IWQW

Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung

Diskussionspapier
Discussion Papers

No. 11/2015

# On the power and size properties of cointegration tests in the light of high-frequency stylized facts

Christopher Krauss
University of Erlangen-Nürnberg

Klaus Herrmann
University of Erlangen-Nürnberg

Stefan Teis
Deutsche Börse AG

ISSN 1867-6707

Friedrich-Alexander-Universität
IWQW
Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung

# On the power and size properties of cointegration tests in the light of high-frequency stylized facts


**Christopher Krauss**
Department of Statistics and Econometrics
University of Erlangen-Nürnberg, Nürnberg

**Klaus Herrmann**
Department of Statistics and Econometrics
University of Erlangen-Nürnberg, Nürnberg

**Stefan Teis**
Deutsche Börse AG, Eschborn*


Wednesday 30th September, 2015


**Abstract**

This paper first establishes a selection of stylized facts for high-frequency cointegration processes in the European equity market. Empirical evidence is given by one minute-binned transaction data of all DAX 30 constituents as traded on Deutsche Börse's Xetra market in 2014. A methodology is introduced to simulate cointegrated stock pairs, following none, some or all of the discussed stylized facts. In particular, AR(1), AR(1)-GARCH(1,1) and multiple regime STAR(1)-GARCH(1,1) processes are used to model the cointegration relationship. Furthermore, this cointegration relationship is contaminated with jumps. Based on these processes, the power and size properties of ten contemporary cointegration tests are assessed. We provide an economic interpretation of our approach by relating cointegration to relative-value arbitrage strategies in near-efficient markets. Quintessentially, we find that in a high-frequency setting typical for stock price data, selected cointegration tests still exhibit high power. Especially the Phillips-Perron and the Pantula, Gonzalez-Farias and Fuller tests perform best at very limited size distortions.



**Keywords:** Cointegration testing, high-frequency, stylized facts, power analysis, conditional heteroskedasticity, smooth transition autoregressive models



*The views expressed here are those of the authors and not necessarily those of affiliated institutions. Please address correspondence to Christopher Krauss per email: christopher.krauss@fau.de

# 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 cointegrating relationships also in high-frequency financial market data [see, among others, Elyasiani and Kocagil (2001), Hasbrouck (2003), Dunis et al. (2010), Pati and Rajib (2011), Yang et al. (2012)]. However, it remains unclear whether standard cointegration tests applied in these studies are truly robust against the specifics of high-frequency settings. In our data set, we find several stylized facts with potential impact on power and size properties of contemporary cointegration tests. Most notably, the one minute return data are highly non-normal, exhibit ARCH effects with intraday seasonalities and jumps. Additionally we find evidence for nonlinear dependencies - even after applying AR(1)-GARCH(1,1) filtrations, see section 2.

Monte Carlo studies of 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 Johansen (1988), Johansen and Juselius (1990), and Johansen (1991). The objective is to compare power and size properties across a wide range of different cointegration tests. Common references are Kremers et al. (1992), Haug (1996) and Hubrich et al. (2001). 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. Boswijk et al. (1999) consider Student's t-distributions with 3 degress of freedom, a truncated Cauchy distribution, Gaussian mixtures and others. Furthermore, bivariate ARCH and GARCH processes are also used to model the innovations. Rahbek et al. (2002) analyze the impact of ARCH innovations from a multivariate BEKK-ARCH DGP on the trace test. Cavaliere et al. (2010a) and Cavaliere et al. (2010b) follow a similar approach. The third group covers the concept of threshold cointegration. The seminal reference is by Balke and Fomby (1997), who were the first to introduce threshold nonlinearities in cointegrating relationships and analyze the power of conventional and enhanced tests in this setting. Further studies and improved tests followed by Hansen and Seo (2002) and Seo (2006).

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 threefold. 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 cointegrating relationships, reflecting different stages of high-frequency stylized facts, i.e., non-normality, GARCH effects, threshold nonlinearities, jumps and regime shifts. Third, ten contemporary cointegration tests are examined in a Monte Carlo simulation with respect to their power and size properties for sample sizes of 510 minutes, i.e., one trading day at Xetra. We find that (1) non-normal innovations and GARCH effects have only 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 increase and regime shifts strongly deteriorate the power. We give an economic interpretation of the models and derive that in the settings to be expected for stock data, the power - especially of Phillips-Perron and Pantula, Gonzalez-Farias and Fuller tests - is still very high.

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 four 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 the light of the existing literature. An economic interpretation of our results is provided in section 5. Finally, section 6 concludes and summarizes directions for further research.

## 2. Data sample and its stylized facts

The data applied in this study consists of all transactions of the DAX 30 constituents traded on Xetra in continuous trading throughout the year 2014. Transaction data is aggregated to one-minute 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 is available as well. We have sanitized 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 the light of well-known stylized facts for financial data.

For each of the 249 trading days, we perform a set of tests[1] for each of the 30 constituents. 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[2]. 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 Engle and Sokalska (2012). 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 index in Herrmann et al. (2014), see figure 1.

Furthermore, the Barndorff-Nielsen and Shephard test indicates jumps. These findings are in line with the existing literature, see for example Cont (2001) for an account of relevant stylized facts in financial markets.

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

Table 1: Stylized facts stock return data. $249 \cdot 30$ processes analyzed. Share of tests with p-value less than 0.05.

In a second step, we check if the stylized facts of the cointegration processes of potentially

[1]The Jarque-Bera test, the BDS test and the Teräsvirta test are implemented in the R package `tseries` by Trapletti and Hornik (2013). The Box test is part of the R package `stats` by the R Core Team (2014). Engle's ARCH test is implemented in the R package `FinTS` by Graves (2014). The Barndorff-Nielsen and Shephard test of Barndorff-Nielsen and Shephard (2006) is implemented in the R package `highfrequency` by Boudt et al. (2014). The Tsay test is implemented in the R package `TSA` by Chan and Ripley (2012). The Luukkonen test of Luukkonen et al. (1988) is implemented in the R package `twinkle` by Ghalanos (2014b).

[2]We use the R package `rugarch` of Ghalanos (2014a) to fit AR(1)-GARCH(1,1) models.

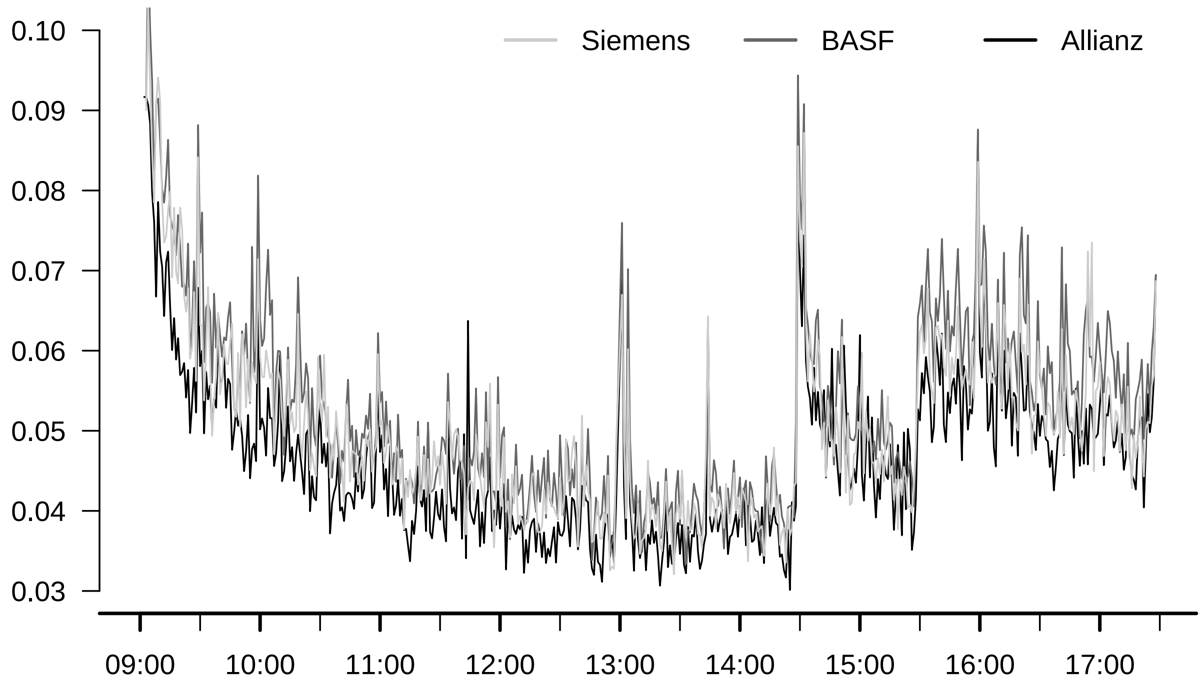


Figure 1: Average realized volatility per one minute of the trading day (in CET) for selected stocks in 2014.

cointegrated stock pairs differ from the results above. We identify the cointegration processes with the Johansen trace test at the five percent significance level. This 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 time 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 Pfaff (2008) to examine all $30 \cdot 29/2 = 435$ pairs of stocks for each of the 249 full trading days (this amounts to a total of 108315 tests). First, we use the Augmented Dickey-Fuller test of Dickey and Fuller (1979) to check if the individual series are I(1) at the 1 percent significance level. If yes, 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 Doornik (1998) and implemented in R by Yang (2013). A total of $435 \cdot 249 = 108315$ processes are analyzed and 41069 are identified as cointegrated at the 5 percent significance level. Finally, we run the same set of tests to see if the stylized facts of the cointegration processes differ from the results of table 1. The results are summarized in table 2. We see that the cointegration processes exhibit similar stylized facts as the return data, i.e., non-normality, ARCH effects, jumps and further evidence of nonlinearities.

| **Type** | **Test** | **Raw data** | **AR(1)** | **AR(1)-GARCH(1,1)** |
|---|---|---|---|---|
| Nonnormality | Jarque-Bera test | 0.89 | 0.99 | 0.99 |
| | Box test | 1.00 | 0.49 | 0.05 |
| ARCH effects | Engle's ARCH test | 1.00 | 0.76 | 0.07 |
| | BDS test | 1.00 | 0.66 | 0.49 |
| | Tsay test | 0.45 | 0.60 | 0.27 |
| | Luukkonen test | 0.55 | 0.39 | 0.15 |
| Nonlinearity | Teräsvirta test | 0.55 | 0.36 | 0.16 |
| Jumps | BNS test | 1.00 | 0.98 | 1.00 |

Table 2: Stylized facts cointegration processes. 41069 processes analyzed. Share of tests with p-value less than 0.05.

# 3. Methodology

The suggested methodology to assess the power and size properties of ten cointegration tests is based on four steps. In the first step, we simulate high-frequency stock price data while retaining their stylized facts. In the second step, we simulate the cointegration process, exhibiting different stylized facts. In the third step, the cointegrating relationship and hence the price of the second stock are defined. In the fourth step, a Monte Carlo simulation is performed to obtain the power and size properties of the tests.

## 3.1 Simulation of stock prices

For the Monte Carlo simulation of cointegration processes in the third step, the $249 \cdot 30 = 7470$ daily price time series in our data set are not sufficient. Hence, already in step one, we need a methodology to effectively simulate high-frequency price time series, while retaining their stylized facts as accurately as possible. We suggest a variant of the stationary bootstrap developed by Politis and Romano (1994):

1. Set the start index $st$ equal to one, 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 random block length $l$ from a geometric distribution with expected value of four. This value 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. We prefer introducing a higher level of diversity.
5. Choose block of length $l$, consisting of returns from stock $s$ from day $d$ for indices $i \in [st, ..., st + l]$. Copy these returns in vector $v$ at positions $[st, ..., st + l]$.
6. Update $st$ with $st + l + 1$. 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.[3] Accumulate the return vector $v$ to a price time series.

This procedure has several advantages. First, it ensures that the return time series remains stationary, see Politis and Romano (1994). Second, since blocks of a random length with 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. 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 processes

In order to reflect the identified stylized facts, five DGPs are considered for the simulation of the cointegration processes: A simple AR(1)-process, an AR(1)-GARCH(1,1) process, an MR(3)-STAR(1)-GARCH(1,1)[4] process, an MR(3)-STAR(1)-GARCH(1,1) process with jumps and an MR(3)-STAR(1)-GARCH(1,1) process with jumps in mean. Each of these processes can be generated either with normally or t-distributed innovations. Tables 5 provides an overview of all models

[3] The starting price has no impact on power as all cointegration tests are invariant towards the absolute price level.
[4] Multiple regime smooth transition autoregressive process

and parameters applied, section 5 gives an economic interpretation of models and argues for the plausibility of chosen parameters. In the following, we will briefly discuss the relevant statisticial properties of each DGP.

**3.2.1 Autoregressive model:** Let $u_t$ denote the cointegration residual at time $t$, such that the AR(1) process of returns with coefficient $\phi_1$ and scale parameter $\sigma$ may be defined as

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

such that the innovations are either normally distributed with mean zero and standard deviation of 1 or t-distributed with mean zero, $\nu$ degrees of freedom and a shape parameter $\omega$. Following Azzalini and Capitanio (2014), the shape parameter $\omega$ is chosen in such a way, that the standard deviation of the t-distribution equals 1. For this purpose, $\nu$ must be strictly greater than 2. For all numerical implementations involved, a lower bound of 2.1 for $\nu$ 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.

**3.2.2 Generalized autoregressive conditional heteroscedasticity model:** The GARCH model of Bollerslev (1986) generalizes equation (1) to an AR(1)-GARCH(1,1) model, accounting for time-dependency in the scale parameter as

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

where the time-dependency is determined by parameters $\alpha_0$, $\alpha_1$ and $\beta_1$. The innovations $\epsilon_t$ may again be either normally or t-distributed in the same manner as for the autoregressive model. The mean follows the same stationarity condition. The variance is stationary, if $\alpha_1 + \beta_1 < 1$. This process models the stylized facts of non-normality and of ARCH effects.

**3.2.3 Multiple regime smooth transition autoregressive model:** Terasvirta (1994) has first introduced STAR models, which have been extended by van Dijk and Franses (1999) 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 (1999) with smoothness parameter $\gamma$ and threshold $c$. $\sigma_t$ is again dependent on time and follows a GARCH(1,1) process as in equation (2). 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 our knowledge, no comprehensive stationarity conditions have been developed yet. However, for our purposes, we will just discuss the following simplified parameter constellations:

*Case A:* Let $\phi_1$ be equal to one, and $c_1 < c_2$. If $\phi_2$ and $\phi_3$ are equal to zero, we wind up having only one regime - a random walk, which clearly is nonstationary.

*Case B:* If $\phi_1$ is element of $[0.95, 0.90, 0.85]$, $\phi_2$ is element of $[0.05, 0.10, 0.15]$ and $\phi_3$ is element of $[-0.05, -0.10, -0.15]$, the middle regime corresponding to $\phi_2$ is nonstationary, but the outer regimes become stationary. The latter is due to a mix effect. For values sufficiently[5] less than $c_1$, the logistic functions $F_{12}$ and $F_{23}$ equal to zero, so the lower regime is stationary with a mixed coefficient $\phi_L$ element of $[0.95, 0.90, 0.85]$. For values sufficiently greater than $c_1$ but sufficiently smaller than $c_2$, the logistic function $F_{12}$ is equal to one but $F_{23}$ is still at zero. Hence, the mixed coefficient for the middle regime $\phi_M$ is element of $[1.00, 1.00, 1.00]$, indicating nonstationarity. For values sufficiently greater than $c_2$, $F_{23}$ also equals to one, leading to a mixed coefficient $\phi_U$ 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. Also, the smaller the parameter $\gamma$, the smoother the logistic function and the larger the influence of the lower and upper regimes.

This process fulfills the stylized facts of non-normality, ARCH effects and further nonlinear dependencies. The latter is related to the fact that the MR-STAR model ”nests several other nonlinear time-series models (...), [f]or example, an artificial neural network”, as van Dijk and

[5] The point at which $F_{12}$ becomes zero depends on the choice of $\gamma$.

Franses (1999, 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 process. We model them by successively drawing waiting times $w_i$ from an exponential distribution with parameter $\lambda_W$, so that the cumulative sum of the waiting times is less than or equal to 510 minutes (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. A jump is then simply defined as multiplying the scale parameter $\sigma_t$ with a fixed factor, in case a jump occurred. In consequence, this model fulfills all the stylized facts outlined in section 2.

#### 3.2.5 Multiple regime smooth transition autoregressive model with nonreversible jumps:

We define nonreversible jumps as extreme events that have a permanent effect on the time series, i.e., a mean shift. We model this series with a compound poisson process defined in Ross (1996), p. 87 ff., as

$$j_t = \sum_{i=1}^{n_t} D_i, \quad \text{with} \quad n_t \sim \text{Pois}(\lambda_P), \quad D_i \sim t\left(0, \omega_P = \sigma_P \sqrt{\frac{\nu_P - 2}{\nu_P}}, \nu_P\right), \tag{4}$$

such that $D_i$ is defined as a sequence of t-distributed random variables independent from $n_t$, i.e., a poisson process with rate $\lambda_P$. The cointegration process $u_t^*$ is then a superposition of the MR(3)-STAR(1)-GARCH(1,1) model of equation (3) and the compound poisson process in equation (4), i.e.,

$$u_t^* = u_t + j_t. \tag{5}$$

In consequence, this model fulfills all the stylized facts outlined in section 2. Note that the compound poisson process has an expected value of zero as well, but its variance increases over time.

#### 3.2.6 Determination of AR(1)-GARCH(1,1) parameters:

The above models require various parameters that we have introduced in equations (1) through (4). We implement the following routine to estimate these parameters based on actual high-frequency data: In section 2, we have obtained a total of 41069 cointegration processes. We use the R package `rugarch` of Ghalanos

(2014a) to fit AR(1)-GARCH(1,1) models to all of these cointegration processes. We use the specification as given in equation 1 with t-distributed innovations. We log all $41069 \times 5$ coefficients in a matrix of these dimensions, notably, the AR(1) coefficient $\phi_1$, the intercept $\alpha_0$ of the GARCH model, the ARCH parameter $\alpha_1$, the GARCH parameter $\beta_1$ and the degrees of freedom of the t-distribution $\nu$. We add an additional column for the unconditional standard deviation $\sigma_G$ of the process, which can be calculated. Next, we sanitize the coefficient matrix by eliminating the 99-th percentile of values as well as a handful of cases 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 40644. In figure 2, we show a histogramm for the above mentioned process parameters and in table 3 their summary statistics. We see that $\alpha_1$ is close to zero (mean of 0.064), $\beta_1$ 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.

For simulation purposes, we perform a core density estimation for all six parameters, with the bandwith chosen according to the approach suggested by Sheather and Jones (1991). We now sample from the 40644 sanitized cointegration processes with replacement, to obtain the five raw parameters. Following Silverman (1986), we add a Gaussian error term with mean zero and standard deviation equal to the bandwidth of the Sheather-Jones method to each parameter. We ensure that the error term does not violate 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 cointegration processes, while introducing additional flexibility for the Monte Carlo Simulations through the Gaussian error terms.

### 3.3 The cointegration relation

Once the prices of the first stock are simulated according to the approach in subsection 3.1 and the cointegration process according to the approach in subsection 3.2, we can finally determine the

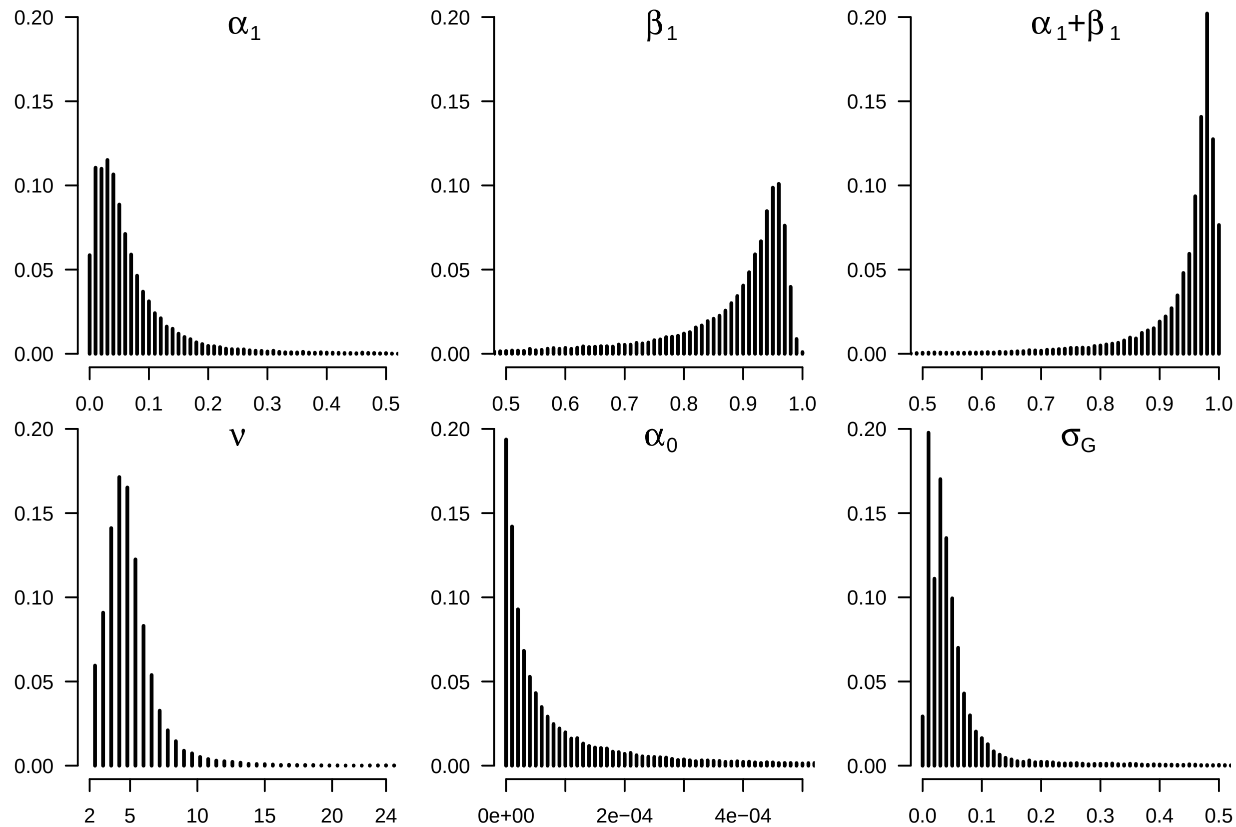


Figure 2: Distribution of parameter estimates.

cointegrating relationship. The latter is defined as

$$p_{2t} = \alpha + \beta p_{1t} + u_t, \tag{6}$$

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 processes as simulated with one of the processes suggested in subsection 3.2. The coefficient $\alpha$ is the intercept, which is drawn from a uniform distribution between $-2$ and 2. The parameter $\beta$ is the coefficient of cointegration, which is 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 Analysis of power and size properties

**3.4.1 Unit root 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 Clegg (2014) to

| | $\phi_1$ | $\alpha_0$ | $\alpha_1$ | $\beta_1$ | $\alpha_1 + \beta_1$ | $\nu$ | $\sigma_G$ |
|---|---|---|---|---|---|---|---|
| Minimum | 0.7519 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 2.1000 | 0.0000 |
| 1st Quartile | 0.9582 | 0.0084 | 0.0224 | 0.8491 | 0.9344 | 3.7480 | 0.0003 |
| Median | 0.9737 | 0.0355 | 0.0449 | 0.9212 | 0.9688 | 4.6280 | 0.0012 |
| Mean | 0.9690 | 0.1620 | 0.0641 | 0.8754 | 0.9396 | 5.0560 | 0.0059 |
| 3rd Quartile | 0.9852 | 0.1215 | 0.0813 | 0.9527 | 0.9828 | 5.6960 | 0.0031 |
| Maximum | 1.0000 | 39.8200 | 0.9290 | 0.9990 | 0.9990 | 99.9940 | 0.2710 |

Table 3: Summary statistics process parameters

our setting. This packages provides p-values for ten different unit root tests that are adjusted for cointegration testing. All tests are listed in table 4 together with their implementation. Details about their functionality can be obtained from the references.

**3.4.2 Definition of size and power:** We aim for analyzing the size and power properties of these tests. In the bivariate setting, all the above tests have a null hypothesis of "no cointegration". In this context, size and power are defined as follows:

*Size:* The size of a test is the probability that the null hypothesis is rejected, when the null hypothesis is true. In other words, the pair is falsely identified as being cointegrated by the test, even though it is not. Hence, we obtain the size of these tests by repeatedly applying them to pairs, that are linked by nonstationary cointegration processes with an AR(1)-coefficient equal to one. The share of pairs with a p-value less than the significance level $\alpha_S = 5\%$ relative to all pairs is the size of the test. If no distortions occur, this value should be equal to the significance level $\alpha_S$, i.e., the type I error of falsely rejecting the null hypothesis.

*Power:* The power of a test is the probability that the null hypothesis is rejected, when the null hypothesis is false. In other words, the pair is correctly identified as being cointegrated by the test. Hence, we obtain the power of these tests by repeatedly applying them to pairs that are linked by stationary cointegration processes. The share of pairs that are correctly identified as cointegrated (i.e., with p-value less than the significance level $\alpha_S = 5\%$) relative to all pairs gives the power of the test.

**3.4.3 Setup of Monte Carlo simulations:** There are two categories of parameters to be set. The first group is determined at random upon every Monte Carlo replication. The underlying logic is described in subsection 3.1 for the parameters affecting $p_{1t}$ and in subsection 3.2 for $\alpha_0$ of the

| **Unit root test** | | **R implementation** | |
|---|---|---|---|
| Augmented Dickey-Fuller test | Dickey and Fuller (1979) and Dickey and Fuller (1981) | `tseries` | Trapletti and Hornik (2013) |
| Phillips-Perron test | Perron (1988) | `tseries` | Trapletti and Hornik (2013) |
| Pantula, Gonzales-Farias and Fuller test | Pantula et al. (1994) | `egcm` | Clegg (2014) |
| Breitung's variance ratio test | Breitung (2002) and Breitung and Taylor (2003) | `egcm` | Clegg (2014) |
| Johansen's eigenvalue test | Johansen (1988), Johansen and Juselius (1990) and Johansen (1991) | `urca`, `vars` | Pfaff (2008) |
| Johansen's trace test | Johansen (1988), Johansen and Juselius (1990) and Johansen (1991) | `urca`, `vars` | Pfaff (2008) |
| Elliott, Rothenberg and Stock point optimal test | Elliott et al. (1996) | `urca` | Pfaff (2008) |
| Elliott, Rothenberg and Stock DF-GLS test | Elliott et al. (1996) | `urca` | Pfaff (2008) |
| Schmidt and Phillips rho statistic | Schmidt and Phillips, C. B. Peter (1992) | `urca` | Pfaff (2008) |
| Based on Hurst exponent | - | `fArma` | Wuertz and Taqqu (2013) |

Table 4: Unit root tests.

GARCH model, the ARCH parameter $\alpha_1$, the GARCH parameter $\beta_1$ the degrees of freedom of the t-distribution $\nu$, the cointegration intercept $\alpha$ and the coefficient of cointegration $\beta$. All other parameters are controlled according to table 5. In particular, we conduct six types of Monte Carlo simulations, each with 10000 replications for each fixed parameter constellation. Each type is based on a different cointegration process, gradually incorporating stylized facts discussed in section 2.

**Type I:** An AR(1) process with normally distributed innovations is used to model the cointegration relationship. We test different values for $\phi_1$. 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 10000 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 four different values for $\phi_1$, 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 process exhibiting non-normality, ARCH effects and nonlinear dependencies. Type IV is split in 36 simulations. We experiment with a total of three different, symmetric threshold levels, i.e., $|c_i| = [1.00, 5.00, 10.00]$, with $i = 1, 2$. For each threshold level, three different values of $\gamma_i$ are tested, i.e., $\gamma_i = [1.00, 5.00, 10.00]$. For each set of threshold and gamma, we jointly vary $\phi_1$, $\phi_2$ and $\phi_3$ 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 10000 replications. Note: The threshold levels $|c_i|$ are a multiple of the unconditional standard deviation of the mixed process, assuming a constant average AR coefficient of $\overline{\phi} = 0.95$.[6] This auxiliary metric is used to define fixed thresholds, even in the light of potential nonstationarities.

**Type V:** A MR(3)-STAR(1)-GARCH(1,1) model with t-distributed innovations is extended by reversible jumps. Type V consists of 12 simulations. We set $|c_i| = 1$ and $\gamma_i = 5$. We iterate through three different values for $\lambda_W$, notably $\lambda_W = [1.00, 5.00, 10.00]$, which corresponds to an expected value of one, five or ten jumps per trading day. Each jump is implemented by multiplying the original scale parameter at the time index of the jump with a factor of 25, which was set ad hoc. For each parameter value of $\lambda_W$, we iterate through four combinations of $\phi_j$, with $j = 1, 2, 3$, as in Type IV. This scenario reflects the stylized facts of non-normality, the ARCH effects, further nonlinearities and reversible jumps.

**Type VI:** A MR(3)-STAR(1)-GARCH(1,1) model with t-distributed innovations is extended by nonreversible jumps. Type VI consists of 36 simulations. As before, we set $|c_i| = 1$ and $\gamma_i = 5$. We iterate through three different values for $\lambda_P$, notably $\lambda_P = [1.00, 5.00, 10.00]$, which corresponds to an expected value of one, five or ten jumps per trading day. For each parameter setting of $\lambda_W$, we test three different values for the standard deviation of the t-distributed innovations for the

[6] Clearly, the mixed AR coefficient varies with the parameters $\phi_1, \phi_2, \phi_3$, but for simplicity reasons and better comparability it was fixed ad hoc at 0.95.

compound poisson process, notably $\sigma_P = [0.01, 0.10, 1.00]$. The degrees of freedom $\nu_P$ are set to five in order to create a leptokurtic distribution with fat tails, thus making larger jumps more probable. For each combination of $\lambda_P$ and $\sigma_P$, we iterate through three combinations of $\phi_j$, as in Type IV. This scenario reflects the stylized facts of non-normality, the ARCH effects, further nonlinearities and nonreversible jumps.

| **MC Type** | **Type I** | | | | **Type II** | | | | **Type III** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Process | AR(1) | | | | AR(1) | | | | AR(1)-GARCH(1,1) | | | |
| Distribution | Normal | | | | t | | | | t | | | |
| $\phi_1$ | 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 | | | | nonreversible | | | |
| Distribution | t | | | | t | | | | t | | | |
| $\phi_L$ | 1.00 | 0.95 | 0.90 | 0.85 | 1.00 | 0.95 | 0.90 | 0.85 | 1.00 | 0.95 | 0.90 | 0.85 |
| $\phi_M$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $\phi_U$ | 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 |
| $\gamma_1$ | 1.00 | 5.00 | 10.00 | | | | | 5.00 | | | | 5.00 |
| $\gamma_2$ | 1.00 | 5.00 | 10.00 | | | | | 5.00 | | | | 5.00 |
| $\lambda_W$ | | | | | 1.00 | 5.00 | 10.00 | | | | | |
| Multiplier | | | | | 25.00 | 25.00 | 25.00 | | | | | |
| $\lambda_P$ | | | | | | | | | 1.00 | 5.00 | 10.00 | |
| $\sigma_P$ | | | | | | | | | 0.01 | 0.10 | 1.00 | |

Table 5: Parameter settings for the Monte Carlo simulations.

# 4. Results

## 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 6 that only very limited size distortions occur for Type

I, i.e., the normally distributed innovations. When running the original routines of Clegg (2014) with 10000 replications, we also obtain very slight size distortion at the level of one percentage point. Whereas Clegg (2014) 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.

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 size distortions beyond those of the AR(1)-GARCH(1,1) process occur, compare table 11 in the appendix. 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.

| Test | $\phi_1$ | 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 |

Table 6: Size Type I, Type II, Type III.

**Power:** The Type I simulation shows undisturbed power properties, see table 7. We see that the power is above 0.80 for all tests for $\phi_1$ equal to 0.85. With the AR(1) coefficient moving towards a random walk, the power deteriorates. For $\phi_1$ 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 (2008). 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 ADF test. Also, we confirm the findings of Pantula et al. (1994) in a cointegration setting, who affirm higher power of their alternative estimators compared to OLS. Along those lines, it makes sense that the Augmented Dickey-Fuller (ADF) test does not perform as good 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 $\phi_1$ 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 Elliott et al. (1996) 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 510 minutes and linear trends are not considered. Hence, the power of these tests vacillates compared to the ADF test. At $\phi_1$ 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 $\phi_1$ equal to 0.95. The Hurst exponent and Breitung's nonparametric variance ratio test (BVR) disappoint.

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.

| Test | $\phi_1$ | 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 |

Table 7: Power Type I, Type II, Type III.

### 4.2 Results Type IV

Type IV introduces threshold nonlinearities as additional stylized fact. Results on power are given in table 8. Overall, we see that 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 reduce the power. Take the PP test as an example. A threshold level of $\pm 1$ leads to a power of 0.78 for $\phi_L = \phi_U = 0.95$ and $\gamma_i = 1.00$. A threshold of $\pm 5$ reduces that power to 0.54 and a threshold of $\pm 10$ to 0.38. This behavior is easy to understand: With increasing threshold levels, a larger share of the total number of observations fall 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. For example, for $\gamma_i = 1.00$, increasing the threshold from $\pm 1$ to $\pm 10$ leads to a deterioration in 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 $\phi_L = \phi_U = 0.85$. Surprisingly, this decline is similar across all tests, so the originally established ranking from table 7 remains intact. Also, we see that threshold effects are getting strong with AR-coefficients approaching nonstationarity. Nonparametric tests, such as BVR, that should perform well in light of nonlinear dynamics, see Breitung (2002), do not gain the upper hand.

The second effect stems from the parameter $\gamma$ of the logistic function in the STAR model. Increasing values of $\gamma$ lead to more abrupt regime shifts and have an adverse effect on the power. The latter can be explained as follows: If $\gamma$ 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 $\gamma$. 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. Speaking in relative terms, the decline in power driven by higher $\gamma$ is also similar across all tests, so the ranking remains unchanged once again.

| Test | $\phi_L = \phi_U$ | 1.00 | 1.00 | 1.00 | 5.00 | 5.00 | 5.00 | 10.00 | 10.00 | 10.00 | $c_i$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1.00 | 5.00 | 10.00 | 1.00 | 5.00 | 10.00 | 1.00 | 5.00 | 10.00 | $\gamma_i$ |
| 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 | |

Table 8: Power Type IV.

### 4.3 Results Type V

Type V introduces reversible jumps as additional stylized fact. Results on power are given in table 9. We see that the power increases with $\lambda_W$, i.e., the expected number of jumps per 510 minutes trading day. 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 0.78 for $\lambda_W = 5$ and 0.80 for $\lambda_W = 10$. 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 process 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, by scaling the shape parameter with a factor of high magnitude as described in section 3, chances are much higher for jumps occurring from the nonstationary to the stationary regime than vice versa. The reason is a much higher variance of the jumps compared to the unconditional variance of the MR(3)-STAR(1)-GARCH(1,1) process. 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 do not change the ranking compared to table 8. In respect to the AR-coefficients, we observe that gain in power is more expressed when moving closer to nonstationarity.

| Test | $\phi_L = \phi_U$ | $\lambda_W = 0$ | $\lambda_W = 1$ | $\lambda_W = 5$ | $\lambda_W = 10$ |
|---|---|---|---|---|---|
| pp | 0.95 | 0.65 | 0.69 | 0.78 | 0.80 |
| pp | 0.90 | 0.97 | 0.97 | 0.98 | 0.99 |
| pp | 0.85 | 0.98 | 0.98 | 0.99 | 0.99 |
| adf | 0.95 | 0.46 | 0.49 | 0.54 | 0.57 |
| adf | 0.90 | 0.87 | 0.90 | 0.93 | 0.95 |
| adf | 0.85 | 0.95 | 0.96 | 0.97 | 0.98 |
| jo-e | 0.95 | 0.38 | 0.40 | 0.45 | 0.48 |
| jo-e | 0.90 | 0.89 | 0.92 | 0.96 | 0.97 |
| jo-e | 0.85 | 0.97 | 0.97 | 0.98 | 0.99 |
| jo-t | 0.95 | 0.36 | 0.39 | 0.42 | 0.45 |
| jo-t | 0.90 | 0.85 | 0.89 | 0.95 | 0.96 |
| jo-t | 0.85 | 0.96 | 0.97 | 0.98 | 0.99 |
| ers-p | 0.95 | 0.58 | 0.64 | 0.71 | 0.73 |
| ers-p | 0.90 | 0.82 | 0.86 | 0.89 | 0.90 |
| ers-p | 0.85 | 0.91 | 0.92 | 0.93 | 0.93 |
| ers-d | 0.95 | 0.57 | 0.62 | 0.70 | 0.71 |
| ers-d | 0.90 | 0.79 | 0.83 | 0.88 | 0.88 |
| ers-d | 0.85 | 0.86 | 0.89 | 0.91 | 0.92 |
| sp-r | 0.95 | 0.47 | 0.50 | 0.54 | 0.57 |
| sp-r | 0.90 | 0.85 | 0.87 | 0.90 | 0.92 |
| sp-r | 0.85 | 0.95 | 0.95 | 0.95 | 0.96 |
| hurst | 0.95 | 0.34 | 0.34 | 0.34 | 0.37 |
| hurst | 0.90 | 0.61 | 0.60 | 0.61 | 0.63 |
| hurst | 0.85 | 0.74 | 0.75 | 0.76 | 0.78 |
| bvr | 0.95 | 0.43 | 0.44 | 0.45 | 0.48 |
| bvr | 0.90 | 0.70 | 0.70 | 0.72 | 0.74 |
| bvr | 0.85 | 0.81 | 0.83 | 0.85 | 0.87 |
| pgff | 0.95 | 0.68 | 0.72 | 0.82 | 0.84 |
| pgff | 0.90 | 0.97 | 0.97 | 0.98 | 0.99 |
| pgff | 0.85 | 0.98 | 0.98 | 0.99 | 0.99 |

Table 9: Power Type V.

### 4.4 Results Type VI

Result for Type VI are given in table 10. Type VI introduces nonreversible jumps through a compound poisson process with expected number of jumps $\lambda_P$ per trading day at standard deviation $\sigma_P$ and degrees of freedom $\nu_P$. 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 10, we see that the power decreases with increasing expected number of jumps and increasing standard deviation of the jumps. In the following, we differentiate the results by jump size: *Small jumps:* It is interesting to see that smaller jumps with standard deviations of 0.01 only have a minor effect on the power - even though this value is more than eight times higher than the median unconditional variance $\sigma_G$ of the innovations, see table 3. For example, the power of the PP test declines from 0.65 to 0.62 for $\lambda_P = 5$ and to 0.49 for $\lambda_P = 10$. *Medium jumps:* Medium-sized jumps with standard deviation of 0.10 are more detrimental. An expected value of five jumps leads to a deterioration in power to 0.41 for the PP test, i.e., a slump of 37 percent. Nevertheless, power is still a respectable level, considering that these jumps exhibit a more than 80 times higher standard deviation than the underlying process. However, higher jump frequencies completely deteriorate the power and render the tests useless - 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 1.00 already do significant damage if they occur just once in expectation. We see that power deteriorates by more than 40 percent across almost all tests compared to the case with no jumps. The latter finding suggests that if a jump actually manifests, it reduces the power to approximately five percent, i.e., the size of the type one error. If the jump does not manifest, it remains at the original level. The mixed effect results in the decline of 40 percent. The salient point is, that the occurrence of one large jump prevents detection of time-varying cointegration relationships.

These results are paramount, meaning that we are able to detect a cointegration relationship in financial market data, even if its mean varies over time either at low frequencies or with only limited shifts in mean - see section 5 for a more comprehensive discussion.

| | | 0.00 | 1.00 | 1.00 | 1.00 | 5.00 | 5.00 | 5.00 | 10.00 | 10.00 | 10.00 | $\lambda_P$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Test | $\phi_L = \phi_U$ | 0.00 | 0.01 | 0.10 | 1.00 | 0.01 | 0.1 | 1.00 | 0.01 | 0.10 | 1.00 | $\sigma_P$ |
| pp | 0.95 | 0.65 | 0.64 | 0.56 | 0.37 | 0.62 | 0.41 | 0.10 | 0.49 | 0.12 | 0.06 | |
| pp | 0.90 | 0.97 | 0.96 | 0.88 | 0.55 | 0.95 | 0.68 | 0.13 | 0.84 | 0.17 | 0.06 | |
| pp | 0.85 | 0.98 | 0.98 | 0.91 | 0.57 | 0.97 | 0.73 | 0.14 | 0.87 | 0.19 | 0.06 | |
| adf | 0.95 | 0.46 | 0.46 | 0.41 | 0.27 | 0.45 | 0.28 | 0.10 | 0.35 | 0.10 | 0.05 | |
| adf | 0.90 | 0.87 | 0.86 | 0.75 | 0.47 | 0.84 | 0.52 | 0.12 | 0.66 | 0.12 | 0.05 | |
| adf | 0.85 | 0.95 | 0.95 | 0.83 | 0.52 | 0.93 | 0.59 | 0.11 | 0.75 | 0.13 | 0.06 | |
| jo-e | 0.95 | 0.38 | 0.38 | 0.31 | 0.23 | 0.35 | 0.21 | 0.10 | 0.26 | 0.08 | 0.06 | |
| jo-e | 0.90 | 0.89 | 0.88 | 0.75 | 0.47 | 0.87 | 0.48 | 0.12 | 0.65 | 0.10 | 0.07 | |
| jo-e | 0.85 | 0.97 | 0.96 | 0.85 | 0.52 | 0.95 | 0.60 | 0.12 | 0.78 | 0.12 | 0.06 | |
| jo-t | 0.95 | 0.36 | 0.36 | 0.31 | 0.21 | 0.34 | 0.21 | 0.08 | 0.26 | 0.07 | 0.06 | |
| jo-t | 0.90 | 0.85 | 0.85 | 0.72 | 0.45 | 0.82 | 0.46 | 0.11 | 0.60 | 0.10 | 0.06 | |
| jo-t | 0.85 | 0.96 | 0.96 | 0.84 | 0.51 | 0.94 | 0.57 | 0.11 | 0.75 | 0.11 | 0.06 | |
| ers-p | 0.95 | 0.58 | 0.58 | 0.51 | 0.31 | 0.56 | 0.35 | 0.08 | 0.45 | 0.10 | 0.05 | |
| ers-p | 0.90 | 0.82 | 0.82 | 0.72 | 0.44 | 0.80 | 0.50 | 0.09 | 0.64 | 0.14 | 0.05 | |
| ers-p | 0.85 | 0.91 | 0.90 | 0.78 | 0.49 | 0.87 | 0.55 | 0.10 | 0.71 | 0.14 | 0.05 | |
| ers-d | 0.95 | 0.57 | 0.55 | 0.48 | 0.30 | 0.54 | 0.33 | 0.08 | 0.43 | 0.10 | 0.05 | |
| ers-d | 0.90 | 0.79 | 0.79 | 0.68 | 0.41 | 0.76 | 0.47 | 0.09 | 0.59 | 0.12 | 0.04 | |
| ers-d | 0.85 | 0.86 | 0.85 | 0.73 | 0.44 | 0.83 | 0.50 | 0.09 | 0.65 | 0.12 | 0.05 | |
| sp-r | 0.95 | 0.47 | 0.47 | 0.40 | 0.25 | 0.45 | 0.27 | 0.07 | 0.36 | 0.10 | 0.05 | |
| sp-r | 0.90 | 0.85 | 0.84 | 0.72 | 0.43 | 0.82 | 0.47 | 0.08 | 0.62 | 0.10 | 0.05 | |
| sp-r | 0.85 | 0.95 | 0.94 | 0.81 | 0.47 | 0.92 | 0.53 | 0.08 | 0.72 | 0.10 | 0.05 | |
| hurst | 0.95 | 0.34 | 0.34 | 0.29 | 0.23 | 0.33 | 0.19 | 0.09 | 0.24 | 0.07 | 0.05 | |
| hurst | 0.90 | 0.61 | 0.58 | 0.46 | 0.34 | 0.54 | 0.26 | 0.09 | 0.34 | 0.07 | 0.05 | |
| hurst | 0.85 | 0.74 | 0.71 | 0.54 | 0.39 | 0.63 | 0.27 | 0.09 | 0.36 | 0.07 | 0.05 | |
| bvr | 0.95 | 0.43 | 0.43 | 0.35 | 0.24 | 0.40 | 0.23 | 0.07 | 0.29 | 0.07 | 0.04 | |
| bvr | 0.90 | 0.70 | 0.66 | 0.52 | 0.34 | 0.63 | 0.29 | 0.07 | 0.40 | 0.08 | 0.05 | |
| bvr | 0.85 | 0.81 | 0.79 | 0.60 | 0.40 | 0.70 | 0.29 | 0.07 | 0.42 | 0.07 | 0.05 | |
| pgff | 0.95 | 0.68 | 0.67 | 0.59 | 0.37 | 0.66 | 0.42 | 0.09 | 0.53 | 0.13 | 0.05 | |
| pgff | 0.90 | 0.97 | 0.97 | 0.89 | 0.55 | 0.96 | 0.71 | 0.13 | 0.85 | 0.19 | 0.06 | |
| pgff | 0.85 | 0.98 | 0.98 | 0.92 | 0.58 | 0.97 | 0.77 | 0.14 | 0.90 | 0.22 | 0.06 | |

Table 10: Power Type VI.

# 5. Economic interpretation

## 5.1 Interpretation of the cointegration processes

From an economic perspective, a cointegration relation reflects the law of one price (LOP); see Gatev et al. (2006) for the subsequent discussion. Ingersoll (1987) states the LOP as the "proposition (...) that two investments with the same payoff in every state of nature must have the same current value." Chen and Knez (1995) expand on this concept and argue that "*closely integrated* markets should assign to similar payoffs prices that are *close*". In other words, the models provided above tie the price time 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 Gatev et al. (2006), ensuring subsequent convergence. The latter is referred to as a "*weak-form market integration*" by Chen and Knez (1995), meaning a "*near-efficient market*". Based on this interpretation of cointegration, we may interpret the models examined as follows:

- The AR(1) and the AR(1)-GARCH(1,1) models are very strict in their formulation, allowing only for one stationary (or nonstationary regime) with arbitrage permanently occurring.

- The MR(3)-STAR(1)-GARCH(1,1) model is a lot more flexible. Specifically, in the middle regime, the cointegration process tying the prices together truly behaves like a random work - arbitrage is not yet profitable. However, once the cointegration process ventures in the outer regimes, arbitrage starts to occur. The (symmetric) thresholds $c_1$ and $c_2$ define at which level arbitrage kicks in and the parameter $\gamma$ 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 Krauss (2015). This diversity creates fuzziness around the threshold levels, which is reflected in the parameter $\gamma$. A value of ten already hinges towards a heaviside function, a value of one is very smooth. As ad hoc choice for Type V and Type VI, we have opted for a value of five. Regarding the thresholds, we have reason to believe that they would amount to the mean value of the cointegration process plus or minus the transaction costs of the arbitrageur. The average bid-ask spread for all DAX 30 stocks in 2014 is in the range of four to five basis points (bps), i.e., approximately 10 bps per round trip of the spread trade. Compared to this estimate, even a threshold level of one time the unconditional standard deviation of a hypothetical mixed process with $\phi = 0.95$ offers sufficient profit opportunities. The unconditional standard deviation $\sigma_U$ for an AR(1) process is defined in e.g., Tsay (2010) as

$$\sigma_U = \frac{\sigma_G}{\sqrt{1-\phi^2}}. \tag{7}$$

As before, $\sigma_G$ denotes the unconditional standard deviation of the GARCH innovations. For a mean value of 0.006 for $\sigma_G$, and $\phi$ equal to 0.95, the unconditional standard deviation $\sigma_U$ amounts to 0.019. Hence, clearly, the return from a spread trade, entered at the threshold level of $|c_i| = 1 \cdot \sigma_U$ and held until mean reversion amounts to 1.9 percent and exceeds transaction costs by far[7]. Considering this thought experiment, it is reasonable to believe that equity markets exhibit arbitrage threshold levels below $|c_i| = 1$. In this light it appears to be a conservative choice to set $|c_i| = 1$ as done for Type V and Type VI simulations.

- Reversible jumps can be interpreted as uninformed buying or selling, as described in Andrade et al. (2005). In this case, clearly, the arbitrage mechanism re-establishes equilibrium over time. The frequency at which uninformed trading occurs in near-efficient markets is not known, so there is no good assessment at hand for reasonable values of $\lambda_W$.

- Nonreversible jumps may reflect idiosyncratic information, affecting only one of the two companies. Such a jump translates into a regime shift in the mean equation, causing further arbitrage to occur at a different level. It is clear that strong and frequent jumps of this type will lead to a process that frequently re-establishes new cointegrating relationships at different levels and of short durations. We believe that during one trading day of 510 minutes, such a frequent and especially strong arrival of idiosyncratic information in an otherwise cointegrated stock pair is rather the exception than the rule - otherwise the LOP simply does not hold for these two securities, they are not cointegrated and arbitrage does not exist. However, clearly, this conjecture is subject for further research.

[7] We have calculated the return on an investment of 1 monetary unit in the long leg of the spread trade. The latter is often the convention, see for example Gatev et al. (2006).

## 5.2 Interpretation of the power analysis

An economic interpretation of the analysis in section 4 is as simple as important: If the LOP applies and relative-value arbitrage is performed between two securities, the ability of a cointegration test to detect the cointegration relationship mainly depends on four factors: First, the parametrization of the MR(3)-STAR(1)-GARCH(1,1) model, i.e., the threshold level at which arbitrage kicks in and the abruptness with which arbitrage kicks in. Second, the expected number and scaling factor of reversible jumps and third, expected number and standard deviation of irreversible jumps. In the light of the economic background outlined in subsection 5.1, we justify choices of models and parameters as follows:

- **Type IV:** Under reasonable arbitrage assumptions in near-efficient markets outlined in subsection 5.1, we would assume that arbitrage kicks in at thresholds below $c_i = 1$ with an abruptness of $\gamma_i \leq 5$. Under these conditions, standard cointegration tests exhibit good power properties, e.g., the PP test with a power of 0.65 or the PGFF with a power of 0.68 - even for an AR(1) coefficient of 0.95. For lower values of $\phi_1$, power for these tests even increases to levels above 0.95. However, once we make more aggressive assumptions about the threshold levels or the abruptness, power quickly deteriorates. It is the basis of future research to evaluate the threshold levels and abruptness coefficients associated with relative-value arbitrageurs.

- **Type V:** Reversible jumps representing uninformed demand shocks actually have a positive effect on the power. Quintessentially, they make the arbitrage process visible, by pushing the cointegration process in its stationary region. Any occurrence is beneficial to the power of the tests. As such, lack of clarity on the parameter $\lambda_W$ is not key.

- **Type VI:** Nonreversible jumps contradict the stationarity assumption of the cointegration process and may thus prevent the identification of such a relation. Therefore, we can neither provide an estimate for the frequency of nonreversible jumps nor for their variance. These two parameters would steer the level of idiosyncratic information, affecting only one of the two stocks of a cointegrated pair vs. the level of common information, affecting both stocks of the pair. However, we can put upper bounds on the parameter values, i.e., until which levels it is still possible to detect a cointegrating relationship with acceptable power. Again, considering the PP test for $\phi_L = \phi_U = 0.95$, we see that at $\sigma_P = 0.01$ and $\lambda_P = 10$, we would still wind up with a power of 0.49. Consequently, some idiosyncratic information may occur, as long as the level shifts in the mean equation are small in absolute value.

## 6. Conclusion

We conduct an in-depth analysis of 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 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 processes are classified as stationary by the Johansen trace test, but still exhibit non-normalities, ARCH effects, further nonlinearities as well as jumps. One could have assumed that these stylized facts cancel each other out, just as the nonstationary components. The data prove us wrong in this case.

Second, we propose an innovative approach for simulating stock prices following the stationary bootstrap of Politis and Romano (1994). 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 cointegration processes based on the current literature. These processes accommodate the above mentioned stylized facts in a staggered approach. To our knowledge, the application of a MR(3)-STAR(1)-GARCH(1,1) model is a novelty in this context. It provides a lot more 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 marginal impact on size and power properties. We can summarize 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 equal to round-trip transaction costs, 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, actually increase the power. This improvement 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, nonreversible jumps disrupt the power with increasing rate of occurrence and increasing variance.

Across all of the above mentioned stylized facts, the PGFF and the PP test exhibit the most favorable power properties. These test strictly dominate all other tests in all settings reflecting stylized facts, where PGFF is slightly better than PP.

We conclude that contemporary cointegration tests may be applied in the high-frequency setting of one minute stock return data.

## References


Alexander, C. (2008). *Practical financial econometrics*. Wiley, Chichester.

Andrade, S., Di Pietro, V., and Seasholes, M. (2005). Understanding the profitability of pairs trading. *Working paper, UC Berkeley, Northwestern University.*

Azzalini, A. and Capitanio, A. (2014). *The skew-normal and related families*. Cambridge University Press, Cambridge.

Balke, N. S. and Fomby, T. B. (1997). Threshold cointegration. *International Economic Review*, 38(3):627–645.

Barndorff-Nielsen, O. E. and Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4(1):1–30.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Boswijk, H. P., Lucas, A., and Taylor, N. (1999). A comparison of parametric, semi-nonparametric, adaptive, and nonparametric cointegration tests. *Tinbergen Institute Discussion Papers.*

Boudt, K., Cornelissen, J., Payseur, S., Nguyen, G., and Schermer, M. (2014). highfrequency: Tools for highfrequency data analysis. *R package.*

Breitung, J. (2002). Nonparametric tests for unit roots and cointegration. *Journal of Econometrics*, 108(2):343–363.

Breitung, J. and Taylor, A. (2003). Corrigendum to "Nonparametric tests for unit roots and cointegration" [J. Econom. 108 (2002) 343–363]. *Journal of Econometrics*, 117(2):401–404.

Cavaliere, G., Rahbek, A., and Am Taylor, R. (2010a). Testing for co-integration in vector autoregressions with non-stationary volatility. *Journal of Econometrics*, 158(1):7–24.

Cavaliere, G., Rahbek, A., and Taylor, A. R. (2010b). Cointegration rank testing under conditional heteroskedasticity. *Econometric Theory*, 26(06):1719–1760.

Chan, K.-S. and Ripley, B. (2012). TSA: Time series analysis. *R package.*

Chen, Z. and Knez, P. J. (1995). Measurement of market integration and arbitrage. *Review of Financial Studies*, 8(2):287–325.

Clegg, M. (2014). On the persistence of cointegration in pairs trading. *SSRN Electronic Journal.*

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431.

Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49(4):1057–1072.

Doornik, J. A. (1998). Approximations to the asymptotic distributions of cointegration tests. *Journal of Economic Surveys*, 12(5):573–593.

Dunis, C. L., Giorgioni, G., Laws, J., and Rudy, J. (2010). Statistical arbitrage and high-frequency data with an application to Eurostoxx 50 equities. *Working paper, Liverpool Business School.*

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64(4):813.

Elyasiani, E. and Kocagil, A. E. (2001). Interdependence and dynamics in currency futures markets: A multivariate analysis of intraday data. *Journal of Banking & Finance*, 25(6):1161–1186.

Engle, R. F. and Sokalska, M. E. (2012). Forecasting intraday volatility in the US equity market: Multiplicative component GARCH. *Journal of Financial Econometrics*, 10(1):54–83.

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

Ghalanos, A. (2014a). rugarch: Univariate GARCH models. *R package.*

Ghalanos, A. (2014b). twinkle: Dynamic smooth transition ARMAX models. *R package.*

Graves, S. (2014). FinTS: Companion to Tsay (2005) Analysis of financial time series. *R package.*

Hansen, B. E. and Seo, B. (2002). Testing for two-regime threshold cointegration in vector error-correction models. *Journal of Econometrics*, 110(2):293–318.

Hasbrouck, J. (2003). Intraday price formation in U.S. equity index markets. *The Journal of Finance*, 58(6):2375–2399.

Haug, A. A. (1996). Tests for cointegration a Monte Carlo comparison. *Journal of Econometrics*, 71(1-2):89–115.

Herrmann, K., Teis, S., and Yu, W. (2014). Components of intraday volatility and their prediction at different sampling frequencies with application to DAX and BUND futures. *IWQW Discussion Paper Series, Friedrich-Alexander-University Erlangen-Nürnberg, Nürnberg.*

Hubrich, K., Lutkepohl, H., and Saikkonen, P. (2001). A review of systems cointegration tests. *Econometric Reviews*, 20(3):247–318.

Ingersoll, J. E. (1987). *Theory of financial decision making.* Rowman & Littlefield, Lanham, Md.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3):231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580.

Johansen, S. and Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration–With applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, 52(2):169–210.

Krauss, C. (2015). Statistical Arbitrage Pairs Trading Strategies: Review and Outlook. *IWQW Discussion Paper Series, Friedrich-Alexander-University Erlangen-Nürnberg, Nürnberg.*

Kremers, J. J., Ericsson, N. R., and Dolado, J. J. (1992). The power of cointegration tests. *Oxford Bulletin of Economics and Statistics*, 54(3):325–348.

Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, 75(3):491–499.

Pantula, S. G., Gonzalez-Farias, G., and Fuller, W. A. (1994). A comparison of unit-root test criteria. *Journal of Business & Economic Statistics*, 12(4):449–459.

Pati, P. C. and Rajib, P. (2011). Intraday return dynamics and volatility spillovers between NSE S&P CNX Nifty stock index and stock index futures. *Applied Economics Letters*, 18(6):567–574.

Perron, P. (1988). Trends and random walks in macroeconomic time series. *Journal of Economic Dynamics and Control*, 12(2-3):297–332.

Pfaff, B. (2008). *Analysis of integrated and cointegrated time series with R.* Springer, New York.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.

R Core Team (2014). R: A language and environment for statistical computing. *R software.*

Rahbek, A., Hansen, E., and Dennis, J. G. (2002). ARCH innovations and their impact on cointegration rank testing. *Working paper, Centre for Analytical Finance*, 22:15.

Ross, S. M. (1996). *Stochastic processes.* Wiley, New York.

Schmidt, P. and Phillips, C. B. Peter (1992). LM tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics*, 54(3):257–287.

Seo, M. (2006). Bootstrap testing for the null of no cointegration in a threshold vector error correction model. *Journal of Econometrics*, 134(1):129–150.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 683–690.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* Chapman and Hall, London and New York.

Terasvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89(425):208–218.

Trapletti, A. and Hornik, K. (2013). tseries: Time series analysis and computational finance. *R package*.

Tsay, R. S. (2010). *Analysis of financial time series*. John Wiley & Sons, Hoboken, N.J.

van Dijk, D. and Franses, P. H. (1999). Modeling multiple regimes in the business cycle. *Macroeconomic dynamics*, 3(03):311–340.

Wuertz, D. and Taqqu, M. S. (2013). fArma: ARMA time series modelling. *R package*.

Yang, F. (2013). CommonTrend: Extract and plot common trends from a cointegration system. Calculate p-value for Johansen statistics. *R package*.

Yang, J., Yang, Z., and Zhou, Y. (2012). Intraday price discovery and volatility transmission in stock index and stock index futures markets: Evidence from China. *Journal of Futures Markets*, 32(2):99–121.

| Test | $\phi_L = \phi_U$ | 1 | 1 | 1 | 5 | 5 | 5 | 10 | 10 | 10 | $c_i$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 5 | 10 | 1 | 5 | 10 | 1 | 5 | 10 | $\gamma_i$ |
| pp | 1 | 0.07 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.07 | 0.06 | 0.07 | |
| adf | 1 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | |
| jo-e | 1 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | |
| jo-t | 1 | 0.07 | 0.06 | 0.06 | 0.07 | 0.06 | 0.06 | 0.06 | 0.07 | 0.06 | |
| ers-p | 1 | 0.05 | 0.05 | 0.05 | 0.05 | 0.06 | 0.05 | 0.05 | 0.05 | 0.06 | |
| ers-d | 1 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.06 | 0.05 | |
| sp-r | 1 | 0.06 | 0.05 | 0.05 | 0.05 | 0.06 | 0.05 | 0.05 | 0.05 | 0.05 | |
| hurst | 1 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.06 | 0.05 | 0.05 | |
| bvr | 1 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | |
| pgff | 1 | 0.06 | 0.06 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.06 | 0.07 | |

| Test | $\phi_L = \phi_U$ | $\lambda_W = 1$ | $\lambda_W = 5$ | $\lambda_W = 10$ |
|---|---|---|---|---|
| pp | 1 | 0.07 | 0.06 | 0.06 |
| adf | 1 | 0.06 | 0.06 | 0.06 |
| jo-e | 1 | 0.07 | 0.07 | 0.07 |
| jo-t | 1 | 0.06 | 0.07 | 0.07 |
| ers-p | 1 | 0.05 | 0.05 | 0.05 |
| ers-d | 1 | 0.05 | 0.05 | 0.05 |
| sp-r | 1 | 0.05 | 0.04 | 0.04 |
| hurst | 1 | 0.06 | 0.05 | 0.06 |
| bvr | 1 | 0.05 | 0.04 | 0.04 |
| pgff | 1 | 0.06 | 0.05 | 0.06 |

| Test | $\phi_L = \phi_U$ | 1 | 1 | 1 | 5 | 5 | 5 | 10 | 10 | 10 | $\lambda_P$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.01 | 0.1 | 1 | 0.01 | 0.1 | 1 | 0.01 | 0.1 | 1 | $\sigma_P$ |
| pp | 1 | 0.07 | 0.06 | 0.06 | 0.07 | 0.06 | 0.07 | 0.06 | 0.06 | 0.06 | |
| adf | 1 | 0.06 | 0.06 | 0.07 | 0.06 | 0.06 | 0.07 | 0.06 | 0.06 | 0.05 | |
| jo-e | 1 | 0.07 | 0.07 | 0.08 | 0.06 | 0.07 | 0.08 | 0.07 | 0.07 | 0.06 | |
| jo-t | 1 | 0.07 | 0.07 | 0.07 | 0.06 | 0.07 | 0.08 | 0.06 | 0.06 | 0.06 | |
| ers-p | 1 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.04 | 0.05 | 0.05 | 0.05 | |
| ers-d | 1 | 0.06 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | |
| sp-r | 1 | 0.05 | 0.05 | 0.05 | 0.06 | 0.05 | 0.04 | 0.05 | 0.05 | 0.04 | |
| hurst | 1 | 0.05 | 0.05 | 0.06 | 0.05 | 0.06 | 0.06 | 0.05 | 0.05 | 0.05 | |
| bvr | 1 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.04 | |
| pgff | 1 | 0.07 | 0.06 | 0.05 | 0.06 | 0.06 | 0.05 | 0.06 | 0.05 | 0.05 | |

Table 11: Size Type IV, V and VI.

## Diskussionspapiere 2015
## Discussion Papers 2015

01/2015 **Seebauer, Michael**: Does Direct Democracy Foster Efficient Policies? An Experimental Investigation of Costly Initiatives

02/2015 **Bünnings, Christian, Schmitz, Hendrik, Tauchmann, Harald and Ziebarth, Nicolas R.**: How Health Plan Enrollees Value Prices Relative to Supplemental Benefits and Service Quality

03/2015 **Schnabel, Claus:** United, yet apart? A note on persistent labour market differences between western and eastern Germany

04/2015 **Dieckmann, Anja, Fischbacher, Urs, Grimm, Veronika, Unfried, Matthias, Utikal, Verena and Valmasoni, Lorenzo:** Trust and Beliefs among Europeans: Cross-Country Evidence on Perceptions and Behavior

05/2015 **Grimm, Veronika, Utikal, Verena and Valmasoni, Lorenzo**: In-group favoritism and discrimination among multiple out-groups

06/2015 **Reichert, Arndt R., Tauchmann, Harald and Wübker, Ansgar**: Weight Loss and Sexual Activity in Adult Obese Individuals: Establishing a Causal Link

07/2015 **Klein, Ingo, Mangold, Benedikt:** Cumulative Paired $\phi$-Entropy

08/2015 **Erbe, Katharina:** Tax Planning of Married Couples in East and West Germany

09/2015 **Krauss, Christopher:** Statistical Arbitrage Pairs Trading Strategies: Review and Outlook

10/2015 **Mangold, Benedikt:** A multivariate linear rank test of independence based on a multiparametric copula with cubic sections

## Diskussionspapiere 2014
## Discussion Papers 2014

01/2014 **Rybizki, Lydia**: Learning cost sensitive binary classification rules accounting for uncertain and unequal misclassification costs

02/2014 **Abbiati, Lorenzo, Antinyan, Armenak and Corazzini, Lucca:** Are Taxes Beautiful? A Survey Experiment on Information, Tax Choice and Perceived Adequacy of the Tax Burden

03/2014 **Feicht, Robert, Grimm, Veronika and Seebauer, Michael**: An Experimental Study of Corporate Social Responsibility through Charitable Giving in Bertrand Markets

04/2014 **Grimm, Veronika, Martin, Alexander, Weibelzahl, Martin and Zoettl, Gregor**: Transmission and Generation Investment in Electricity Markets: The Effects of Market Splitting and Network Fee Regimes

05/2014 **Cygan-Rehm, Kamila and Riphahn, Regina**: Teenage Pregnancies and Births in Germany: Patterns and Developments

06/2014 **Martin, Alexander and Weibelzahl, Martin:** Where and when to Pray? - Optimal Mass Planning and Efficient Resource Allocation in the Church

07/2014 **Abraham, Martin, Lorek, Kerstin, Richter, Friedemann and Wrede, Matthias**: Strictness of Tax Compliance Norms: A Factorial Survey on the Acceptance of Inheritance Tax Evasion in Germany

08/2014 **Hirsch, Boris, Oberfichtner, Michael and Schnabel Claus**: The levelling effect of product market competition on gender wage discrimination

09/1014 **Mangold, Benedikt:** Plausible Prior Estimation

10/2014 **Gehrke, Britta:** Fiscal Rules and Unemployment

11/2014 **Gehrke, Britta** and **Yao, Fang**: Phillips Curve Shocks and Real Exchange Rate Fluctuations: SVAR Evidence

12/2014 **Mäder, Miriam, Müller, Steffen, Riphahn, Regina and Schwientek, Caroline**: Intergenerational transmission of unemployment - evidence for German sons

13/2014 **Casal, Sandro, Ploner, Matteo and Sproten, Alec N.**: Fostering the Best Execution Regime. An Experiment about Pecuniary Sanctions and Accountability in Fiduciary Money Management

14/2014 **Lochner, Benjamin**: Employment protection in dual labor markets - Any amplification of macroeconomic shocks?

15/2014 **Herrmann, Klaus, Teis, Stefan and Yu, Weijun**: Components of Intraday Volatility and Their Prediction at Different Sampling Frequencies with Application to DAX and BUND Futures

## Diskussionspapiere 2013
## Discussion Papers 2013

01/2013 **Wrede, Matthias**: Rational choice of itemized deductions

02/2013 **Wrede, Matthias**: Fair Inheritance Taxation in the Presence of Tax Planning

03/2013 **Tinkl, Fabian**: Quasi-maximum likelihood estimation in generalized polynomial autoregressive conditional heteroscedasticity models

04/2013 **Cygan-Rehm, Kamila**: Do Immigrants Follow Their Home Country's Fertility Norms?

05/2013 **Ardelean, Vlad and Pleier, Thomas**: Outliers & Predicting Time Series: A comparative study

06/2013 **Fackler, Daniel and Schnabel, Claus**: Survival of spinoffs and other startups: First evidence for the private sector in Germany, 1976-2008

07/2013 **Schild, Christopher-Johannes:** Do Female Mayors Make a Difference? Evidence from Bavaria

08/2013 **Brenzel, Hanna, Gartner, Hermann and Schnabel Claus**: Wage posting or wage bargaining? Evidence from the employers' side

09/2013 **Lechmann, Daniel S. and Schnabel Claus**: Absence from work of the self-employed: A comparison with paid employees

10/2013 **Bünnings, Ch. and Tauchmann, H.**: Who Opts Out of the Statutory Health Insurance? A Discrete Time Hazard Model for Germany

## Diskussionspapiere 2012
## Discussion Papers 2012

01/2012 **Wrede, Matthias**: Wages, Rents, Unemployment, and the Quality of Life

02/2012 **Schild, Christopher-Johannes:** Trust and Innovation Activity in European Regions - A Geographic Instrumental Variables Approach

03/2012 **Fischer, Matthias:** A skew and leptokurtic distribution with polynomial tails and characterizing functions in closed form

04/2012 **Wrede, Matthias**: Heterogeneous Skills and Homogeneous Land: Segmentation and Agglomeration

05/2012 **Ardelean, Vlad**: Detecting Outliers in Time Series

# Diskussionspapiere 2011
# Discussion Papers 2011

01/2011 **Klein, Ingo, Fischer, Matthias and Pleier, Thomas:** Weighted Power Mean Copulas: Theory and Application

02/2011 **Kiss, David:** The Impact of Peer Ability and Heterogeneity on Student Achievement: Evidence from a Natural Experiment

03/2011 **Zibrowius, Michael:** Convergence or divergence? Immigrant wage assimilation patterns in Germany

04/2011 **Klein, Ingo and Christa, Florian:** Families of Copulas closed under the Construction of Generalized Linear Means

05/2011 **Schnitzlein, Daniel**: How important is the family? Evidence from sibling correlations in permanent earnings in the US, Germany and Denmark

06/2011 **Schnitzlein, Daniel**: How important is cultural background for the level of intergenerational mobility?

07/2011 **Steffen Mueller**: Teacher Experience and the Class Size Effect - Experimental Evidence

08/2011 **Klein, Ingo:** Van Zwet Ordering for Fechner Asymmetry

09/2011 **Tinkl, Fabian and Reichert Katja:** Dynamic copula-based Markov chains at work: Theory, testing and performance in modeling daily stock returns

10/2011 **Hirsch, Boris and Schnabel, Claus:** Let's Take Bargaining Models Seriously: The Decline in Union Power in Germany, 1992 – 2009

11/2011 **Lechmann, Daniel S.J. and Schnabel**, **Claus** : Are the self-employed really jacks-of-all-trades? Testing the assumptions and implications of Lazear's theory of entrepreneurship with German data

12/2011 **Wrede, Matthias**: Unemployment, Commuting, and Search Intensity

13/2011 **Klein, Ingo**: Van Zwet Ordering and the Ferreira-Steel Family of Skewed Distributions

## Diskussionspapiere 2010
## Discussion Papers 2010

01/2010 **Mosthaf, Alexander, Schnabel, Claus and Stephani, Jens:** Low-wage careers: Are there dead-end firms and dead-end jobs?

02/2010 **Schlüter, Stephan and Matt Davison:** Pricing an European Gas Storage Facility using a Continuous-Time Spot Price Model with GARCH Diffusion

03/2010 **Fischer, Matthias, Gao, Yang and Herrmann, Klaus:** Volatility Models with Innovations from New Maximum Entropy Densities at Work

04/2010 **Schlüter, Stephan and Deuschle, Carola:** Using Wavelets for Time Series Forecasting – Does it Pay Off?

05/2010 **Feicht, Robert and Stummer, Wolfgang:** Complete closed-form solution to a stochastic growth model and corresponding speed of economic recovery.

06/2010 **Hirsch, Boris and Schnabel, Claus:** Women Move Differently: Job Separations and Gender.

07/2010 **Gartner, Hermann, Schank, Thorsten and Schnabel, Claus:** Wage cyclicality under different regimes of industrial relations.

08/2010 **Tinkl, Fabian**: A note on Hadamard differentiability and differentiability in quadratic mean.

## Diskussionspapiere 2009
## Discussion Papers 2009

01/2009 **Addison, John T. and Claus Schnabel:** Worker Directors: A German Product that Didn't Export?

02/2009 **Uhde, André and Ulrich Heimeshoff:** Consolidation in banking and financial stability in Europe: Empirical evidence

03/2009 **Gu, Yiquan and Tobias Wenzel:** Product Variety, Price Elasticity of Demand and Fixed Cost in Spatial Models

04/2009 **Schlüter, Stephan:** A Two-Factor Model for Electricity Prices with Dynamic Volatility

05/2009 **Schlüter, Stephan and Fischer, Matthias**: A Tail Quantile Approximation Formula for the Student t and the Symmetric Generalized Hyperbolic Distribution

06/2009 **Ardelean, Vlad:** The impacts of outliers on different estimators for GARCH processes: an empirical study

07/2009 **Herrmann, Klaus:** Non-Extensitivity versus Informative Moments for Financial Models - A Unifying Framework and Empirical Results

08/2009 **Herr, Annika:** Product differentiation and welfare in a mixed duopoly with regulated prices: The case of a public and a private hospital

09/2009 **Dewenter, Ralf, Haucap, Justus and Wenzel, Tobias**: Indirect Network Effects with Two Salop Circles: The Example of the Music Industry

10/2009 **Stuehmeier, Torben and Wenzel, Tobias**: Getting Beer During Commercials: Adverse Effects of Ad-Avoidance

11/2009 **Klein, Ingo, Köck, Christian and Tinkl, Fabian**: Spatial-serial dependency in multivariate GARCH models and dynamic copulas: A simulation study

12/2009 **Schlüter, Stephan**: Constructing a Quasilinear Moving Average Using the Scaling Function

13/2009 **Blien, Uwe, Dauth, Wolfgang, Schank, Thorsten and Schnabel, Claus:** The institutional context of an "empirical law": The wage curve under different regimes of collective bargaining

14/2009 **Mosthaf, Alexander, Schank, Thorsten and Schnabel, Claus**: Low-wage employment versus unemployment: Which one provides better prospects for women?

## Diskussionspapiere 2008
## Discussion Papers 2008

01/2008 **Grimm, Veronika and Gregor Zoettl**: Strategic Capacity Choice under Uncertainty: The Impact of Market Structure on Investment and Welfare

02/2008 **Grimm, Veronika and Gregor Zoettl**: Production under Uncertainty: A Characterization of Welfare Enhancing and Optimal Price Caps

03/2008 **Engelmann, Dirk and Veronika Grimm**: Mechanisms for Efficient Voting with Private Information about Preferences

04/2008 **Schnabel, Claus and Joachim Wagner**: The Aging of the Unions in West Germany, 1980-2006

05/2008 **Wenzel, Tobias**: On the Incentives to Form Strategic Coalitions in ATM Markets

06/2008 **Herrmann, Klaus**: Models for Time-varying Moments Using Maximum Entropy Applied to a Generalized Measure of Volatility

07/2008 **Klein, Ingo and Michael Grottke:** On J.M. Keynes' "The Principal Averages and the Laws of Error which Lead to Them" - Refinement and Generalisation