Journal of
Risk and Financial Management

MDPI

Article

On the Power and Size Properties of Cointegration Tests in the Light of High-Frequency Stylized Facts

Christopher Krauss *,† **and Klaus Herrmann** †

University of Erlangen-Nürnberg, Lange Gasse 20, 90403 Nürnberg, Germany; herrmannsklaus@posteo.de

* Correspondence: christopher.krauss@fau.de; Tel.: +49-0911-5302-278

† The views expressed here are those of the authors and not necessarily those of affiliated institutions.

Academic Editor: Teodosio Perez-Amaral

Received: 10 October 2016; Accepted: 31 January 2017; Published: 7 February 2017

Abstract: This paper establishes a selection of stylized facts for high-frequency cointegrated processes, based on one-minute-binned transaction data. A methodology is introduced to simulate cointegrated stock pairs, following none, some or all of these stylized facts. AR(1)-GARCH(1,1) and MR(3)-STAR(1)-GARCH(1,1) processes contaminated with reversible and non-reversible jumps are used to model the cointegration relationship. In a Monte Carlo simulation, the power and size properties of ten cointegration tests are assessed. We find that in high-frequency settings typical for stock price data, power is still acceptable, with the exception of strong or very frequent non-reversible jumps. Phillips–Perron and PGFF tests perform best.

Keywords: cointegration testing; high-frequency; stylized facts; conditional heteroskedasticity; smooth transition autoregressive models

1. Introduction

The concept of cointegration has been empirically applied to a wide range of financial and macroeconomic data. In recent years, interest has surged in identifying cointegration relationships in high-frequency financial market data (see, among others, [1–5]). However, it remains unclear whether standard cointegration tests are truly robust against the specifics of high-frequency settings. In our dataset, we find several stylized facts with potential impact on the power and size properties of contemporary cointegration tests. Most notably, the one-minute return data are highly non-normal and exhibit ARCH effects with intraday seasonalities and jumps. Additionally, we find evidence for nonlinear dependencies, even after applying AR(1)-GARCH(1,1) filtration; see Section 2.

Monte Carlo studies of the power and size properties of cointegration tests are no novelty to the literature. Existing approaches may be clustered as follows: The first group uses vector autoregressive (VAR) models with Gaussian innovations as data generating processes (DGPs). These are in line with the assumptions of the commonly-applied Johansen procedure; see [6–8]. The objective is to compare power and size properties across a wide range of different cointegration tests. Common references are [9–11]. The latter study constitutes the most comprehensive contribution and provides an excellent literature review. The second group analyzes the effect of non-Gaussian innovations on power and size properties. The work in [12] considers Student's t-distributions with three degrees of freedom, a truncated Cauchy distribution, Gaussian mixtures and others. Furthermore, bivariate ARCH and GARCH processes are also used to model the innovations. The work in [13] analyzes the impact of ARCH innovations from a multivariate Baba-Engle-Kraft-Kroner-ARCH DGP on the trace test. The works in [14,15] follow a similar approach. The third group covers the concept of threshold cointegration. The seminal reference is by [16], who were the first to introduce threshold nonlinearities in cointegration relationships and to analyze the power of conventional and enhanced tests in this setting. Further studies and improved tests follow [17,18].

However, none of these studies has considered the joint impact of high-frequency stylized facts on the power and size properties of contemporary cointegration tests. In this respect, the contribution of this paper is three-fold:

First, a procedure is developed to simulate high-frequency stock prices from actual market data, while retaining most of their idiosyncrasies.

Second, suitable DGPs are considered to simulate cointegration relationships, reflecting different stages of high-frequency stylized facts, i.e., non-normality, GARCH effects, threshold nonlinearities, reversible and non-reversible jumps. Our choice of DGPs is not only driven by the known stylized facts, but also by the economic interpretation of the cointegration relation. Similar approaches are known to the literature (see, e.g., [19,20]), but new in the context of testing. Here, we follow the notion that the cointegration relation reflects the law of one price (LOP); see [21] for the subsequent discussion. The work in [22] states the LOP as the "proposition (...) that two investments with the same payoff in every state of nature must have the same current value." The work in [23] expands on this concept and argues that "*closely integrated* markets should assign to similar payoffs prices that are *close*." In other words, the cointegration residual ties the price series of two securities together, ensuring that they are close. Whenever they diverge too far, the stationary models describe a relative value arbitrage mechanism as in [21], ensuring subsequent convergence. The latter is referred to as a "weak-form market integration" by [23], meaning a "near-efficient market". Consequently, we investigate the following models for the cointegration residual in our simulations:

- In a basic approach, an AR(1) and an AR(1)-GARCH(1,1) model are chosen, representing one stationary (or nonstationary) regime with arbitrage permanently occurring.
- A sophistication is given by a MR(3)-STAR(1)[1]-GARCH(1,1) with its different regimes to model the impact of transaction cost on arbitrage: In its middle regime, the cointegration residual truly behaves like a random walk; in this domain, arbitrage is not yet profitable. However, once the cointegration residual ventures into the outer regimes, arbitrage starts to occur.
- An addition of reversible jumps represents uninformed buying or selling, as described in [24], or in the sense of [25], who associate "transitory" or "observable jumps" in continuous time trading with temporary liquidity shortages.
- An alternative addition of non-reversible jumps represents idiosyncratic information, affecting only one of the two companies. Such a jump translates into a regime shift, causing further arbitrage to occur, but at a different level. A similar motivation is given in [25], who define "permanent" or "innovation jumps" as regime shifts in the fundamental value of a firm.

Third, ten contemporary cointegration tests are examined in a Monte Carlo simulation with respect to their power and size properties for sample sizes of $n = 510$ minutes, i.e., one trading day at Xetra. We find that: (1) non-normal innovations and GARCH effects have only a limited impact on the tests; (2) threshold nonlinearities lead to significant distortions with increasing threshold levels and abruptness of regime shifts; and (3) reversible jumps do not harm or even increase the power, and non-reversible jumps strongly deteriorate the power.

The remainder of this paper is organized as follows: Section 2 covers the high-frequency sample provided by Deutsche Börse AG. Section 3 reviews the three-step approach for simulating a cointegrated stock pair and assessing the power and size properties of the different cointegration tests. Section 4 presents the results and discusses the key findings in light of the existing literature. Finally, Section 5 concludes and summarizes directions for further research.

2. Data Sample and Its Stylized Facts

The data applied in this study consist of all transactions of the DAX 30 constituents traded on Xetra in continuous trading throughout the year 2014. Transaction data are aggregated to one-minute

[1] Multiple regime smooth transition autoregressive process.

bins, so we have more than 3.8 million data points at hand. The total value of all transactions amounts to EUR 910 billion, representing a major share of all trades in the respective stocks. Corporate actions data are available, as well. In order to smooth the data for the deterministic price changes implied by corporate actions (time and magnitude are largely known up front), we sanitize the returns by the approximate amount of their corporate actions, i.e., by setting the largest return in absolute value within a ten-minute window around the exact time of the corporate action to zero. In a first step, we analyze the data in light of well-known stylized facts for financial data. For each of the 249 trading days, we perform a set of tests for each of the 30 constituents. (The Jarque–Bera test, the Broock-Dechert-Scheinkman (BDS) test and the Teräsvirta test are implemented in the R package tseries by [26]. The Box test is part of the R package stats by the [27]. Engle's ARCH test is implemented in the R package FinTS by [28]. The Barndorff-Nielsen and Shephard (BNS) test of [29] is implemented in the R package highfrequency by [30]. The Tsay test is implemented in the R package TSA by [31]. The Luukkonen test of [32] is implemented in the R package twinkle by [33].) These tests are first run on the raw returns and then on the residuals of adequately-fitted AR(1) and AR(1)-GARCH(1,1) processes - the latter are fitted with the R package rugarch of [34]. Table 1 summarizes the results and depicts the share of tests with p-values less than 0.05. The raw and the filtered returns are highly non-normal, as indicated by the Jarque–Bera tests. ARCH effects are present in the raw returns, but largely filtered out by the AR(1)-GARCH(1,1) processes. The raw returns are highly nonlinear, as indicated by the BDS test, the Tsay test, the Luukkonen test and the Teräsvirta test. Even after the GARCH filter, there remains evidence of nonlinearity in the data. Volatility does not only show strong GARCH(1,1) effects, but exhibits significant intraday seasonality, as discussed, e.g., in [35]. Specifically, we observe the typical U-shaped pattern across all constituents. The clock time dependency of this seasonality pattern for each individual stock strongly resembles the pattern identified for the DAX future in [36]; see Figure 1.

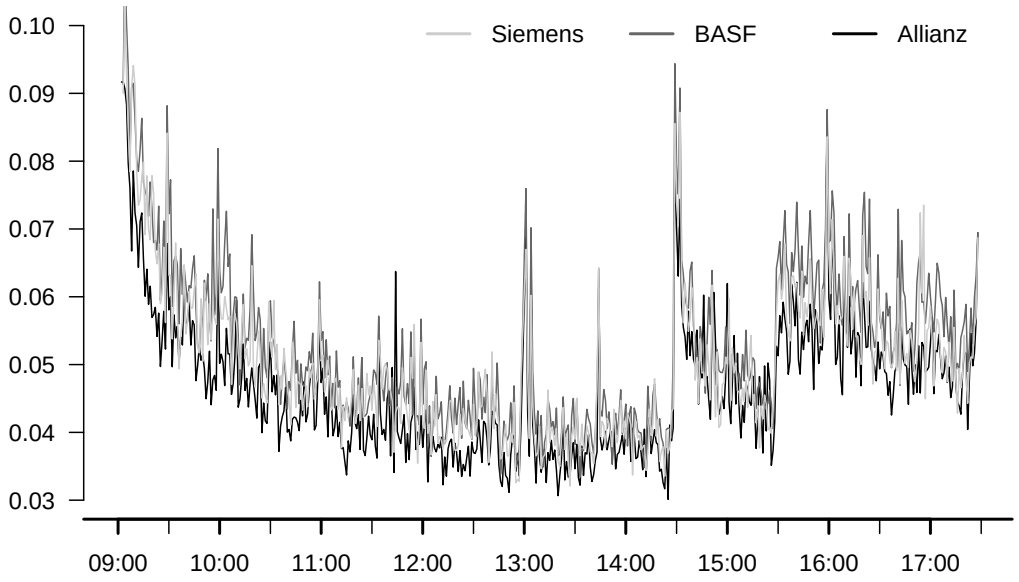

Figure 1. Average realized volatility per one minute of the trading day (in CET) for selected stocks in 2014.

Furthermore, the BNS test indicates jumps. These findings are in line with the existing literature; see, for example [37] for an account of relevant stylized facts in financial markets.

Table 1. Evaluation of minute-binned stock return data of the DAX 30 constituents over the 249 trading days in 2014. Subject to testing are raw returns, the residuals after AR(1)-filtration and the residuals after AR(1)-GARCH(1,1)-filtration. Depicted is the share of tests with a p-value less than 0.05.

Type	Test	Raw Returns	AR(1)	AR(1)-GARCH(1,1)
Non-normality	Jarque–Bera test	1.00	1.00	1.00
ARCH effects	Box test	0.57	0.55	0.03
	Engle's ARCH test	0.71	0.69	0.04
Nonlinearity	BDS test	0.84	0.76	0.24
	Tsay test	0.50	0.45	0.23
	Luukkonen test	0.41	0.40	0.08
	Teräsvirta test	0.37	0.36	0.07
Jumps	BNS test	1.00	0.99	1.00

In a second step, we check if the stylized facts of the cointegration residuals of potentially cointegrated stock pairs differ from the results above. We identify the cointegration relationship with the Johansen trace test at the five percent significance level. This approach is delicate, since one of the objectives of this paper is to examine the power of cointegration tests. However, we do not claim that the price series are really cointegrated, but much rather identify processes that are typically identified as cointegrated by one of the most commonly-used cointegration tests. We use the implementation of the Johansen procedure by [38] to examine all $30 \times 29/2 = 435$ pairs of stocks for each of the 249 full trading days (this amounts to a total of 108,315 tests). First, we use the augmented Dickey–Fuller test of [39] to check if the individual series are I(1) at the one percent significance level. We follow the routine outlined in [40], where first the augmented Dickey–Fuller (ADF) test is deployed on the two nonstationary processes and second the Johansen test. A more conservative approach could be to use the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test in the first step, with a null hypothesis of stationarity. Rejecting the I(0) null hypothesis of a KPSS test instead of accepting the I(1) null hypothesis of a unit root test may further restrict the number of pairs. If the series are considered I(1), the lag order of the VAR models is selected with the Bayesian information criterion (BIC). Next, the test statistic of the Johansen trace test is calculated, and the p-values are approximated with the routine described in [41] and implemented in R by [42]. A total of $435 \times 249 = 108{,}315$ processes are analyzed, and 41,069 are identified as cointegrated at the five percent significance level. Finally, we run the same set of tests to see if the stylized facts of the cointegration residuals differ from the results of Table 1. The results are summarized in Table 2. We see that the cointegration residuals exhibit similar stylized facts as the return data, i.e., non-normality, ARCH effects, jumps and further evidence of nonlinearities.

Table 2. Evaluation of 41,069 sets of cointegration residuals between the DAX 30 constituents identified on 249 trading days in 2014. Subject to testing are the raw data, the residuals after AR(1)-filtration and the residuals after AR(1)-GARCH(1,1)-filtration. Depicted is the share of tests with a p-value less than 0.05.

Type	Test	Raw Data	AR(1)	AR(1)-GARCH(1,1)
Non-normality	Jarque–Bera test	0.89	0.99	0.99
ARCH effects	Box test	1.00	0.70	0.13
	Engle's ARCH test	1.00	0.75	0.07
Nonlinearity	BDS test	1.00	0.64	0.49
	Tsay test	0.45	0.54	0.27
	Luukkonen test	0.55	0.38	0.15
	Teräsvirta test	0.55	0.35	0.16
Jumps	BNS test	1.00	0.98	1.00

Since the cointegration residuals are one of the focal points of this paper, we provide further evidence on GARCH effects. We deploy the R package rugarch of [34] to fit AR(1)-GARCH(1,1) models to all identified cointegration relationships, following a standard specification further detailed in Equation (2) with t-distributed innovations. We log all 41,069 $\times 5$ coefficients in a matrix of these dimensions, notably, the AR(1) coefficient ϕ_1, the intercept α_0 of the GARCH model, the ARCH parameter α_1, the GARCH parameter β_1 and the degrees of freedom of the t-distribution ν. We add an additional column for the unconditional standard deviation σ_G of the process, which can be calculated. Next, we sanitize the coefficient matrix by eliminating the 99th percentile of values of σ_G, as well as a handful of estimates with $\sigma_G < 0$. The latter cases usually correspond to processes, where the solver does not find a proper optimum. This sanitization eliminates a total of 424 cases, reducing the number of observations to 40,644. In Figure 2, we show a histogram for the above mentioned process parameters and in Table 3 their summary statistics. We see that α_1 is close to zero (mean of 0.064); β_1 is close to one (mean of 0.875); and $\alpha_1 + \beta_1$ even closer to one (mean of 0.940), i.e., close to variance nonstationarity. The t-distributed innovations exhibit low degrees of freedom; most of the probability mass is concentrated to the left of $\nu = 10$. The latter observations are typical for financial data. The unconditional standard deviation is very low; most of the probability mass is left of 0.1, which is in line with the short time period of one-minute returns.

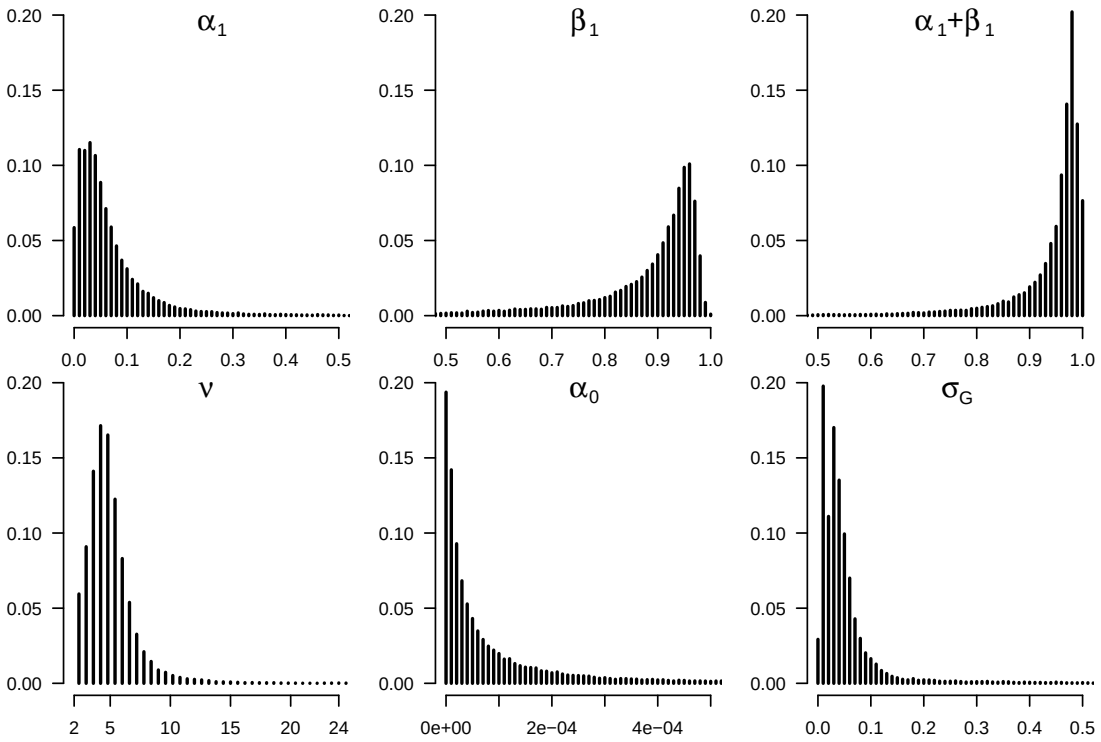

Figure 2. Distribution of parameter estimates, when fitting AR(1)-GARCH(1,1) models to the 41,069 sets of cointegration residuals between the DAX 30 constituents identified on 249 trading days in 2014.

Next, we submit the cointegration residuals to a detailed jump analysis, loosely following the methodology [25] applied to stocks. In particular, we difference each of the cointegration residuals, consisting of $n = 510$ observations. Then, we estimate the left and right tail index over the differenced series, using Hill's estimator (compare [43]) with a threshold of five percent, relying on the R package fExtremes. For simplicity's sake, we assume that the innovations of the differenced cointegration residuals are drawn from a t-distribution whose degrees of freedom are equal to the average of our left and right tail index estimates. We characterize a jump as an innovation outside of the one percent or outside of the 99 percent quantile. It is reversible, if the jump persists for only one period before the

process bounces back into its 98 percent confidence interval. Conversely, it is defined as non-reversible if the new level is sustained for at least two periods. Further details can be found in [25], Appendix 6.1. Table 4 reports the descriptive statistics for reversible and non-reversible jumps, with $q^{k\%}$ denoting the k-% quantile. We see that, on average, 3.66 reversible and 6.88 non-reversible jumps occur per day in the cointegration residuals. Non-reversible jumps are larger, with a mean of 0.17 EUR as opposed to 0.12 EUR for reversible jumps. Note that we observe more jumps than the theoretical quantiles of the t-distribution would suggest, indicating even fatter tails than estimated. This result is in analogy to [25].

Table 3. Summary statistics of parameter estimates, when fitting AR(1)-GARCH(1,1) models to the 41,069 sets of cointegration residuals between the DAX 30 constituents identified on 249 trading days in 2014.

	ϕ_1	α_0	α_1	β_1	$\alpha_1 + \beta_1$	ν	σ_G
Minimum	0.75187	0.00000	0.00000	0.00000	0.00000	2.10000	0.00000
1st Quartile	0.95820	0.00001	0.02244	0.84915	0.93438	3.74772	0.00029
Median	0.97370	0.00004	0.04491	0.92124	0.96879	4.62810	0.00119
Mean	0.96904	0.00016	0.06414	0.87542	0.93956	5.05612	0.00585
3rd Quartile	0.98516	0.00012	0.08126	0.95268	0.98277	5.69619	0.00312
Maximum	1.00000	0.03982	0.92903	0.99900	0.99900	99.99361	0.27104

Table 4. Summary statistics for reversible and non-reversible jumps detected in the 41,069 sets of cointegration residuals between the DAX 30 constituents identified on 249 trading days in 2014.

Panel A: Number of Jumps per Day	$q^{5\%}$	$q^{25\%}$	$q^{50\%}$	**Mean**	$q^{75\%}$	$q^{95\%}$
Reversible jumps	0.00	2.00	3.00	3.66	5.00	8.00
Non-reversible jumps	2.00	4.00	6.00	6.88	8.00	13.00
Panel B: Size of Jumps in EUR	$q^{5\%}$	$q^{25\%}$	$q^{50\%}$	**Mean**	$q^{75\%}$	$q^{95\%}$
Reversible jumps	0.02	0.05	0.10	0.12	0.16	0.31
Non-reversible jumps	0.02	0.06	0.11	0.17	0.19	0.45

3. Methodology

The suggested methodology to assess the power and size properties of ten cointegration tests is based on three steps. In the first step in Subsection 3.1, we present our approach to simulate high-frequency stock prices while retaining their stylized facts. In the second step in Subsection 3.2, we show how to simulate cointegration residuals, exhibiting different stylized facts. In the third step in Subsection 3.3, we finally present the methodology for simulating the cointegration relationship and, hence, the price of the second stock. Steps 1 to 3 are then repeated in different Monte Carlo settings in Subsection 3.4 to obtain the power and size properties of the tests under different conditions. This univariate design for assessing cointegration tests is loosely based on the setups of [44,45]. The parameter choices common to all Monte Carlo settings are discussed along Subsection 3.1 through Subsection 3.3, and specific choices are detailed in Subsection 3.4.

3.1. Simulation of Stock Prices

For an effective Monte Carlo simulation, we need a methodology to generate artificial high-frequency price series, while retaining their stylized facts as accurately as possible. We suggest a variant of the stationary bootstrap developed by Politis and Romano [46] that we apply across all Monte Carlo settings:

1. Set the return index i equal to one for initiation, i.e., to the first return of the day. Initialize a vector v with length 509 with zeros.
2. Draw one stock s out of the 30 DAX 30 constituents.

3. Draw one day d out of 249 full trading days.
4. Draw a random block length l from a geometric distribution with expected value of four [2].
5. Choose a block of length l, consisting of returns from stock s from day d for indices $i, i+1, ..., i+l$. Copy these returns in vector v at positions $[i, i+1, ..., i+l]$.
6. Update i with $i+l+1$. Go back to Step 1, until vector v consists of 509 returns.
7. Draw a random starting price between 5 and 40 from a uniform distribution. Accumulate the return vector v to a price series.

This procedure has several advantages. First, it ensures that the return series remains stationary; see [46]. Second, since blocks of a random length with an expected value of four are drawn, volatility clusters are partially preserved. Third, considering that the time ordering within a trading day is retained, intraday volatility seasonalities, i.e., the U-shaped pattern, are also reflected in the simulated prices. Fourth, also all further stylized facts can be observed in simulated price paths [3]. For example, jumps in the raw input data are drawn with some positive probability corresponding to the actual occurrence of jumps in high-frequency financial market data. Subsuming all of these advantages justifies the ad hoc nature of this nonparametric approach.

3.2. Simulation of Cointegration Residuals

In order to reflect the identified stylized facts, five DGPs are considered for the simulation of the cointegration residuals: a simple AR(1)-process, an AR(1)-GARCH(1,1) process, an MR(3)-STAR(1)-GARCH(1,1) process, an MR(3)-STAR(1)-GARCH(1,1) process with reversible jumps and an MR(3)-STAR(1)-GARCH(1,1) process with non-reversible jumps. In the following, we briefly discuss the relevant statistical properties of each DGP.

3.2.1. Autoregressive Model

Let u_t denote the cointegration residual at time t. An AR(1) process of this residual with coefficient ϕ_1 and scale parameter σ_G may be defined as:

$$u_t = \phi_1 u_{t-1} + \sigma_G \epsilon_t, \quad \text{with} \quad \epsilon_t \sim \mathcal{N}(0,1) \quad \text{or} \quad \epsilon_t \sim \mathbf{t}\left(0, \omega = \sqrt{\frac{\nu-2}{\nu}}, \nu\right), \tag{1}$$

such that the innovations are either normally distributed with mean zero and standard deviation of one or t-distributed with mean zero, ν degrees of freedom and a shape parameter ω. Following [47], the shape parameter ω is chosen in such a way that the standard deviation of the t-distribution equals one. For this purpose, ν must be strictly greater than two, so for all numerical implementations involved, a lower bound of 2.1 for ν is applied. Clearly, this process is stationary if $|\phi_1|$ is less than one. In case the normal distribution is chosen, this process does not reflect any of the high-frequency stylized facts. The t-distribution introduces non-normality. Note that in the AR(1)-case, the cointegrated processes share the same stylized facts, e.g., even jumps would happen simultaneously. This behavior is typical for dual-listed companies; see, for example, [48].

---

[2] A value of four is chosen ad hoc, as a compromise between partially preserving serial dependence in returns or squared returns and sufficient randomization. The latter refers to the fact that setting large block lengths leads to the risk of creating the simulated time series just on the basis of a few selected stocks. A value of four puts more emphasis on introducing a higher level of diversity.

[3] We replicate the analyses of Table 1 on bootstrap simulated data in order to see if the bootstrap preserves the stylized facts. It turns out that almost all results are very similar. We mainly observe a decrease in detected ARCH effects, as volatility patterns cannot be perfectly replicated with small block lengths.

3.2.2. Generalized Autoregressive Conditional Heteroscedasticity Model

The GARCH model of [49] generalizes Equation (1) to an AR(1)-GARCH(1,1) model, accounting for time-dependency in the scale parameter as:

$$u_t = \phi_1 u_{t-1} + x_t, \quad \text{with} \quad x_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad \alpha_0 > 0, \alpha_1, \beta_1 \geq 0 \tag{2}$$

where the time-dependency is determined by parameters α_0, α_1 and β_1. The innovations ϵ_t may again be either normally or t-distributed in the same manner as for the autoregressive model. The mean follows the same stationarity condition as Equation (1). A sufficient condition for strong stationarity for GARCH(1,1) is given if $E(\log(\beta_1 + \alpha_1 \epsilon_t^2)) < 0$ (compare [50]), whereas the necessary and sufficient condition for weak stationarity is $\alpha_1 + \beta_1 < 1$; see [49]. The given process is used to model the stylized facts of non-normality and of ARCH effects. Several popular extensions of GARCH, such as Glosten-Jagannathan-Runkle-GARCH ([51]) or exponential-GARCH ([52]), allowing for leverage/asymmetry in conditional volatility, have been neglected for the sake of simplicity and interpretability. The same applies to highly flexible multivariate volatility models, such as the diagonal Baba-Engle-Kraft-Kroner model for multivariate conditional volatilities and co-volatilities; see [53].

3.2.3. Multiple Regime Smooth Transition Autoregressive Model

Based on TAR models developed in [54,55], which first introduced STAR models, which have again been extended by [56] to three regimes, for our purpose, we use a simplified version of the latter model, i.e., a MR(3)-STAR(1) process defined as:

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-1} F_{12}(\gamma_1, c_1) + \phi_3 u_{t-1} F_{23}(\gamma_2, c_2) + \sigma_t \epsilon_t. \tag{3}$$

Thereby, $\phi_i, i \in [1,2,3]$ are the AR(1) coefficients of the three regimes, and F denotes the logistic transition function as suggested in van Dijk and Franses [56] with smoothness parameter γ and threshold c. σ_t follows again Equation (2) with innovations being either normally or t-distributed. With three regimes, threshold nonlinearities and GARCH effects, the model may be denoted as MR(3)-STAR(1)-GARCH(1,1). Its variance follows the same stationarity condition as above. Regarding the mean equation, to the best of our knowledge, no comprehensive stationarity conditions have been developed yet; compare the discussions in [57] or [58]. However, for our purposes, we just discuss simplified parameter constellations, following the ideas of [59] for a multiple-threshold AR(1) model:

Case A: Let ϕ_1 be equal to one, and $c_1 < c_2$. If ϕ_2 and ϕ_3 are equal to zero, we wind up having only one regime: a random walk, which clearly is nonstationary.

Case B: If ϕ_1 is an element of $[0.95, 0.90, 0.85]$, ϕ_2 is an element of $[0.05, 0.10, 0.15]$, ϕ_3 is an element of $[-0.05, -0.10, -0.15]$ and $c_1 < c_2$, the middle regime corresponding to ϕ_2 is nonstationary, but the outer regimes become stationary, making the entire process stationary. The latter is due to a mix effect. For values sufficiently [4] less than c_1, the logistic transition functions F_{12} and F_{23} converge towards zero, so the lower regime is stationary with a mixed coefficient ϕ_L element of $[0.95, 0.90, 0.85]$. For values sufficiently greater than c_1, but sufficiently smaller than c_2, the logistic transition function F_{12} converges towards one, but F_{23} still converges towards zero. Hence, the mixed coefficient for the middle regime ϕ_M is an element of $[1.00, 1.00, 1.00]$, indicating nonstationarity. For values sufficiently greater than c_2, F_{23} also converges towards one, leading to a mixed coefficient ϕ_U equal to $[0.95, 0.90, 0.85]$. Let us assume that we choose symmetric threshold levels, i.e., $c_1 = -c_2$, with $c_2 > 0$. The farther the thresholds are apart, the larger the nonstationary share of the process becomes. Furthermore, the smaller the parameter γ, the smoother the logistic function and the larger the influence of the lower and upper regimes.

[4] The point at which F_{12} converges to zero depends on the choice of γ.

The MR(3)-STAR(1)-GARCH(1,1) can thus model the stylized facts not only of non-normality and ARCH effects, but also nonlinear dependencies. The latter is related to the fact that the MR-STAR model "nests several other nonlinear time-series models (...), for example, an artificial neural network", as ([56], p. 317) point out.

3.2.4. Multiple Regime Smooth Transition Autoregressive Model with Reversible Jumps

We define reversible jumps as extreme events that are reversed over time after their initial occurrence, meaning they only have a temporary effect on the cointegration relationship. We model them by successively drawing waiting times w_i from an exponential distribution with parameter λ_w, so that the cumulative sum of the waiting times is less than or equal to $n = 510$ min (we round to the next full integer). The cumulative sum over the waiting times w_i provides the time index t of the current jump. At these time indexes, the innovations $\sigma_G \epsilon_t$ of Equation (1) and $\sigma_t \epsilon_t$ of Equations (2) or (3) are replaced by a draw from a Student's t-distribution defined as:

$$D_w \sim t\left(0, \omega_w = s_w m_w \sqrt{\frac{\nu_w - 2}{\nu_w}}, \nu_w\right), \tag{4}$$

where the shape parameter ω_w fixes the median of the absolute value of the t-distribution to the desired level m_w, i.e., the median jump size (s_w first standardizes it to one). In stationary models, these jumps only have a temporary impact. This model fulfills all of the stylized facts outlined in Section 2, except non-reversible jumps.

3.2.5. Multiple Regime Smooth Transition Autoregressive Model with Non-reversible Jumps

We define non-reversible jumps as extreme events that have a permanent effect on the time series, i.e., a mean shift. We model the jump series j_t as a compound Poisson process defined in ([60], p. 87) ff., as:

$$j_t = \sum_{i=1}^{N(t)} D_i, \quad \text{with} \quad D_i \sim t\left(0, \omega_p = s_p m_p \sqrt{\frac{\nu_p - 2}{\nu_p}}, \nu_p\right), \tag{5}$$

where $\{N(t), t \geq 0\}$ is a Poisson process with parameter λ_P, the expected number of jumps per trading day, and D_i is a sequence iid of t-distributed random variables independent from $N(t)$. The shape parameter ω_p fixes the median of the absolute value of the t-distribution to the desired level m_p, i.e., the median jump size (s_p first standardizes it to one). The cointegration residual u_t^* is then a superposition of the MR(3)-STAR(1)-GARCH(1,1) model of Equation (3) and the compound Poisson process in Equation (5), i.e.,

$$u_t^* = u_t + j_t. \tag{6}$$

As a consequence, this model fulfills all of the stylized facts outlined in Section 2, except reversible jumps. Note that the compound Poisson process has an expected value of zero, as well, but its variance increases over time.

3.2.6. Parameter Choices Common to All Monte Carlo Variants

The above models require various parameters that we have introduced in Equations (1) through (5). Generic parameter settings for all Monte Carlo variants with respect to the cointegration residuals are discussed below. In particular, the five GARCH parameters are sampled from actual high-frequency data, i.e., the ARCH parameter α_1, the GARCH parameter β_1, the degrees of freedom of the t-distribution ν and the unconditional standard deviation σ_G. Specifically, we sample with replacement from the 40,644 sets of sanitized cointegration residuals elaborated in Section 2. Each resampling provides us with five raw GARCH parameters specific to a particular set of cointegration residuals. Following [61], we fuzzify these parameters by adding Gaussian error terms with mean zero and standard deviation equal to the bandwidth determined with a method suggested

by [62]. We ensure that the error term does not lead to the violation of the general assumptions about the process parameters, e.g., among others, $\alpha_1 + \beta_1 < 1$ or $\sigma_G > 0$. This routine allows for effective resampling from realistic GARCH parameters, while introducing additional flexibility for the Monte Carlo simulations through the Gaussian error terms.

3.3. The Cointegration Relationship

Once the prices of the first stock are simulated according to the approach in Subsection 3.1 and the cointegration residual according to the approach in Subsection 3.2, we can finally determine the cointegration relationship. The latter is defined as:

$$p_{2t} = \alpha + \beta p_{1t} + u_t, \tag{7}$$

where p_{2t} is the price of the second stock, p_{1t} is the price of the first stock as simulated with the approach outlined in Subsection 3.1 and u_t is the cointegration residual as simulated with one of the processes suggested in Subsection 3.2. The coefficient α is the intercept, which is always drawn from a uniform distribution between -2 and 2. The parameter β is the coefficient of cointegration, which is always drawn from a uniform distribution between 1 and 5. The latter parameters are chosen ad hoc, as they have no impact on the power and size properties of cointegration tests.

3.4. Study Design

3.4.1. Cointegration Tests

In order to assess the power and size properties of ten different cointegration tests, we have adapted the R implementation of the package egcm of [45] to our setting. This package provides p-values for ten different unit root tests that are adjusted for cointegration testing. In other words, for all cointegration tests outlined in Table 5, we use the corresponding critical values suitable for a cointegration setting from egcm (see [45]), in order to not over-reject the null hypothesis. All tests are listed in Table 5 together with their implementation. Details about their functionality can be obtained from the references.

Table 5. Set of cointegration tests applied in this study. Note that critical values for cointegration testing and corresponding p-values for all deployed tests are taken from egcm; see [45].

Cointegration Test		**R Implementation**	
Augmented Dickey–Fuller test	[39,63]	tseries	[26]
Phillips–Perron test	[64,65]	tseries	[26]
Pantula, Gonzalez-Farias and Fuller test	[66]	egcm	[45]
Breitung's variance ratio test	[67,68]	egcm	[45]
Johansen's eigenvalue test	[6–8]	urca, vars	[38]
Johansen's trace test	[6–8]	urca, vars	[38]
Elliott-Rothenberg-Stock point optimal test	[69]	urca	[38]
Elliott-Rothenberg-Stock Dickey-Fuller Generalized Least Squares test	[69]	urca	[38]
Schmidt and Phillips rho statistic	[70]	urca	[38]
Based on Hurst exponent	[71]	fArma	[72]

3.4.2. Setup of Monte Carlo Simulations

We run six different types of Monte Carlo simulations of price pairs, differing in the setup of the cointegration relationship. The types are chosen to gradually incorporate more of the stylized facts discussed in Section 2. Each simulation is run with 10,000 replications per fixed parameter constellation and a path length of $n = 510$, i.e., one trading day. The parameters for these simulation may be distinguished into two groups. The first group is the parameters applied in a common way for all types. The choices for these parameters are outlined in Subsections 3.1 to 3.3. The second group

consists of parameters specific to each simulation type, discussed below, where Table 6 provides an overview.

Table 6. Parameter settings for the six types of Monte Carlo simulations conducted in this study and described in Subsection 3.4.2.

MC Type	**Type I**				**Type II**				**Type III**			
Process	AR(1)				AR(1)				AR(1)-GARCH(1,1)			
Distribution	Normal				t				t			
ϕ_1	1.00	0.95	0.90	0.85	1.00	0.95	0.90	0.85	1.00	0.95	0.90	0.85
MC Type	**Type IV**				**Type V**				**Type VI**			
Process	STAR(1)-GARCH(1,1)				STAR(1)-GARCH(1,1)				STAR(1)-GARCH(1,1)			
Regimes	3				3				3			
Jumps	-				reversible				non-reversible			
Distribution	t				t				t			
ϕ_L	1.00	0.95	0.90	0.85	1.00	0.95	0.90	0.85	1.00	0.95	0.90	0.85
ϕ_M	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
ϕ_U	1.00	0.95	0.90	0.85	1.00	0.95	0.90	0.85	1.00	0.95	0.90	0.85
c_1	–1.00	–5.00	–10.00					–1.00				–1.00
c_2	1.00	5.00	10.00					1.00				1.00
γ_1	1.00	5.00	10.00					5.00				5.00
γ_2	1.00	5.00	10.00					5.00				5.00
λ_w					2.00	3.00	8.00					
m_w					0.05	0.10	0.30					
λ_p									4.00	6.00	12.00	
m_p									0.05	0.10	0.45	

Type I: An AR(1) process with normally-distributed innovations is used to model the cointegration relationship. We test different values for ϕ_1. This setting reflects none of the stylized facts and should return largely undisturbed size and power properties for all tests. As mentioned before, $\phi_1 = 1$ returns the size of the tests, and $\phi_1 = [0.95, 0.90, 0.85]$ returns the power for different AR(1) coefficients. Hence, in total, four simulations are performed, each with 10,000 replications.

Type II: An AR(1) process with t-distributed innovations is applied in the same configuration as above, thereby reflecting the stylized fact of non-normality and its influence on size and power.

Type III: An AR(1)-GARCH(1,1) process with t-distributed innovations introduces non-normality and ARCH effects to the cointegration relationship. As in the scenarios above, we iterate through the four different values for ϕ_1, accounting for four simulations.

Type IV: A MR(3)-STAR(1)-GARCH(1,1) model with t-distributed innovations is the workhorse for a cointegration residual exhibiting non-normality, ARCH effects and nonlinear dependencies. Type IV is split into 36 simulations. Our simulations are based on three different, symmetric threshold levels c_i^*, defined as:

$$c_i^* = c_i \cdot \sigma_M, \quad \text{where} \quad |c_i| = [1.00, 5.00, 10.00], \quad \sigma_M = \frac{\sigma_G}{\sqrt{1-\phi^2}} \quad \text{and} \quad i = 1, 2. \tag{8}$$

In other words, the threshold c_i^* is a multiple c_i of the unconditional standard deviation σ_M of a hypothetical mixed process, assuming $\phi = 0.95 = \text{const.}$ across all applications [5,6]. For each of these

[5] Clearly, the mixed AR coefficient varies with the parameters ϕ_1, ϕ_2, ϕ_3, but for simplicity reasons and better comparability, it was fixed ad hoc at 0.95.

[6] This auxiliary metric is used to define fixed thresholds, even in light of potential non-stationarities.

threshold multiples c_i, three different values of γ_i are tested, i.e., $\gamma_i = [1.00, 5.00, 10.00]$. For each set of threshold multiple and gamma, we jointly vary ϕ_1, ϕ_2 and ϕ_3 through a total of four combinations, i.e., $\phi_1 = [1.00, 0.95, 0.90, 0.85]$, $\phi_2 = [0.00, 0.05, 0.10, 0.15]$ and $\phi_3 = -\phi_2$. The latter leads to $\phi_L = [1.00, 0.95., 0.90, 0.85] = \phi_U$ and $\phi_M = 1$. As before, each core simulation has 10,000 replications.

This choice of parameters is also in line with their economic interpretation: The (symmetric) thresholds c_1^* and c_2^* define at which level arbitrage kicks in, and the parameter γ determines how abruptly arbitrage kicks in. Regarding the abruptness of arbitrage kicking in, we can argue for a certain smoothness. Arbitrageurs most likely have different operating costs and different relative-value arbitrage models; for an overview of such strategies, see [73]. This diversity creates fuzziness around the threshold levels, which is reflected in the parameter γ. A value of ten already hinges towards a Heaviside function; a value of one is very smooth. As an ad hoc choice, we opt for a value of five. Regarding the thresholds, we have reason to believe that they would amount to the mean value of the cointegration residual plus or minus the transaction costs of the arbitrageur. The median bid-ask spread for all DAX 30 stocks in 2014 is in the range of four basis points (bps) or approximately 0.02 EUR, taking into account the price levels. Using Equation (8), the unconditional standard deviation of a hypothetical mixed process with $\sigma_G = 0.006$ (mean value observed in our data in Table 2) and $\phi = 0.95$ (median value corresponding to our data as in Table 7) amounts to 0.019 EUR; hence, clearly, the profit from a spread trade, entered at the threshold level of $|c_i^*| = 1 \cdot \sigma_M$ and held until the mean reversion amounts to 0.019 EUR and approximately covers transaction costs, given that the bid-ask spread has to be crossed only for one of the stocks on average. Considering this thought experiment, it is reasonable to believe that equity markets exhibit arbitrage threshold multiples approximately at $|c_i| = 1$.

Table 7. Ten thousand Type VI cointegration residuals with path length 510 and given mean-reversion coefficient ϕ are generated and contaminated with six non-reversible jumps ($\lambda_p = 6$) with median size $m_p = 0.10$. Depicted are the average estimated mean-reversion coefficients $\hat{\phi}$ compared to their true values ϕ.

$\phi_L = \phi_U$	0.850	0.900	0.950
$\hat{\phi}$	0.956	0.963	0.971

Type V: A MR(3)-STAR(1)-GARCH(1,1) model with t-distributed innovations is extended by reversible jumps. Type V consists of 36 simulations. We set the threshold multiple $|c_i| = 1$ and $\gamma_i = 5$. We iterate through three different values for λ_w, notably $\lambda_w = [2.00, 3.00, 8.00]$, which correspond to an expected value of two, three or eight jumps per trading day. For each parameter setting of λ_w, we test three different values for the scale parameter m_w of the t-distributed innovations, notably $m_w = [0.05, 0.10, 0.30]$. Both parameters, jump frequency λ_w and median jump size m_w, are motivated from the descriptive statistics in Table 4, effectively representing a best case (low jump frequency or jump size close to the 25% quantile), a base case (medium jump frequency or size close to the 50% quantile) and a worst case (high jump frequency or size close to the 95% quantile). The degrees of freedom ν_w are set to five in order to create a leptokurtic distribution with fat tails. For each combination of λ_w and m_w, we iterate through three combinations of ϕ_j, as in Type IV. This scenario reflects the stylized facts of non-normality, the ARCH effects, nonlinear dependencies and reversible jumps.

Type VI: An MR(3)-STAR(1)-GARCH(1,1) model with t-distributed innovations is extended by non-reversible jumps. Type VI consists of 36 simulations. As before, we set the threshold multiple $|c_i| = 1$ and $\gamma_i = 5$. We iterate through three different values for λ_p, notably $\lambda_p = [4.00, 6.00, 12.00]$, which correspond to an expected value of four, six or twelve jumps per trading day. For each parameter setting of λ_p, we test three different values for the scale parameter m_p of the t-distributed innovations for the compound Poisson process, notably $m_p = [0.05, 0.10, 0.45]$. Both parameters, jump frequency λ_p and median jump size m_p, are motivated from the descriptive statistics in Table 4, effectively representing a best case (low jump frequency or jump size close to the 25% quantile), a base case

(medium jump frequency or size close to the 50% quantile) and a worst case (high jump frequency or size close to the 95% quantile). These values are also in line with the findings in [25]. Applying a variant of their methodology suitable for cointegration relationships, we find a median number of six non-reversible jumps per day with a median jump size of approximately 0.10 EUR in absolute value. The degrees of freedom ν_p are again set to five. For each combination of λ_p and m_p, we iterate through three combinations of ϕ_j, as in Type IV. This scenario reflects the stylized facts of non-normality, the ARCH effects, nonlinear dependencies and non-reversible jumps.

Please note that, like the majority of the other parameters, the mean-reversion coefficients ϕ_1 (Types I and II) and $\phi_L = \phi_U$ (Types III to VI) selected for the power and size analysis are motivated based on our data; compare Table 3. Here, the descriptive statistics for the AR(1)-coefficients estimated from the empirical cointegration residuals are significantly closer to one than the values we apply, i.e., $\phi_1 = \phi_L = \phi_U = [1.00, 0.95., 0.90, 0.85]$. This alleged discrepancy reflects the fact that estimates for AR(1)-coefficients are biased towards the nonstationary case, if a true underlying AR(1)-process is contaminated with non-reversible jumps [7]. The fit of our choice of AR(1)-coefficients with the estimates in Table 3 is therefore illustrated by a simple simulation: We run 10,000 replications in order to see how the true coefficients translate to their empirical estimates in our Type VI case for stock market data, i.e., six non-reversible jumps of median size of 0.10 EUR in absolute value. Table 7 reports the results. We see that a mean-reversion coefficient $\phi_L = \phi_U$ of the true DGP translates to an estimate $\hat{\phi}$ of 0.956. This value approximately corresponds to the 25% quantile of the empirically-observed data (see Table 3), i.e, the least conservative case. A mean-reversion coefficient of 0.95 corresponds to estimates of 0.971, which is in line with the 50% quantile of the empirically observed data and, thus, defines the base case for stock market data.

4. Results

Henceforth, we use the following acronyms for cointegration tests in the subsequent tables: pp is the Phillips-Perron test, adf the Augmented Dickey-Fuller test, jo-e the Johansen Eigenvalue test, jo-t the Johansen Trace test, ers-p the Elliott-Rothenberg-Stock point optimal test, ers-d the Elliott-Rothenberg-Stock Dickey-Fuller Generalized Least Squares test, sp-r the Schmidt and Phillips rho statistic, hurst the Hurst exponent, bvr Breitung's variance ratio test, and pgff the Pantula-Gonzalez-Farias and Fuller test. The same acronyms are used in capital letters to refer to these tests in the main text of the results section.

4.1. Results Type I through Type III

Size: The Type I simulation unveils the undisturbed size properties of the cointegration tests. All simulations are run at a significance level of five percent. Since Type I does not violate any assumptions of any of the tests, the size should amount to five percent, as well, as long as the tests are properly calibrated. We see in Table 8 that only very limited size distortions occur for Type I, i.e., the normally-distributed innovations. When running the original routines of [45] with 10,000 replications, we also obtain a very slight size distortion at the level of one percentage point. Whereas [45] uses a standard deviation of one, we simulate the standard deviation as described in Subsection 3.2.6, with a mean value of 0.006 and a first quartile of 0.0003. The latter should not make any difference for the considered cointegration tests, so we attribute the fluctuations around the five percent level to chance alone.

---

[7] To see this effect, estimate the AR(1)-coefficient of two processes: (1) a simple stationary AR(1)-process with coefficient ϕ_1; and (2) the same process contaminated by a single large non-reversible jump. The estimate of the latter is biased towards the nonstationary case.

Table 8. Size of cointegration tests for Monte Carlo settings of Type I, Type II and Type III, as described in Subsection 3.4.2.

Test	ϕ_1	Type I	Type II	Type III
pp	1.00	0.06	0.06	0.07
adf	1.00	0.05	0.05	0.06
jo-e	1.00	0.06	0.06	0.06
jo-t	1.00	0.06	0.06	0.07
ers-p	1.00	0.05	0.05	0.06
ers-d	1.00	0.05	0.05	0.05
sp-r	1.00	0.05	0.05	0.05
hurst	1.00	0.05	0.05	0.06
bvr	1.00	0.05	0.05	0.05
pgff	1.00	0.06	0.06	0.06

For t-distributed innovations, the size remains unchanged to the second decimal across all tests. Hence, t-distributed innovations have no effect whatsoever on the size of these tests. The latter is expected, given the strong asymptotics for a sample size of $n = 510$. For Type III, the size increases on average by 0.01, meaning that GARCH models on the brink of variance nonstationarity, as incurred by high-frequency data, only lead to very limited size distortions. We note at this stage that for types IV through VI, no further significant size distortions beyond those of the AR(1)-GARCH(1,1) process occur; compare Tables 10, 12 and 14. The size properties are an important finding. They suggest that the stylized facts of high-frequency financial data have virtually no impact on the Type I error of these tests.

Power: The Type I simulation shows undisturbed power properties; see Table 9. We see that the power is above 0.80 for all tests for ϕ_1 equal to 0.85. With the AR(1) coefficient moving towards a random walk, the power deteriorates. For ϕ_1 equal to 0.95, the Phillips–Perron (PP) and the Pantula, Gonzalez-Farias and Fuller (PGFF) test still have excellent power of 0.84 and 0.87, respectively. This result confirms the widespread opinion of practitioners that the PP test performs better on financial data; see Alexander [74]. Their reasoning is based on the fact that the PP test allows for dependent errors with heteroscedastic variance by introducing a correction term to the Dickey–Fuller statistic. This correction evidently translates to improved power properties vis-a-vis the augmented Dickey–Fuller (ADF) test. Furthermore, we confirm the findings of [66] in a cointegration setting, who affirm the higher power of their alternative estimators compared to OLS. Along those lines, it makes sense that the ADF test does not perform as well as the PGFF test with a power of 0.64. However, it is surprising to see that the well-established Johansen procedure with the Johansen eigenvalue (JO-E) and the Johansen trace test (JO-T) shows unfavorable power properties at ϕ_1 equal to 0.95. Even though the Johansen test is still comparable to PP and PGFF for $\phi_1 = [0.90, 0.85]$, it significantly deteriorates when moving closer towards nonstationarity, exhibiting powers of 0.56 and 0.52 respectively. As [69] point out, the Elliot–Rothenberg–Stock (ERS-P, ERS-D) tests perform well in small samples and in case the series has an unknown mean or linear trend compared to the standard ADF test. Here, we face large sample sizes of $n = 510$ minutes, and linear trends are not considered. Hence, the power of these tests vacillates compared to the ADF test. At ϕ_1 equal to 0.95, the ERS tests show slightly better power, but for more stationary values, the ADF test dominates. The Schmidt–Phillips test (SP-R) performs almost equivalently to the ADF, with slight disadvantages at the most critical case of ϕ_1 equal to 0.95. The Hurst exponent and Breitung's nonparametric variance ratio test (BVR) disappoint.

Table 9. Power of cointegration tests for Monte Carlo settings of Type I, Type II and Type III, as described in Subsection 3.4.2.

Test	ϕ_1	Type I	Type II	Type III
pp	0.95	0.84	0.85	0.83
pp	0.90	1.00	1.00	1.00
pp	0.85	1.00	1.00	1.00
adf	0.95	0.64	0.63	0.64
adf	0.90	0.98	0.98	0.98
adf	0.85	1.00	1.00	1.00
jo-e	0.95	0.56	0.56	0.58
jo-e	0.90	1.00	1.00	0.99
jo-e	0.85	1.00	1.00	1.00
jo-t	0.95	0.52	0.53	0.54
jo-t	0.90	0.99	0.99	0.98
jo-t	0.85	1.00	1.00	1.00
ers-p	0.95	0.69	0.70	0.70
ers-p	0.90	0.91	0.91	0.90
ers-p	0.85	0.97	0.96	0.96
ers-d	0.95	0.65	0.67	0.68
ers-d	0.90	0.86	0.86	0.87
ers-d	0.85	0.92	0.92	0.92
sp-r	0.95	0.63	0.62	0.62
sp-r	0.90	0.97	0.97	0.96
sp-r	0.85	1.00	1.00	0.99
hurst	0.95	0.43	0.44	0.44
hurst	0.90	0.73	0.72	0.73
hurst	0.85	0.84	0.85	0.85
bvr	0.95	0.53	0.53	0.54
bvr	0.90	0.81	0.81	0.81
bvr	0.85	0.91	0.91	0.91
pgff	0.95	0.87	0.89	0.87
pgff	0.90	1.00	1.00	1.00
pgff	0.85	1.00	1.00	1.00

For Type II, this picture remains unchanged. The *t*-distributed innovations lead to fluctuations of the power of approximately 0.01 to 0.02, which seem random and go in both directions. The same applies to GARCH effects in Type III. So far, we conclude that non-normality and ARCH effects, as present in high-frequency financial data, still allow for cointegration testing with very respectable size and power properties. We can particularly recommend the PGFF and the PP tests. The very commonly-applied ADF test also exhibits adequate power.

4.2. Results Type IV

Type IV introduces threshold nonlinearities as additional stylized fact. Results on size are given in Table 10 and power in Table 11. Size: As size is stable at around 0.05 to 0.07 for all tests, the analysis may be focused on power.

Table 10. Size of cointegration tests for Monte Carlo settings of Type IV.

Test	$\phi_L = \phi_U$	1.00 1.00	1.00 5.00	1.00 10.00	5.00 1.00	5.00 5.00	5.00 10.00	10.00 1.00	10.00 5.00	10.00 10.00	c_i γ_i
pp	1.00	0.07	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.07	
adf	1.00	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	
jo-e	1.00	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	
jo-t	1.00	0.07	0.06	0.06	0.07	0.06	0.06	0.06	0.07	0.06	
ers-p	1.00	0.05	0.05	0.05	0.05	0.06	0.05	0.05	0.05	0.06	
ers-d	1.00	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.05	
sp-r	1.00	0.06	0.05	0.05	0.05	0.06	0.05	0.05	0.05	0.05	
hurst	1.00	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.05	0.05	
bvr	1.00	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	
pgff	1.00	0.06	0.06	0.07	0.07	0.07	0.07	0.07	0.06	0.07	

Table 11. Power of cointegration tests for Monte Carlo settings of Type IV, as described in Subsection 3.4.2.

Test	$\phi_L = \phi_U$	1.00 1.00	1.00 5.00	1.00 10.00	5.00 1.00	5.00 5.00	5.00 10.00	10.00 1.00	10.00 5.00	10.00 10.00	c_i γ_i
pp	0.95	0.78	0.65	0.59	0.54	0.23	0.16	0.38	0.13	0.10	
pp	0.90	0.99	0.97	0.95	0.90	0.41	0.27	0.70	0.22	0.13	
pp	0.85	1.00	0.98	0.97	0.95	0.51	0.34	0.81	0.28	0.16	
adf	0.95	0.58	0.46	0.42	0.39	0.17	0.13	0.28	0.11	0.09	
adf	0.90	0.96	0.87	0.81	0.80	0.33	0.20	0.58	0.18	0.11	
adf	0.85	1.00	0.95	0.93	0.92	0.43	0.27	0.74	0.23	0.13	
jo-e	0.95	0.50	0.38	0.34	0.33	0.13	0.11	0.22	0.09	0.08	
jo-e	0.90	0.97	0.89	0.84	0.80	0.30	0.18	0.55	0.16	0.10	
jo-e	0.85	1.00	0.97	0.95	0.92	0.40	0.25	0.74	0.22	0.11	
jo-t	0.95	0.48	0.36	0.32	0.31	0.13	0.10	0.21	0.09	0.08	
jo-t	0.90	0.96	0.85	0.79	0.76	0.28	0.18	0.52	0.15	0.09	
jo-t	0.85	0.99	0.96	0.94	0.90	0.38	0.23	0.71	0.20	0.11	
ers-p	0.95	0.67	0.58	0.55	0.52	0.24	0.17	0.36	0.13	0.10	
ers-p	0.90	0.88	0.82	0.79	0.78	0.39	0.26	0.61	0.21	0.13	
ers-p	0.85	0.95	0.91	0.88	0.87	0.47	0.31	0.73	0.25	0.16	
ers-d	0.95	0.65	0.57	0.53	0.50	0.23	0.16	0.35	0.13	0.09	
ers-d	0.90	0.86	0.79	0.75	0.74	0.36	0.24	0.60	0.20	0.13	
ers-d	0.85	0.92	0.86	0.83	0.84	0.46	0.30	0.70	0.25	0.14	
sp-r	0.95	0.57	0.47	0.43	0.41	0.18	0.13	0.28	0.10	0.08	
sp-r	0.90	0.94	0.85	0.78	0.78	0.32	0.20	0.57	0.18	0.11	
sp-r	0.85	0.99	0.95	0.91	0.90	0.41	0.26	0.72	0.21	0.13	
hurst	0.95	0.41	0.34	0.32	0.31	0.14	0.11	0.22	0.09	0.08	
hurst	0.90	0.69	0.61	0.55	0.56	0.25	0.17	0.42	0.15	0.09	
hurst	0.85	0.83	0.74	0.69	0.71	0.32	0.22	0.55	0.19	0.11	
bvr	0.95	0.51	0.43	0.41	0.39	0.19	0.14	0.29	0.12	0.08	
bvr	0.90	0.77	0.70	0.63	0.65	0.31	0.20	0.49	0.17	0.11	
bvr	0.85	0.90	0.81	0.76	0.78	0.38	0.25	0.63	0.22	0.12	
pgff	0.95	0.81	0.68	0.64	0.58	0.23	0.17	0.39	0.14	0.10	
pgff	0.90	1.00	0.97	0.96	0.91	0.42	0.28	0.71	0.23	0.14	
pgff	0.85	1.00	0.98	0.97	0.95	0.52	0.35	0.82	0.28	0.17	

Power: Overall, we see that the top of the ranking order of the tests remains as in Subsection 4.1, with PGFF and PP performing best, albeit with reduced power. We observe two dominant effects that negatively affect the power. First, increasing threshold levels significantly reduces the power. Take the PP test as an example. A threshold multiple of $|c_i| = 1.00$ leads to a power of 0.78 for

$\phi_L = \phi_U = 0.95$ and $\gamma_i = 1.00$. A threshold multiple of $|c_i| = 5.00$ reduces that power to 0.54 and a threshold multiple of $|c_i| = 10.00$ to 0.38. This behavior is easy to understand: With increasing threshold levels, a larger share of the total number of observations falls in the nonstationary middle regime. Hence, it is natural for the cointegration tests to exhibit lower power with the nonstationary share of the process gaining in weight. Speaking in relative terms, the decline in power driven by higher thresholds is approximately similar across all tests, with some exceptions. For example, for $\gamma_i = 1.00$, increasing the threshold multiples from ± 1 to ± 10 leads to a deterioration in the power of on average 50 percent for $\phi_L = \phi_U = 0.95$, 36 percent for $\phi_L = \phi_U = 0.90$ and of approximately 25 percent for $\phi_L = \phi_U = 0.85$. We see that threshold effects are getting stronger with AR coefficients approaching nonstationarity.

The second effect stems from the parameter γ of the logistic function in the STAR model. Increasing values of γ lead to more abrupt regime shifts and have an adverse effect on the power. The latter can be explained as follows: If γ is small, the logistic function is very smooth and "drags" the stationary behavior far into the otherwise nonstationary middle regime. In other words, the stationary outer regimes gain in weight at the expense of the nonstationary inner regime. The opposite is true for large values of γ. Then, the logistic function approaches a Heaviside function and induces very abrupt regime changes. In that case, the outer regimes lose weight relative to the inner regime. Logically, more realizations fall in the nonstationary inner regime, with a negative effect on the power.

The nonparametric BVR test, that should perform well in light of nonlinear dynamics (see [67]), does not gain the upper hand in this setting, but slightly improves its relative positioning for $\phi_L = \phi_U = 0.95$ and stronger threshold nonlinearities. PGFF and PP remain at the top of the ranking.

4.3. Results Type V

Type V introduces reversible jumps as additional stylized fact. Results on size are given in Table 12 and power in Table 13. Size: As size varies only between 0.04 and 0.07 for all tests, the analysis may be focused on power.

Power: We see that the power increases with λ_w, i.e., the expected number of jumps per 510 min trading day, and m_w, i.e., the median jump size in absolute value. For example, the PP test exhibits a power of 0.65 for $\phi_L = \phi_U = 0.95$ and $\lambda_w = 0$, which increases to 0.74 for $\lambda_w = 8$ and $m_w = 0.30$, i.e., the most extreme case. This result is not surprising. A reversible jump is defined as an innovation with much higher variance. However, the mechanics of the MR(3)-STAR(1)-GARCH(1,1) process still apply. As such, depending on the sign, there is a given chance that a jump boosts the process from the nonstationary middle regime to one of the stationary outer regimes. Then, mean-reversion occurs, driving the cointegration relationship back towards the middle regime. Of course, the contrary may occur, as well, meaning that the process jumps from one of the stationary outer regimes to the nonstationary inner regime. However, with increasing jump sizes, chances are becoming much higher for jumps occurring from the nonstationary to the stationary regimes than vice versa. The reason is the high median jump size in absolute value compared to the magnitude of the threshold levels. Clearly, we see that reversible jumps enlarge the proportion of observations in the stationary regimes, so it is quite intuitive that the cointegration tests exhibit higher power. We further note that this behavior is similar across all tests, i.e., reversible jumps leave the ranking of Table 11 largely unaffected. With respect to the AR-coefficients, we observe that gain in power is more expressed when moving closer to nonstationarity. However, most importantly, we note that significant gains in power can only be observed for AR(1)-coefficients of 0.95 and the more extreme jump scenarios with median jump sizes of 0.30 and an occurrence of eight jumps per trading day in expectation. The middle scenario with jump sizes of 0.10 and frequencies of 3.0 only shows very marginal gains in power, if any. However, this scenario corresponds to a typical cointegration residual exhibiting reversible jumps; see Table 4. We conclude that reversible jumps do not have an adverse effect on cointegration testing. On the contrary, in the unlikely case that they manifest at high rates and in high magnitude, they actually enhance the power.

Table 12. Size of cointegration tests for Monte Carlo settings of Type V.

Test	$\phi_L = \phi_U$	2.00 0.05	2.00 0.10	2.00 0.30	3.00 0.05	3.00 0.10	3.00 0.30	8.00 0.05	8.00 0.10	8.00 0.30	λ_w m_w
pp	1.00	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.05	
adf	1.00	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	
jo-e	1.00	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	
jo-t	1.00	0.07	0.07	0.07	0.07	0.06	0.07	0.06	0.07	0.07	
ers-p	1.00	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	
ers-d	1.00	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	
sp-r	1.00	0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.05	0.05	
hurst	1.00	0.05	0.05	0.06	0.05	0.05	0.06	0.05	0.05	0.06	
bvr	1.00	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	
pgff	1.00	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.05	

Table 13. Power of cointegration tests for Monte Carlo settings of Type V, as described in Subsection 3.4.2.

Test	$\phi_L = \phi_U$	0.00 0.00	2.00 0.05	2.00 0.10	2.00 0.30	3.00 0.05	3.00 0.10	3.00 0.30	8.00 0.05	8.00 0.10	8.00 0.30	λ_w m_w
pp	0.95	0.65	0.65	0.67	0.70	0.65	0.66	0.71	0.66	0.67	0.74	
pp	0.90	0.97	0.96	0.97	0.97	0.97	0.97	0.97	0.96	0.97	0.98	
pp	0.85	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	
adf	0.95	0.46	0.47	0.47	0.49	0.47	0.47	0.49	0.47	0.47	0.51	
adf	0.90	0.87	0.87	0.87	0.89	0.88	0.88	0.90	0.87	0.89	0.91	
adf	0.85	0.95	0.96	0.96	0.96	0.95	0.95	0.96	0.96	0.96	0.96	
jo-e	0.95	0.38	0.38	0.38	0.39	0.38	0.39	0.39	0.38	0.38	0.43	
jo-e	0.90	0.89	0.89	0.90	0.92	0.89	0.90	0.92	0.90	0.91	0.95	
jo-e	0.85	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.98	
jo-t	0.95	0.36	0.36	0.36	0.37	0.37	0.36	0.37	0.36	0.36	0.39	
jo-t	0.90	0.85	0.85	0.86	0.89	0.86	0.86	0.89	0.86	0.87	0.93	
jo-t	0.85	0.96	0.96	0.96	0.97	0.96	0.96	0.97	0.96	0.97	0.98	
ers-p	0.95	0.58	0.59	0.61	0.65	0.60	0.61	0.66	0.59	0.61	0.68	
ers-p	0.90	0.82	0.83	0.84	0.85	0.83	0.84	0.86	0.83	0.84	0.87	
ers-p	0.85	0.91	0.90	0.90	0.91	0.90	0.91	0.91	0.90	0.91	0.92	
ers-d	0.95	0.57	0.56	0.58	0.62	0.58	0.58	0.64	0.57	0.60	0.65	
ers-d	0.90	0.79	0.80	0.80	0.83	0.79	0.81	0.83	0.80	0.80	0.85	
ers-d	0.85	0.86	0.87	0.87	0.89	0.87	0.87	0.89	0.87	0.88	0.90	
sp-r	0.95	0.47	0.47	0.47	0.50	0.47	0.47	0.50	0.47	0.48	0.51	
sp-r	0.90	0.85	0.85	0.85	0.86	0.84	0.85	0.87	0.84	0.86	0.88	
sp-r	0.85	0.95	0.95	0.94	0.94	0.95	0.94	0.94	0.95	0.95	0.94	
hurst	0.95	0.34	0.34	0.34	0.32	0.34	0.33	0.32	0.34	0.34	0.35	
hurst	0.90	0.61	0.61	0.60	0.60	0.60	0.60	0.60	0.59	0.60	0.62	
hurst	0.85	0.74	0.74	0.74	0.74	0.75	0.74	0.75	0.74	0.74	0.76	
bvr	0.95	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.42	0.44	0.45	
bvr	0.90	0.70	0.70	0.69	0.71	0.69	0.69	0.70	0.69	0.70	0.72	
bvr	0.85	0.81	0.82	0.82	0.82	0.82	0.82	0.83	0.82	0.83	0.84	
pgff	0.95	0.68	0.69	0.68	0.73	0.68	0.69	0.74	0.68	0.70	0.78	
pgff	0.90	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.98	
pgff	0.85	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.99	

4.4. Results Type VI

Results for Type VI are given for size in Table 14 and for power in Table 15. Size: As size varies only between 0.04 and 0.08 for all tests, the analysis may be focused on power.

Table 14. Size of cointegration tests for Monte Carlo settings of Type VI (see Subsection 3.4.2).

Test	$\phi_L = \phi_U$	4.00 0.05	4.00 0.10	4.00 0.45	6.00 0.05	6.00 0.10	6.00 0.45	12.00 0.05	12.00 0.10	12.00 0.45	λ_p m_p
pp	1.00	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.06	
adf	1.00	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	
jo-e	1.00	0.07	0.07	0.08	0.07	0.07	0.07	0.07	0.07	0.07	
jo-t	1.00	0.06	0.06	0.07	0.06	0.06	0.07	0.06	0.07	0.07	
ers-p	1.00	0.05	0.05	0.05	0.06	0.05	0.05	0.05	0.05	0.05	
ers-d	1.00	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	
sp-r	1.00	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.05	0.04	
hurst	1.00	0.05	0.05	0.06	0.05	0.05	0.06	0.05	0.05	0.05	
bvr	1.00	0.05	0.05	0.04	0.05	0.05	0.05	0.04	0.05	0.04	
pgff	1.00	0.07	0.06	0.05	0.06	0.06	0.06	0.06	0.06	0.05	

Table 15. Power of cointegration tests for Monte Carlo settings of Type VI, as described in Subsection 3.4.2.

Test	$\phi_L = \phi_U$	0.00 0.00	4.00 0.05	4.00 0.10	4.00 0.45	6.00 0.05	6.00 0.10	6.00 0.45	12.00 0.05	12.00 0.10	12.00 0.45	λ_p m_p
pp	0.95	0.65	0.45	0.34	0.14	0.41	0.28	0.11	0.32	0.20	0.08	
pp	0.90	0.97	0.74	0.56	0.19	0.68	0.46	0.13	0.56	0.33	0.08	
pp	0.85	0.98	0.79	0.62	0.22	0.74	0.52	0.15	0.62	0.37	0.08	
adf	0.95	0.46	0.32	0.25	0.12	0.28	0.20	0.09	0.23	0.15	0.07	
adf	0.90	0.87	0.58	0.41	0.16	0.51	0.33	0.11	0.40	0.22	0.07	
adf	0.85	0.95	0.66	0.47	0.17	0.58	0.38	0.11	0.45	0.25	0.07	
jo-e	0.95	0.38	0.24	0.18	0.11	0.21	0.15	0.09	0.17	0.12	0.07	
jo-e	0.90	0.89	0.55	0.38	0.15	0.48	0.30	0.11	0.36	0.20	0.08	
jo-e	0.85	0.97	0.68	0.48	0.17	0.60	0.38	0.12	0.45	0.24	0.08	
jo-t	0.95	0.36	0.22	0.17	0.11	0.21	0.14	0.08	0.16	0.11	0.07	
jo-t	0.90	0.85	0.53	0.36	0.14	0.45	0.28	0.10	0.33	0.18	0.08	
jo-t	0.85	0.96	0.65	0.46	0.17	0.58	0.36	0.11	0.43	0.22	0.08	
ers-p	0.95	0.58	0.39	0.28	0.12	0.35	0.24	0.08	0.27	0.17	0.06	
ers-p	0.90	0.82	0.56	0.40	0.13	0.50	0.33	0.09	0.39	0.23	0.07	
ers-p	0.85	0.91	0.62	0.45	0.15	0.55	0.37	0.09	0.44	0.26	0.07	
ers-d	0.95	0.57	0.37	0.26	0.11	0.33	0.22	0.08	0.26	0.17	0.06	
ers-d	0.90	0.79	0.52	0.37	0.13	0.47	0.31	0.09	0.36	0.20	0.06	
ers-d	0.85	0.86	0.57	0.40	0.13	0.50	0.34	0.09	0.40	0.22	0.06	
sp-r	0.95	0.47	0.32	0.23	0.09	0.27	0.19	0.07	0.22	0.13	0.05	
sp-r	0.90	0.85	0.53	0.36	0.12	0.46	0.30	0.08	0.34	0.19	0.05	
sp-r	0.85	0.95	0.62	0.40	0.13	0.52	0.31	0.08	0.39	0.20	0.05	
hurst	0.95	0.34	0.22	0.16	0.10	0.20	0.13	0.08	0.15	0.10	0.06	
hurst	0.90	0.61	0.30	0.21	0.11	0.26	0.17	0.09	0.19	0.11	0.07	
hurst	0.85	0.74	0.32	0.22	0.13	0.27	0.17	0.09	0.18	0.11	0.06	
bvr	0.95	0.43	0.25	0.19	0.08	0.23	0.15	0.06	0.18	0.11	0.05	
bvr	0.90	0.70	0.34	0.22	0.09	0.28	0.18	0.07	0.20	0.11	0.05	
bvr	0.85	0.81	0.35	0.23	0.09	0.29	0.17	0.06	0.21	0.11	0.05	
pgff	0.95	0.68	0.47	0.35	0.13	0.42	0.29	0.10	0.34	0.20	0.07	
pgff	0.90	0.97	0.77	0.58	0.20	0.71	0.51	0.13	0.60	0.35	0.08	
pgff	0.85	0.98	0.82	0.65	0.22	0.76	0.56	0.15	0.67	0.42	0.09	

Power: Type VI introduces non-reversible jumps through a compound Poisson process with expected number of jumps λ_p per trading day and median jumps size m_p in absolute value. The superposition of the compound Poisson process with the previously discussed STAR process of Equation (3) results in a regime shift of the mean equation with each jump. In other words, each jump

technically leads to a rupture of the previous cointegration relationship and the establishment of a new one at another mean level. The shift in mean depends on the size of the jump, which is driven by its variance. From Table 15, we see that the power decreases with increasing expected number of jumps and increasing jump size in absolute value. In the following, we differentiate the results by jump size: Small jumps: Already smaller jumps with median size of 0.05 have an expressed effect on the power, considering that this jump size is already more than 15-times higher than the median size of the average GARCH-innovation of an average cointegration residual [8]. For example, the power of the PP test declines from 0.65 to 0.41 for $\lambda_p = 6$ and to 0.32 for $\lambda_p = 12$. Medium jumps: Medium jumps with a median size of 0.10 are even more detrimental. An expected value of six jumps leads to a deterioration in power to 0.28 for the PP test, i.e., a relative slump of 57 percent. This scenario corresponds to a cointegration residual exhibiting typically-sized non-reversible jumps at a typical frequency; see Table 4 for the quantile estimates. Higher jump frequencies further deteriorate the power, which is fair, considering that cointegration relationships only exist for a shorter time until the next mean shift occurs. Large jumps: Large jumps with standard deviation of 0.45 already do significant damage if they occur just four times a day in expectation. We see that power deteriorates by 80 percent across almost all tests compared to the case with no jumps. Increasing the frequency to six or even 12 jumps deteriorates the power to levels close to the Type I errors, rendering the tests ineffective.

These results are paramount, meaning that we are able to detect a (time-varying) cointegration relationship in financial market data with classical tests, even if its mean varies over time either at low frequencies or with only limited shifts in mean. Furthermore, in these settings, relative to the other tests, PGFF and PP are still recommended.

5. Conclusions

We conduct an in-depth analysis of the size and power properties of ten contemporary cointegration tests, given the stylized facts of high-frequency trade data. We make several contributions to the existing literature.

First, we test for different stylized facts on a large database of one-minute return data. In line with the existing literature, we find non-normality, ARCH effects, jumps and evidence of further nonlinearities. The same stylized facts are established for the cointegration residuals processes linking some of the DAX 30 constituents. To our knowledge, the latter analysis has not yet been performed in the literature. It is interesting to see that the identified cointegration residuals that are classified as stationary by the Johansen trace test still exhibit non-normalities, ARCH effects, further nonlinearities, as well as jumps. In particular, we find a median number of three reversible and six non-reversible jumps with median sizes of 0.10 and 0.11 EUR in absolute value. One could have assumed that these stylized facts cancel each other out, just as the nonstationary components. The data prove this to be wrong.

Second, we propose an innovative approach for simulating stock prices following the stationary bootstrap of [46]. This procedure allows for simulating stock prices in a high-frequency setting, while retaining the majority of their stylized facts.

Third, we suggest six different processes for cointegration residuals based on the current literature. These processes accommodate the above-mentioned stylized facts in a staggered approach. To our knowledge, the application of an MR(3)-STAR(1)-GARCH(1,1) model is a novelty in this context. It provides significantly higher flexibility for modeling relative-value arbitrage strategies and may also be an interesting direction for further research in that respect.

Fourth, we have performed Monte Carlo simulations to assess the power and size properties of ten different cointegration tests. We find that both non-normality and GARCH effects have none or only a

[8] For such an estimation, the median of the absolute value of t-distributed innovations with five degrees of freedom and standard deviation of 0.0059 can be calculated as 0.0033. See Table 3 for the relevant parameter values used in this estimation.

marginal impact on size and power properties. We conclude that non-normality and ARCH effects, as present in high-frequency financial data, still allow for cointegration testing with very respectable size and power properties.

STAR nonlinearities have an adverse effect on the power, which decreases with rising threshold levels and increasing values for gamma. The latter is driven by the fact that increasing parameter values push a higher share of observations in the nonstationary middle regime. However, with the plausible assumption of relative-value arbitrage kicking in at threshold levels estimated via bid-ask spreads, the power of cointegration tests still remains acceptable.

The effect of jumps differs by their nature: Reversible jumps, which may be driven by uninformed buying or liquidity shortages, do not have an adverse effect on the power or actually increase it. This phenomenon is easily explained by the fact that a reversible jump is just a temporary disruption, which usually pushes the observations in the stationary outer regimes. On the contrary, non-reversible jumps disrupt the power with an increasing rate of occurrence and increasing jump size. Assuming typical jump rates and sizes for our cointegration residuals, power deteriorates to levels of approximately 0.29 for PGFF and 0.28 for PP for AR(1)-coefficients of 0.95.

We conclude that contemporary cointegration tests may be applied with caution in the high-frequency setting of one-minute stock return data. In particular, we recommend pre-evaluating the cointegration relationship with the procedure described in this paper and in [25] to assess the size and magnitude of jumps. Then, our tables provide a sensible orientation, whether cointegration testing has acceptable power. Across all of the above-mentioned stylized facts, the PGFF and the PP tests exhibit the most favorable power properties. These tests generally dominate all other tests in all settings, where PGFF is slightly better than PP.

Acknowledgments: We are grateful to Matthew Clegg, Thomas Fischer, Ingo Klein, Helmut Lütkepohl, Benedikt Mangold, Michael McAleer, Johannes Stübinger, the participants of the CEQURA Conference on Advances in Financial and Insurance Risk Management 2015 and two anonymous referees for many helpful discussions and suggestions on this topic.

Author Contributions: The research method was conceived of by C.K. The data were provided by Deutsche Börse, transferred to the database KDB by K.H. and preprocessed in R by C.K. The analyses were performed by C.K. and reviewed by C.K. and K.H. The paper was initially drafted by C.K. and revised, refined and finalized by C.K. and KH.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Elyasiani, E.; Kocagil, A.E. Interdependence and dynamics in currency futures markets: A multivariate analysis of intraday data. *J. Bank. Financ.* **2001**, *25*, 1161–1186.
2. Hasbrouck, J. Intraday price formation in U.S. equity index markets. *J. Finance* **2003**, *58*, 2375–2399.
3. Dunis, C.L.; Giorgioni, G.; Laws, J.; Rudy, J. *Statistical Arbitrage and High-Frequency Data with an Application to Eurostoxx 50 Equities*; Working Paper, Liverpool Business School: Liverpool, UK, 2010.
4. Pati, P.C.; Rajib, P. Intraday return dynamics and volatility spillovers between NSE S&P CNX Nifty stock index and stock index futures. *Appl. Econ. Lett.* **2011**, *18*, 567–574.
5. Yang, J.; Yang, Z.; Zhou, Y. Intraday price discovery and volatility transmission in stock index and stock index futures markets: Evidence from China. *J. Futur. Mark.* **2012**, *32*, 99–121.
6. Johansen, S. Statistical analysis of cointegration vectors. *J. Econ. Dyn. Control* **1988**, *12*, 231–254.
7. Johansen, S.; Juselius, K. Maximum likelihood estimation and inference on cointegration–With applications to the demand for money. *Oxf. Bull. Econ. Stat.* **1990**, *52*, 169–210.
8. Johansen, S. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **1991**, *59*, 1551–1580.
9. Kremers, J.J.; Ericsson, N.R.; Dolado, J.J. The power of cointegration tests. *Oxf. Bull. Econ. Stat.* **1992**, *54*, 325–348.
10. Haug, A.A. Tests for cointegration a Monte Carlo comparison. *J. Econom.* **1996**, *71*, 89–115.
11. Hubrich, K.; Lütkepohl, H.; Saikkonen, P. A review of systems cointegration tests. *Econom. Rev.* **2001**, *20*, 247–318.

12. Boswijk, H.P.; Lucas, A.; Taylor, N. *A Comparison of Parametric, Semi-Nonparametric, Adaptive, and Nonparametric Cointegration Tests*; Tinbergen Institute Discussion Papers: Amsterdam, The Netherlands, 1999.
13. Rahbek, A.; Hansen, E.; Dennis, J.G. *ARCH Innovations and Their Impact on Cointegration Rank Testing*; Working Paper, Department of Theoretical Statistics, University of Copenhagen: Copenhagen, Denmark, 2002.
14. Cavaliere, G.; Rahbek, A.; Taylor, A.M.R. Testing for co-integration in vector autoregressions with non-stationary volatility. *J. Econom.* **2010**, *158*, 7–24.
15. Cavaliere, G.; Rahbek, A.; Taylor, A.M.R. Cointegration rank testing under conditional heteroskedasticity. *Econometr. Theor.* **2010**, *26*, 1719–1760.
16. Balke, N.S.; Fomby, T.B. Threshold cointegration. *Int. Econ. Rev.* **1997**, *38*, 627–645.
17. Hansen, B.E.; Seo, B. Testing for two-regime threshold cointegration in vector error-correction models. *J. Econom.* **2002**, *110*, 293–318.
18. Seo, M. Bootstrap testing for the null of no cointegration in a threshold vector error correction model. *J. Econom.* **2006**, *134*, 129–150.
19. Alexander, C. Optimal hedging using cointegration. *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* **1999**, *357*, 2039–2058.
20. Alexander, C.; Dimitriu, A. Indexing and statistical arbitrage. *J. Portf. Manag.* **2005**, *31*, 50–63.
21. Gatev, E.; Goetzmann, W.N.; Rouwenhorst, K.G. Pairs trading: Performance of a relative-value arbitrage rule. *Rev. Financial Stud.* **2006**, *19*, 797–827.
22. Ingersoll, J.E. *Theory of Financial Decision Making*; Rowman & Littlefield: Totowa, NJ, USA, 1987.
23. Chen, Z.; Knez, P.J. Measurement of market integration and arbitrage. *Rev. Financial Stud.* **1995**, *8*, 287–325.
24. Andrade, S.; Di Pietro, V.; Seasholes, M. *Understanding the Profitability of Pairs Trading*; Working Paper, UC Berkeley, Northwestern University: Evanston, IL, USA, 2005.
25. Jondeau, E.; Lahaye, J.; Rockinger, M. Asymmetry in the price impact of trades in an high-frequency microstructure model with jumps. *J. Bank. Finance.* **2015**, *61, Supplement 2*, 205–224.
26. Trapletti, A.; Hornik, K.; LeBaron, B. tseries: Time series analysis and computational finance. Available online: https://cran.r-project.org/web/packages/tseries/index.html (accessed on 6 February 2017).
27. R Development Core Team. R: A language and environment for statistical computing. Available online: https://www.r-project.org/ (accessed on 6 February 2017).
28. Graves, S. FinTS: Companion to Tsay (2005) Analysis of financial time series. Available online: https://rdrr.io/cran/FinTS/ (accessed on 6 February 2017).
29. Barndorff-Nielsen, O.E.; Shephard, N. Econometrics of testing for jumps in financial economics using bipower variation. *J. Financial Econom.* **2006**, *4*, 1–30.
30. Boudt, K.; Cornelissen, J.; Payseur, S.; Nguyen, G.; Schermer, M. High frequency: Tools for high frequency data analysis. Available online: https://rdrr.io/cran/highfrequency/ (accessed on 6 February 2017).
31. Chan, K.S.; Ripley, B. TSA: Time series analysis. Available online: https://rdrr.io/cran/TSA/ (accessed on 6 February 2017).
32. Luukkonen, R.; Saikkonen, P.; Teräsvirta, T. Testing linearity against smooth transition autoregressive models. *Biometrika* **1988**, *75*, 491–499.
33. Ghalanos, A. Twinkle: Dynamic smooth transition ARMAX models. Available online: http://past.rinfinance.com/agenda/2014/talk/AlexiosGhalanos.pdf (accessed on 6 February 2017).
34. Ghalanos, A. Rugarch: Univariate GARCH models. Available online: https://rdrr.io/cran/rugarch/ (accessed on 6 February 2017).
35. Engle, R.F.; Sokalska, M.E. Forecasting intraday volatility in the US equity market: Multiplicative component GARCH. *J. Financial Econom.* **2012**, *10*, 54–83.
36. Herrmann, K.; Teis, S.; Yu, W. *Components of Intraday Volatility and Their Prediction at Different Sampling Frequencies with Application to DAX and BUND Futures*; IWQW Discussion Paper Series, University of Erlangen-Nürnberg: Erlangen, Germany, 2014.
37. Cont, R. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantit. Finance* **2001**, *1*, 223–236.
38. Pfaff, B. *Analysis of Integrated and Cointegrated Time Series with R*; Springer: New York, NY, USA, 2008.
39. Dickey, D.A.; Fuller, W.A. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* **1979**, *74*, 427–431.

40. Tsay, R.S. *Multivariate Time Series Analysis: With R and Financial Applications*; Wiley Series in Probability and Statistics; John Wiley & Sons: Hoboken, NJ, USA, 2014.
41. Doornik, J.A. Approximations to the asymptotic distributions of cointegration tests. *J. Econ. Surv.* **1998**, *12*, 573–593.
42. Yang, F. CommonTrend: Extract and plot common trends from a cointegration system. Calculate p-value for Johansen statistics. Available online: https://rdrr.io/cran/CommonTrend/ (accessed on 6 February 2017)
43. Hill, B.M. A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **1975**, *3*, 1163–1174.
44. Breitung, J. Rank tests for nonlinear cointegration. *J. Bus. Econ. Stat.* **2001**, *19*, 331–340.
45. Clegg, M. On the persistence of cointegration in pairs trading. Available online: https://rdrr.io/cran/CommonTrend/ (accessed on 6 February 2017)
46. Politis, D.N.; Romano, J.P. The stationary bootstrap. *J. Am. Stat. Assoc.* **1994**, *89*, 1303–1313.
47. Azzalini, A.; Capitanio, A. *The Skew-Normal and Related Families*; Cambridge University Press: Cambridge, UK, 2014.
48. Hong, G.; Susmel, R. *Pairs-Trading in the Asian ADR Market*; Working Paper, University of Houston: Houston, TX, USA, 2003.
49. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *J. Econom.* **1986**, *31*, 307–327.
50. Nelson, D.B. Stationarity and persistence in the GARCH(1,1) model. *Econometr. Theor.* **1990**, *6*, 318–334.
51. Glosten, L.R.; Jagannathan, R.; Runkle, D.E. On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Finance* **1993**, *48*, 1779–1801.
52. Nelson, D.B. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **1991**, *59*, 347–370.
53. McAleer, M.; Chan, F.; Hoti, S.; Lieberman, O. Generalized autoregressive conditional correlation. *Econometr. Theor.* **2008**, *24*, 1554–1583.
54. Tong, H. On a threshold model. In *Pattern Recognition and Signal Processing*; Nato Advanced Study Institutes: Series E, Applied Sciences; Chen, C.H., Ed.; Sijthoff & Noordhoff: Alphen aan den Rijn, The Netherlands, 1978; Volume 29.
55. Teräsvirta, T. Specification, estimation, and evaluation of smooth transition autoregressive models. *J. Am. Stat. Assoc.* **1994**, *89*, 208–218.
56. van Dijk, D.; Franses, P.H. Modeling multiple regimes in the business cycle. *Macroecon. Dyn.* **1999**, *3*, 311–340.
57. Chan, F.; McAleer, M. Maximum likelihood estimation of STAR and STAR-GARCH models: Theory and Monte Carlo evidence. *J.Appl. Econom.* **2002**, *17*, 509–534.
58. Chan, F.; McAleer, M. Estimating smooth transition autoregressive models with GARCH Errors in the presence of extreme observations and outliers. *Appl. Financial Econ.* **2003**, *13*, 581–592.
59. Chan, K.S.; Petruccelli, J.D.; Tong, H.; Woolford, S.W. A multiple-threshold AR(1) model. *J. Appl. Probab.* **1985**, *22*, 267.
60. Ross, S.M. *Stochastic Processes*, 2nd ed.; Wiley Series in Probability and Statistics. Probability and Statistics; Wiley: New York, NY , USA, 1996.
61. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: London, UK; New York, NY , USA, 1986.
62. Sheather, S.J.; Jones, M.C. A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Stat. Soc. Ser. B* **1991**, 683–690.
63. Dickey, D.A.; Fuller, W.A. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **1981**, *49*, 1057–1072.
64. Perron, P. Trends and random walks in macroeconomic time series. *J. Econ. Dyn. Control* **1988**, *12*, 297–332.
65. Phillips, P.C.B.; Ouliaris, S. Asymptotic properties of residual based tests for cointegration. *Econometrica* **1990**, *58*, 165–193.
66. Pantula, S.G.; Gonzalez-Farias, G.; Fuller, W.A. A comparison of unit-root test criteria. *J. Bus. Econ. Stat.* **1994**, *12*, 449–459.
67. Breitung, J. Nonparametric tests for unit roots and cointegration. *J. Econom.* **2002**, *108*, 343–363.
68. Breitung, J.; Taylor, A.M.R. Corrigendum to "Nonparametric tests for unit roots and cointegration" [J. Econom. 108 (2002) 343–363]. *J. Econom.* **2003**, *117*, 401–404.
69. Elliott, G.; Rothenberg, T.J.; Stock, J.H. Efficient tests for an autoregressive unit root. *Econometrica* **1996**, *64*, 813–836.

70. Schmidt, P.; Phillips, P.C.B. LM tests for a unit root in the presence of deterministic trends. *Oxf. Bull. Econ. Stat.* **1992**, *54*, 257–287.
71. Cavaliere, G. Testing the unit root hypothesis using generalized range statistics. *Econom. J.* **2001**, *4*, 70–88.
72. Wuertz, D.; Taqqu, M.S. fArma: ARMA time series modelling. Available online: https://rdrr.io/cran/fArma/ (accessed on 6 February 2017).
73. Krauss, C. Statistical arbitrage pairs trading strategies: Review and outlook. *J. Econ. Surv. Forthcom.* **2016**.
74. Alexander, C. *Practical Financial Econometrics*, Reprinted March 2009 ed.; Volume 2—Market Risk Analysis; Wiley: Chichester, UK, 2009.