223.

Chevallier, J. and Goutte, S., On the estimation of regime-switching Lévy models. *Stud. Nonlinear Dyn. Econom.*, 2017, **21**, 3–29.

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

Cont, R. and Mancini, C., Nonparametric tests for pathwise properties of semimartingales. *Bernoulli*, 2011, **17**, 781–813.

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

Dahlquist, M. and Gray, S.F., Regime-switching and interest rates in the European monetary system. *J. Int. Econ.*, 2000, **50**, 399–419.

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

Do, B. and Faff, R., Does simple pairs trading still work?. *Financ. Anal. J.*, 2010, **66**, 83–95.

Elliott, R.J. and Bradrania, R., Estimating a regime switching pairs trading model. *Quant. Finance*, 2018, **18**, 877–883.

Elliott, R.J., van der Hoek, J. and Malcolm, W.P., Pairs trading. *Quant. Finance*, 2005, **5**, 271–276.

Endres, S. and Stübinger, J., Optimal trading strategies for Lévy-driven Ornstein–Uhlenbeck processes. *Appl. Econ.*, forthcoming, 2019.

Esquivel, M.L. and Mota, P.P., On some auto-induced regime switching double-threshold glued diffusions. *J. Stat. Theory Pract.*, 2014, **8**, 760–771.

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

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

Gatev, E., Goetzmann, W.N. and Rouwenhorst, K.G., Pairs trading: Performance of a relative-value arbitrage rule. *Rev. Financ. Stud.*, 2006, **19**, 797–827.

Göncü, A. and Akyildirim, E., Statistical arbitrage with pairs trading. *Int. Rev. Finance*, 2016, **16**, 307–319.

Göncü, A. and Akyildirim, E., A stochastic model for commodity pairs trading. *Quant. Finance*, 2016, **16**, 1843–1857.

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

Hamilton, J.D, Regime switching models. In *Macroeconometrics and Time Series Analysis*, pp. 202–209, 2010 (Palgrave Macmillan: London).

Hamilton, J.D. and Susmel, R., Autoregressive conditional heteroskedasticity and changes in regime. *J. Econom.*, 1994, **64**, 307–333.

Hardy, M.R., A regime-switching model of long-term stock returns. *N. Am. Actuar. J.*, 2001, **5**, 41–53.

Iacus, S.M., *Simulation and Inference for Stochastic Differential Equations*, 2008 (Springer Series in Statistics: New York).

Jondeau, E., Lahaye, J. and Rockinger, M., Estimating the price impact of trades in a high-frequency microstructure model with jumps. *J. Bank. Finance*, 2015, **61**, 205–224.

Knoll, J., Stübinger, J. and Grottke, M., Exploiting social media with higher-order factorization machines: Statistical arbitrage on high-frequency data of the S&P 500. *Quant. Finance*, 2018. Advance online publication. doi: 10.1080/14697688.2018.1521002.

Kou, S., Yu, C. and Zhong, H., Jumps in equity index returns before and during the recent financial crisis: A Bayesian analysis. *Manage. Sci.*, 2017, **63**, 988–1010.

Krauss, C. and Stübinger, J., Non-linear dependence modelling with bivariate copulas: Statistical arbitrage pairs trading on the S&P 100. *Appl. Econ.*, 2017, **49**, 5352–5369.

Larsson, S., Lindberg, C. and Warfheimer, M., Optimal closing of a pair trade with a model containing jumps. *Appl. Math.*, 2013, **58**, 249–268.

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

Lintilhac, P.S. and Tourin, A., Model-based pairs trading in the bitcoin markets. *Quant. Finance*, 2017, **17**, 703–716.

Liu, B., Chang, L.B. and Geman, H., Intraday pairs trading strategies on high frequency data: The case of oil companies. *Quant. Finance*, 2017, **17**, 87–100.

Liu, F., Pantelous, A.A. and von Mettenheim, H.J., Forecasting and trading high frequency volatility on large indices. *Quant. Finance*, 2018, **18**, 737–748.

Liu, Z., Waggoner, D.F. and Zha, T., Sources of macroeconomic fluctuations: A regime-switching DSGE approach. *Quant. Econom.*, 2011, **2**, 251–301.

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

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

Mancini, C., Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scand. J. Stat.*, 2009, **36**, 270–296.

Masuda, H., Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes. *Electron. J. Stat.*, 2010, **4**, 525–565.

Miao, G.J., High frequency and dynamic pairs trading based on statistical arbitrage using a two-stage correlation and cointegration approach. *Int. J. Econ. Finance.*, 2014, **6**, 96–110.

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

Mota, P.P. and Esquivel, M.L., Model selection for stock prices data. *J. Appl. Stat.*, 2016, **43**, 2977–2987.

Nath, G.C. and Dalvi, M., 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, 2004.

Pinson, P., Christensen, L.E., Madsen, H., Sørensen, P.E., Donovan, M.H. and Jensen, L.E., Regime-switching modelling of the fluctuations of offshore wind generation. *J. Wind Eng. Ind. Aerodyn.*, 2008, **96**, 2327–2347.

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

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

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

Ramezani, C.A. and Zeng, Y., Maximum likelihood estimation of asymmetric jump-diffusion processes: Application to security prices. Working paper, California Polytechnic State University, 1998.

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

Stübinger, J., Statistical arbitrage with optimal causal paths on high-frequency data of the S&P 500. *Quant. Finance*, 2018. Advance online publication. doi: 10.1080/14697688.2018.1537503.

Stübinger, J. and Bredthauer, J., Statistical arbitrage pairs trading with high-frequency data. *Int. J. Econom. Financ. Issues*, 2017, **7**, 650–662.

Stübinger, J. and Endres, S., Pairs trading with a mean-reverting jump-diffusion model on high-frequency data. *Quant. Finance*, 2018, **18**, 1735–1751.

Stübinger, J., Mangold, B. and Krauss, C., Statistical arbitrage with vine copulas. *Quant. Finance*, 2018, **18**, 1831–1849.

Tourin, A. and Yan, R., Dynamic pairs trading using the stochastic control approach. *J. Economic Dyn. Control*, 2013, **37**, 1972–1981.

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

Vidyamurthy, G., *Pairs Trading: Quantitative Methods and Analysis*, 2004 (John Wiley & Sons: Hoboken, NJ).

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

Yeo, J. and Papanicolaou, G., Risk control of mean-reversion time in statistical arbitrage. *Risk Decis. Anal.*, 2017, **6**, 263–290.

Zeng, Z. and Lee, C.G., Pairs trading: Optimal thresholds and profitability. *Quant. Finance*, 2014, **14**, 1881–1893.