ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 15 • 2016
ISSN 1642-168X
AOC, 2016; 15: 75–89
DOI: 10.15678/AOC.2016.1505

Przemysław Jaśko

STATISTICAL ARBITRAGE: A CRITICAL VIEW

Abstract

Statistical arbitrage dynamics is driven by a stationary, autoregressive process known as mispricing. This process approximates the value in time of a portfolio weighted equally to the elements of a cointegration vector of the log-prices processes of related instruments. Statistical arbitrage involves taking either long or short positions on a portfolio according to predictions of mispricing. This paper offers a theoretical analysis of cointegration testing under the conditional heteroscedasticity of the innovations process. Cointegration testing is used in the procedure of searching for the log-price processes of the related instruments that will form a statistical arbitrage portfolio. We also investigate dynamic characteristics of the mispricing process, which is a linear combination (cointegration vector elements are coefficients of it) of related log--prices processes for which the (T)VECM-MGARCH model class is assumed. Under this model assumptions making precise predictions on mispricing process based on past realizations are difficult. This paper can be treated as a starting point for an empirical analysis of statistical arbitrage portfolio construction. Reference is made to theory to describe the challenges which can be faced in constructing a statistical arbitrage portfolio based on cointegration, in modelling the dynamics of mispricing, and in prediction where the innovation process is conditionally heteroscedastic.

Keywords: statistical arbitrage, cointegration, conditional heteroscedasticity, VECM--MGARCH, Breitung cointegration test.
JEL Classification: C320, C580.

1. Introduction. A General Description of the Problem of Statistical Arbitrage

Statistical arbitrage[1] is a form of quantitative trading method which can be classified as a long-short, market neutral and relative pricing strategy. It is based on the assumption that the log-prices of related financial instruments,

Przemysław Jaśko, Cracow University of Economics, Faculty of Management, Computational Systems Department, Rakowicka 27, 31-510 Kraków, Poland, e-mail: jaskop@uek.krakow.pl

[1] As developed in Burgess (2000).

such as a subgroup of index constituent stocks or a term structure of interest rates, are driven by a reduced number of common stochastic trends, and that there is an equilibrium relation between the log-prices of these instruments. Deviations from the levels suggested by the equilibrium relation, caused by idiosyncratic shocks to the log-prices of a particular instrument or subgroup of instruments, are subjected to reversion by arbitrageurs and the related log-prices tend towards new levels at which the equilibrium relationship is satisfied.

Assuming that the equilibrium relation is given by the linear function $\boldsymbol{\beta}'\mathbf{x}_t = 0$ of related log-prices in vector $\mathbf{x}_t$, the process of deviations (also called the mispricing process) defined as $\{y_t = \boldsymbol{\beta}'\mathbf{x}_t\}$ should be a stationary, autoregressive process. In this case, the vector $\boldsymbol{\beta}$ elements are taken as portfolio weights and the value of y_t represents an approximate[2] value of such a portfolio over time. A portfolio with a structure of this kind is known as a statistical arbitrage portfolio or a Beta portfolio. In statistical arbitrage theory $\{y_t\}$, which approximates portfolio value, is a stationary, autoregressive process. When the y_t value deviates from 0, it is expected to move towards zero, which is informed by the level of expected value conditional on the process past. Anyone observing positive – or negative – deviations can then take a short – or long – position in a statistical arbitrage portfolio and make a profit by taking the opposite position when equilibrium is subsequently restored.

We demonstrate in this article that using information only on the expected value of y_t, conditional on the process past, is not sufficient to precisely forecast future movements of y_t values. According to the stylized facts about financial log-return processes (and therefore of log--prices as their cumulative sums), their innovation processes (stochastic input processes to dynamic models) are characterized by conditional heteroscedasticity, which is often of the MGARCH or MSV type, and sometimes also by unconditional heteroscedasticity. Because of this, the same idiosyncratic shocks (innovations) that cause deviations of y_t from the equilibrium level also inflate future conditional variances and covariances of innovations. This is in turn reflected in increased conditional variance of y_t, which is a linear combination of log-prices as shaped by innovations. This increased conditional variance makes it difficult to precisely forecast future movements of y_t values – despite the autoregressive property of y_t.

In statistical arbitrage problem, when we treat the log-prices of related instruments (for example the daily closing log-prices of stocks) as belonging

[2] The approximation is derived in Chan (2011).

to the class of integrated processes (most frequently as $I(1)$ vector processes), cointegration is applied to describe the equilibrium relations between log-prices and the VECM model (including its extensions) as a tool for modelling the dynamics of the log-prices vector process. This process is driven by the common stochastic trends, which makes it a $I(1)$ process, and the $I(0)$ temporary component shaped by an error correction mechanism and the short term dynamics of log-returns (the first differences of log-prices).

This paper offers a theoretical analysis of cointegration testing under the conditional heteroscedasticity of the innovations process. We also investigate dynamic characteristics of the mispricing process, which is a linear combination (coefficients of this combination are equal to cointegration vector $\boldsymbol{\beta}$ elements) of related log-prices processes for which the (T)VECM-MGARCH model class is assumed.

This paper can be treated as a starting point for an empirical analysis in statistical arbitrage portfolio construction. Reference is made to theory to describe the challenges which can be faced in constructing a statistical arbitrage portfolio based on cointegration, in modelling the dynamics of mispricing and in prediction, under a conditionally heteroscedastic innovation process.

We first present a formal definition of statistical arbitrage trading strategy and then consider the impact of innovations with conditional heteroscedasticity on cointegration based statistical arbitrage ability, and their influence on cointegration testing according to the frequentist approach.

2. Statistical Arbitrage

We define (after Jarrow et al. 2012) statistical arbitrage as a zero initial cost, self-financing trading strategy with a discounted cumulative trading profit value $V(n) = \sum_{i=1}^{n} \Delta V(i)$ (also called investor's wealth) for which:

1. $V(0) = 0$,
2. $\lim_{n \to \infty} E^P[V(n)] > 0$,
3. $\lim_{n \to \infty} P(V(n) < 0) = 0$,
4. $\lim_{n \to \infty} \frac{Var^P[V(n)]}{n} = 0 \text{ if } P(V(n) < 0) > 0 \ \forall n < \infty$.

According to this definition, the expected value of discounted cumulative value in statistical arbitrage trading must, asymptotically, be positive. Statistical arbitrage strategy is different from traditional deterministic arbitrage in that it can exhibit negative discounted cumulative value – with positive probability in intermediate finite times – under conditions where the time-averaged variance of cumulative value for infinite time tends to zero and, asymptotically, the probability of a negative value for a trading strategy is zero.

The proponents of statistical arbitrage (Jarrow et al. 2012) assume that the dynamics of the incremental trading profits of statistical arbitrage (investor's wealth) can be described by the process:

$$\Delta V(i) = \mu i^{\theta} + \sigma i^{\lambda} Z_i,$$

where $\{Z_i\} \sim iiN(0,1)$ or $\{Z_i\} \sim MA(1)$.

Inference, if the constructed trading strategy can be considered statistical arbitrage, is based on testing a conjunction of hypotheses on the parameters of an incremental trading profits process: $H_1\colon \mu > 0, H_2\colon \lambda < 0$ and $H3\colon\ \theta > \max\left\{\lambda - \frac{1}{2}, -1\right\}$. An empirical series of investor's wealth deriving from statistical arbitrage trading is used in the testing.

3. Cointegration, the Heteroscedasticity of Model Innovations and Statistical Arbitrage

Before considering cointegrated processes it is necessary to define integrated n-dimensional (vector) processes.

We call the n-dimensional process $\{\mathbf{x}_t\}$ integrated of order 0 the process: $\{\mathbf{x}_t\} \sim I(0) \overset{df}{\Leftrightarrow} \mathbf{x}_t = \sum_{i=0}^{\infty} \boldsymbol{\gamma}_i L^i \boldsymbol{\epsilon}_t$, where L is a lag operator, $\{\boldsymbol{\epsilon}_t\} \sim WN(\mathbf{0}, \boldsymbol{\Sigma})$ (n-dimensional white noise process) and $\sum_{i=0}^{\infty} \boldsymbol{\gamma}_i \neq \mathbf{0}$.

We call the n-dimensional process $\{\mathbf{x}_t\}$ integrated of order d ($d \in \mathbf{Z}$) the processes:

$$\{\mathbf{x}_t\} \sim I(d) \overset{df}{\Leftrightarrow} \{\Delta^d \mathbf{z}_t\} \sim I(0) \text{ and } \{\Delta^{d-1}\mathbf{z}_t\} \nsim I(0).$$

Let us now assume n-dimensional process $\{\mathbf{x}_t\} \sim I(1)$ given by the VAR(k) model:

$$\mathbf{x}_t = \sum_{i=1}^{k} \boldsymbol{\Pi}_i \mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t, t = 1, \ldots, t,$$

with $\{\boldsymbol{\epsilon}_t\} \sim iiN(\mathbf{0}, \boldsymbol{\Sigma})$, represented equivalently by:

$$\Delta\mathbf{x}_t = \boldsymbol{\Pi}\mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i \Delta\mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t, t = 1, \ldots, T,$$

where $\mathbf{\Pi} = \sum_{i=1}^{k} \mathbf{\Pi}_i - \mathbf{I}_n$, $\mathbf{\Pi}_i = -\sum_{j=i+1}^{k} \mathbf{\Pi}_j$, with the characteristic polynomial matrix $\mathbf{A}(z) = (1-z)\mathbf{I}_n - \mathbf{\Pi}\, z - \sum_{i=1}^{k-1} \mathbf{\Gamma}_i (1-z) z^i$.

Additionally we assume $|\mathbf{A}(z)| = 0$ for z such that $|z| > 1$ or $z = 1$. The number of unit roots $z = 1$, is exactly $n - r$. For $z = 1$ we have $|\mathbf{A}(1)| = |-\mathbf{\Pi}| = 0$, implying that $\mathbf{\Pi}$ has reduced rank: $rk(\mathbf{\Pi}) = r < n$. We can thus make factorization $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ where $\dim(\boldsymbol{\alpha}) = \dim(\boldsymbol{\beta}) = n \times r$ and $rk(\boldsymbol{\alpha}) = rk(\boldsymbol{\beta}) = r$.

For the processes $\Delta \mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \mathbf{\Gamma}_i \Delta \mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t$ and $\boldsymbol{\beta}'\mathbf{x}_t$ (which is an r-dimensional process) to have initial conditions such that both will be $I(0)$ processes, it is necessary and sufficient that $|-\boldsymbol{\alpha}_\perp' \dot{\mathbf{A}}(1) \boldsymbol{\beta}_\perp| = |\boldsymbol{\alpha}_\perp' \mathbf{\Gamma} \boldsymbol{\beta}_\perp| \neq 0$, where $\dot{\mathbf{A}}(1) = \frac{\mathrm{d}}{\mathrm{d}z}\mathbf{A}(z)\big|_{z=1}$, $\mathbf{\Gamma} = \mathbf{I}_n - \sum_{i=1}^{k-1} \mathbf{\Gamma}_i$, and $\boldsymbol{\alpha}_\perp$, $\boldsymbol{\beta}_\perp$ are respectively $n \times (n-r)$ matrices of orthogonal complements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, with rank $rk(\boldsymbol{\alpha}_\perp) = rk(\boldsymbol{\beta}_\perp) = n - r$.

When these conditions are met, the Johansen version of the Granger Representation Theorem (Johansen 1995) states that $I(1)$ process $\{\mathbf{x}_t\}$ is cointegrated of order $(1,1)$: $\{\mathbf{x}_t\} \sim CI(1,1)$ and can be equivalently represented as (for $t = 1, \ldots, T$):

$$\Delta \mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \mathbf{\Gamma}_i \Delta \mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t,$$
$$\mathbf{x}_t = \mathbf{C} \sum_{i=1}^{t} \boldsymbol{\epsilon}_i + \mathbf{C}_1(L)\boldsymbol{\epsilon}_t + \mathbf{A},$$

where $\mathbf{C} = \boldsymbol{\beta}_\perp (\boldsymbol{\alpha}_\perp' \mathbf{\Gamma} \boldsymbol{\beta}_\perp)^{-1} \boldsymbol{\alpha}_\perp'$, $\mathbf{C}_1(L)\boldsymbol{\epsilon}_t \sim I(0)$ and $\boldsymbol{\beta}'\mathbf{A} = \mathbf{0}$ ($\mathbf{A}$ is associated with the initial value).

The column vectors from the $\boldsymbol{\beta}$ matrix form the basis of a cointegration space which is the r-dimensional subspace of $\mathbf{R}^n$, where $0 < r < n$ and, for any vector $\mathbf{b} \in \mathrm{sp}(\boldsymbol{\beta})$, we have $\{\mathbf{b}'\mathbf{x}_t\} \sim I(0)$, because $\mathbf{b}'\mathbf{C} = 0$, specifically $\boldsymbol{\beta}'\mathbf{x}_t$ forms an r-dimensional $I(0)$ process.

Summarizing for $\{\mathbf{x}_t\} \sim CI(1,1)$, we have: $\{\mathbf{x}_t\} \sim I(1)$, $\{\Delta \mathbf{x}_t\} \sim I(0)$, $\{\mathbf{y}_t = \boldsymbol{\beta}'\mathbf{x}_t\} \sim I(0)$, additionally $\{\boldsymbol{\beta}_\perp' \Delta \mathbf{x}_t\} \sim I(0)$.

Once the related log-prices have been identified, the central problem in statistical arbitrage is to model and forecast the deviations process. When we assume $r = 1$ (a higher cointegration rank may suggest that the chosen group of assets includes some mutually exclusive subgroups of related log-prices), the deviations process is represented by a scalar process $\{y_t = \boldsymbol{\beta}'\mathbf{x}_t\}$, which is a stationary, autoregressive process.

Unfortunately, when heteroscedastic variance is present in y_t, the autoregressive property is not a sufficient condition for a precise directional forecast of y_t and hence for taking profitable positions on a Beta portfolio based on it.

To demonstrate this, let us make further assumptions that incorporate stylized facts about financial log-returns by extending the VECM model with the *iiN* innovations process.

For most financial log-returns, innovations processes $\{\boldsymbol{\epsilon}_t\}$ show conditional heteroscedasticity, which is usually modelled by one of the many MGARCH variants, and are no longer strict white noise processes. Innovations processes are composed of variables that are not correlated in time, but are not independent in time. Unconditional heteroscedasticity, caused for example by structural breaks that permanently increase the mean dispersion level from a particular moment in time, is also sometimes observed. The heteroscedastic innovations referred to above are embraced by a group of martingale difference sequence (MDS) processes.

Let us consider a VECM-MGARCH[3] model for the log-returns of related stocks with a $CI(1,1)$ cointegrated n-dimensional log-prices process, where $r = 1$ implies $\boldsymbol{\beta}$ composed of only one cointegrating vector. For ease of interpretation we assume that there are no short-term dynamics in the model i.e. $\boldsymbol{\Gamma}_i = \mathbf{0}$, $i = 1, \ldots, k-1$.

VECM-MGARCH model ($t = 1, \ldots, T$):

$$\left.\Delta\mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t \right\} \text{ VECM part}$$

$$\left.\begin{array}{l} \boldsymbol{\epsilon}_t = \mathbf{H}_t^{1/2}\boldsymbol{\eta}_t \\ \mathbf{H}_t = \mathbf{H}(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1}, \ldots, \mathbf{H}_{t-1}, \ldots) \\ \{\boldsymbol{\eta}_t\} \sim iid(\mathbf{0}, \mathbf{I}_n) \end{array}\right\} \text{ MGARCH part,}$$

where $\mathbf{H}_t = \mathbf{H}_t^{1/2}(\mathbf{H}_t^{1/2})'$ is the "square root" decomposition of $\mathbf{H}_t = [h_{ij,t}]_{i,j=1,\ldots,n}$, representing a covariance matrix in moment t conditional on the past of the process, $\mathbf{H}$ is a matrix function representing MGARCH, with some previous values of $\boldsymbol{\epsilon}_{t-j}\boldsymbol{\epsilon}'_{t-j}$ and $\mathbf{H}_{t-j}$ as arguments, $\{\boldsymbol{\eta}_t\}$ is an n-dimensional process of independent standardized variables, having for example a multivariate normal distribution or a multivariate t-Student

[3] So that more general statements can be made, the variant of the MGARCH model is not precisely specified.

distribution with vector mean $\mathbf{0}$ and covariance $\mathbf{I}_n$ (n-dimensional unit matrix), but also with asymmetric counterparts of these distributions.

The deviation process (mispricing process) for this model, with cointegration rank $r = 1$ and cointegration vector $\boldsymbol{\beta} = [\beta_1 \dots \beta_n]'$, is a scalar process $\{y_t\}$ given by:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t = (1 + \boldsymbol{\beta}'\boldsymbol{\alpha})\boldsymbol{\beta}'x_{t-1} + \boldsymbol{\beta}'\boldsymbol{\epsilon}_t$$
$$y_t = \phi y_{t-1} + \varepsilon_t^y,$$

where $\phi = (1 + \boldsymbol{\beta}'\boldsymbol{\alpha})$, $\phi \in (-1,1)$ for $\{\mathbf{x}_t\} \sim CI(1,1)$ and $\varepsilon_t^y = \boldsymbol{\beta}'\boldsymbol{\epsilon}_t$.

The deviations process $\{y_t\}$ is in fact stationary and autoregressive, but let us investigate its properties, such as its expected value and variance conditional on the past of the process.

Let $\Psi_t = \sigma(\mathbf{x}_s, s \leq t)$ be a σ-algebra generated by the process $\{\mathbf{x}_s\}$ up to moment t.

$$E(y_t \mid \Psi_{t-1}) = \phi y_{t-1}$$

$$V(y_t \mid \Psi_{t-1}) = V(\varepsilon_t^y \mid \Psi_{t-1}) = V(\boldsymbol{\beta}'\boldsymbol{\epsilon}_t \mid \Psi_{t-1}) =$$
$$= \sum_{i=1}^{n} \beta_i^2 V(\varepsilon_{ti} \mid \Psi_{t-1}) + 2\sum_{i=1}^{n}\sum_{j>i} \beta_i \beta_j Cov(\varepsilon_{ti}, \varepsilon_{tj} \mid \Psi_{t-1}) \Leftrightarrow$$
$$\Leftrightarrow V(y_t \mid \Psi_{t-1}) = \sum_{i=1}^{n} \beta_i^2 h_{ii,t} + 2\sum_{i=1}^{n}\sum_{j>i} \beta_i \beta_j h_{ij,t}.$$

The conditional variance form for y_t shows that in general conditions $\{\varepsilon_t^y\}$ is not given by a univariate GARCH model. The first component in y_t conditional variance $\sum_{i=1}^{n} \beta_i^2 h_{ii,t}$ is always positive and cumulates (with positive multipliers β_i^2) the conditional variances $h_{ii,t}$ of the univariate constituents of $\boldsymbol{\epsilon}_t$ from the innovations process, thus increasing the value of $V(y_t \mid \Psi_{t-1})$. The second component, which is twice $\sum_{i=1}^{n}\sum_{j>i} \beta_i \beta_j h_{ij,t}$, can – but does not have to – take negative values and, in some conditions, can reduce the level of conditional variance of y_t. The sign of the second component depends on the signs of parameters β_i, β_j and on the conditional covariances $h_{ij,t}$ for the constituents of $\boldsymbol{\epsilon}_t$.

These findings confirm that, because of increased conditional variance $V(y_{t+1} \mid \Psi_t)$, information about $E(y_{t+1} \mid \Psi_t)$ is not on its own a precise indicator of future y_{t+1} value movements. Moreover, the conditional distribution $\boldsymbol{\epsilon}_{t+1} \mid \Psi_t$ type and parameters strongly affect the conditional distribution of $y_{t+1} \mid \Psi_t$ as a linear combination of $\mathbf{x}_{t+1} \mid \Psi_t$ constituents.

If useful predictions are to be made in a case such as this, all of the information available on the conditional distribution of $y_{t+1} \mid \Psi_t$ should be exploited rather than selected parameters only. From the conditional distribution of $y_{t+1} \mid \Psi_t$ we can derive quantile forecasts or assess the probability of up or down movement from the current y_t value. Because of the complex shape of the conditional distribution $y_{t+1} \mid \Psi_t$, which can be asymmetric, and owing to the complicated relations describing its parameters, there may occur a situation in which $\operatorname{sgn}[E(y_{t+1} \mid \Psi_t) - y_t]$ gives a specific direction for future movement while the information on the conditional distribution of $y_{t+1} \mid \Psi_t$ suggests that movement in the opposite direction is more probable. Here, the autoregressive tendency to revert, which was expected, is dominated by overdispersion and statistical arbitrage cannot be realized.

We have simulated a sample series of a length of $T = 1000$ simulated from VECM-DCC-GARCH ($n = 2$, $r = 1$ with a 2×1 cointegration vector $\boldsymbol{\beta}$; the model has no short-term dynamics) for $\mathbf{x}_t$, $\Delta\mathbf{x}_t$, $y_t = \boldsymbol{\beta}'\mathbf{x}_t$. Figure 1 shows scatter plot for $\mathbf{x}_t = (x_{t1}, x_{t2})'$, Figures 2 to 4 are plots of the time series concerned.

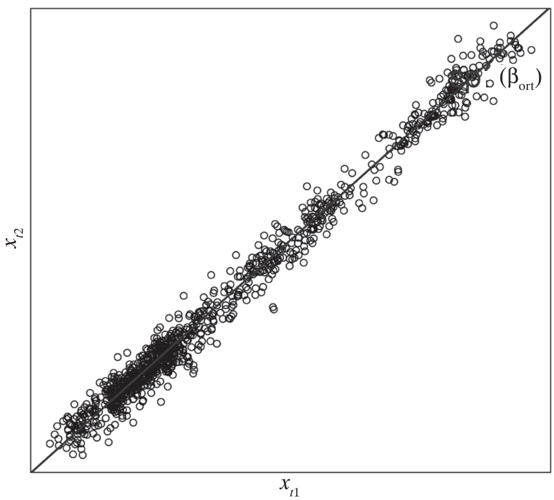

Fig. 1. Scatter Plot for $(x_{t1}, x_{t2})'$ with Marked Attractor Given by the Subspace $\mathrm{sp}(\boldsymbol{\beta}_\perp)$

Source: author’s own research.

The one-dimensional subspace spanned by the $\boldsymbol{\beta}$ orthogonal complement, denoted by $\mathrm{sp}(\boldsymbol{\beta}_\perp)$, forms an attractor for process $\{\mathbf{x}_t\}$; as for $\mathbf{x}_t^* = c \cdot \boldsymbol{\beta}_\perp$ with arbitrary $c \neq 0$, we have $y_t^* = \boldsymbol{\beta}'\mathbf{x}_t^* = c \cdot \boldsymbol{\beta}'\boldsymbol{\beta}_\perp = 0$ and, for the assumed model, $y_t = \boldsymbol{\beta}'\mathbf{x}_t$ is driven towards 0.

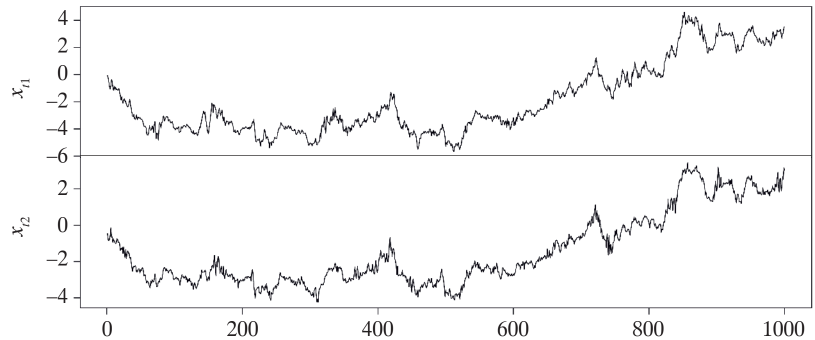

Fig. 2. Simulated Log-prices $\mathbf{x}_t = (x_{t1}, x_{t2})'$ Time Series

Source: author's own research.

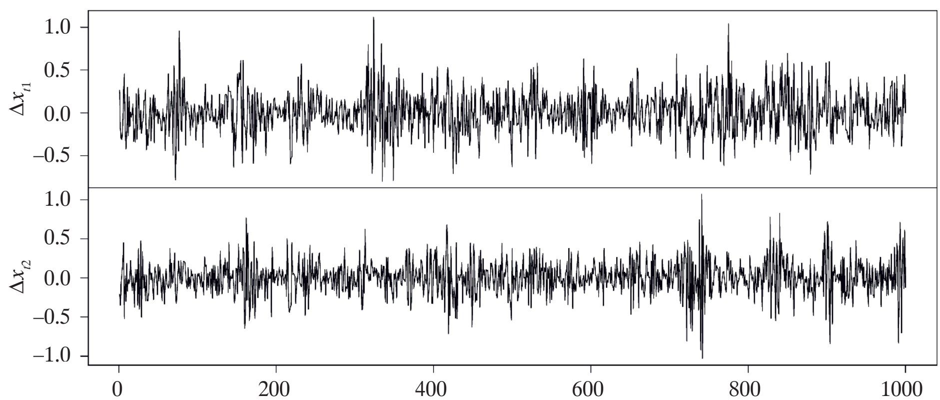

Fig. 3. Simulated Log-returns $\Delta\mathbf{x}_t = (\Delta x_{t1}, \Delta x_{t2})'$ Time Series

Source: author's own research.

The VECM-MGARCH may be too restrictive in its construction, since it is suggested that because of transaction costs, only higher absolute deviations from the equilibrium relation are corrected by arbitrageurs. An extension to the VECM part of the model, known as TVECM or Threshold VECM, was proposed to take account of this (Balke & Fomby 1997). In this case TVECM assumes three regimes and one cointegrating vector, $r = 1$:

$$\Delta\mathbf{x}_t = \sum_{m=1}^{3}\left(\boldsymbol{\alpha}^{(m)}\boldsymbol{\beta}^{(m)}{}'\mathbf{x}_{t-1} + \sum_{i=1}^{k-1}\boldsymbol{\Gamma}_i^{(m)}\Delta\mathbf{x}_{t-i} + \boldsymbol{\epsilon}_t^{(m)}\right)\cdot I(c_{m-1} < y_{t-1} \le c_m),$$

where I is an indicator function and for the middle regime $m = 2$ we have: $0 \in (c_1, c_2]$, $\boldsymbol{\alpha}^{(2)} \equiv 0$ which means there is no cointegration in the middle regime and $y_t = \boldsymbol{\beta}'\mathbf{x}_t \sim I(1)$ for $c_1 < y_{t-1} \le c_2$.

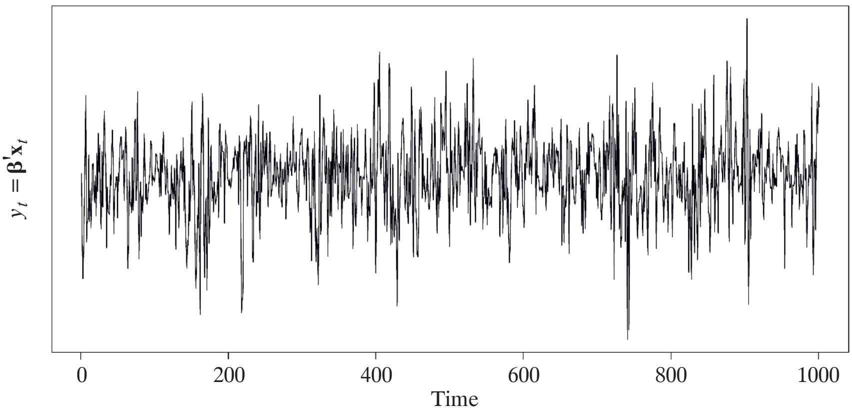

Fig. 4. Simulated Realization of Deviations (Mispricing) Process $y_t = \boldsymbol{\beta}'\mathbf{x}_t$

Source: author's own research.

In this case, because of nonlinear dynamics, the model does not have the representation stated by the Granger Representation Theorem.

To analyze the properties of vector processes with non-linear dynamics, concerning order of integration and cointegration, the definitions of integrated and cointegrated processes need to be extended.

The extended definition of the $I(0)$ n-dimensional (vector) process makes use of the functional central limit theorem (FCLT), whose formal aspects are described by Davidson (1994).

We call the n-dimensional process $\{\mathbf{x}_t\}$ an $I(0)$ process $\Leftrightarrow \forall\, a \in [0,1]$ and $T \to \infty$: $T^{-\frac{1}{2}} \sum_{t=1}^{\lfloor aT \rfloor} \mathbf{x}_t \xrightarrow{d} \boldsymbol{\Sigma}_{\mathbf{x}}^{1/2} \mathbf{W}(a)$, where d symbolizes weak convergence (convergence in distribution), $\lfloor \cdot \rfloor$ is a floor function, $\mathbf{W}(a)$ is an n-dimensional standard Wiener process, $\boldsymbol{\Sigma}_{\mathbf{x}} = \lim_{T \to \infty} T^{-1} Cov\left(\sum_{t=1}^{T} \mathbf{x}_t\right)$ is called a long-term covariance matrix and $\boldsymbol{\Sigma}_{\mathbf{x}}^{1/2}$ is its "square root" matrix. The definition of the processes for vector $I(d)$ remains unchanged.

In this extended approach, cointegration is defined without appealing to an explicitly specified model. In this way it can embrace models with different types of short-term and error-correction dynamics.

Let $\{\mathbf{x}_t\} \sim I(1)$ with respect to the extended definition. We additionally assume decomposition of the invertible matrix $\tilde{\boldsymbol{\beta}} = [\boldsymbol{\beta}_\perp, \boldsymbol{\beta}]$, where $\dim(\boldsymbol{\beta}) = n \times r$, $\dim(\boldsymbol{\beta}_\perp) = n \times (n - r)$, and $0 < r < n$ and $\boldsymbol{\beta}'\boldsymbol{\beta}_\perp = \mathbf{0}$.

Process $\{\mathbf{x}_t\}$ is $CI(1,1)$ if we can decompose it into two components: $\tilde{\boldsymbol{\beta}}'\mathbf{x}_t = \begin{bmatrix} \boldsymbol{\beta}_\perp' \\ \boldsymbol{\beta}' \end{bmatrix} \mathbf{x}_t = \begin{bmatrix} \mathbf{u}_t \\ \mathbf{y}_t \end{bmatrix}$, for which $T^{-\frac{1}{2}} \mathbf{u}_{[aT]} \xrightarrow{d} \mathbf{W}(a) \sim I(1)$ and $T^{-2} \sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}_t' = o_p(1)$, where $\mathbf{W}(a)$ is a $(n - r)$-dimensional standard Wiener process.

Here $\{\mathbf{y}_t = \boldsymbol{\beta}'\mathbf{x}_t\}$ represents a transitory component, which can also be generated by a nonlinear process with a short memory. In addition, $\boldsymbol{\beta}$ spans an r-dimensional cointegration space. $\{\mathbf{u}_t = \boldsymbol{\beta}_\perp'\mathbf{x}_t\}$, on the other hand, is a stochastic trend component, which is “variance dominating”. This means that $\{\mathbf{u}_t\}$ diverges at a faster rate than $\{\mathbf{y}_t\}$.

4. Difficulties with Inference on Cointegration in the Case of Heteroscedastic Innovations

This paper offers a brief discussion of only the frequentist approach to testing cointegration under the heteroscedastic innovations of a specific type.

Classical Johansen cointegration rank tests associated with the $CI(1,1)$ process VECM model with *iiN* innovations, known as the maximum eigenvalue test (cointegration rank: H_0: r vs. H_1: $r + 1$) and the trace test (cointegration rank: H_0: r vs. H_1: $n \Leftrightarrow \{\mathbf{x}_t\} \sim I(0)$), under the null hypotheses have asymptotic distributions, which are derived with the use of FCLT and specified as the functionals of the standard Wiener process.

It has been shown (Cavaliere, Rahbek & Taylor 2010) that when we attenuate assumptions about an innovations process from *iiN* to one that belongs to the MDS class of processes, which includes conditionally and unconditionally heteroscedastic processes, Johansen tests will weakly converge to the same asymptotic distributions.

In VECM models with heteroscedastic innovations, Johansen cointegration rank tests for finite-length samples are regarded as quasi-likelihood ratio tests because they use a likelihood function for the VECM model with *iiN* innovations. These Quasi-LR tests use asymptotic critical values, which is reflected in moderate to high test-size distortions. In a simulation study of Johansen tests using innovations with an MGARCH type of conditional heteroscedasticity (Maki 2013), a true null hypothesis of no cointegration ($r = 0$) was more frequently rejected than the nominal critical level assumed.

To improve the performance of the Johansen Quasi-LR tests for finite--length samples, a wild bootstrap procedure was suggested (Cavaliere, Rahbek & Taylor 2010). Unlike other bootstrap methods, such as the *iid* bootstrap (Swensen 2006), wild bootstrap makes it possible to retain the heteroscedasticity structure of the original series. In a single wild bootstrap replication, Quasi-Maximum Likelihood (QML) estimated VECM model errors $\{\boldsymbol{\epsilon}_t\}_{t=1}^{T}$ are multiplicatively distorted by a univariate[4] $iid(0, 1)$ process $\{\omega_t\}_{t=1}^{T}$ and a new series of $\Delta\mathbf{x}_t^b$ is constructed using $\Delta\mathbf{x}_t^b = \hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\beta}}'\mathbf{x}_{t-1}^b +$ $+\sum_{i=1}^{k}\boldsymbol{\Gamma}_i\Delta\mathbf{x}_{t-1}^b + \boldsymbol{\epsilon}_t^b$, $t = 1, \ldots, T$, where $\boldsymbol{\epsilon}_t^b = \omega_t \cdot \boldsymbol{\epsilon}_t$ with $\{\omega_t\}_{t=1}^{T} \sim iid(0,1)$, $\Delta\mathbf{x}_0^b = (\mathbf{x}_0, \mathbf{x}_{-1}, \ldots, \mathbf{x}_{-k+1})'$.

A wild bootstrap p-value of a Johansen quasi-LR test with a null hypothesis of cointegration rank r, for B replications of wild bootstrap and sample length T, is calculated by: $\tilde{p}_{r,T} = B^{-1}\sum_{b=1}^{B} I(Q_{r,b} > Q_r)$, where I is an indicator function, $Q_{r,\,b}$ is a quasi-LR test value calculated for a VECM model estimated using series $\Delta\mathbf{x}_t^b$ constructed in a b-th replication of the wild bootstrap procedure, and Q_r is a quasi-LR test value calculated for a VECM model estimated using the genuine series $\Delta\mathbf{x}_t$.

Simulations (Cavaliere, Rahbek & Taylor 2008, 2010) under the null hypothesis of no cointegration and MGARCH heteroscedasticity innovations or unconditional heteroscedasticity innovations, have shown a reduction in test size distortion for the presented wild bootstrap variant in comparison to tests using asymptotic critical values for quasi-LR Johansen rank tests. These bootstrap tests are associated with a VECM model that assumes linear error-correction and short-term dynamics.

Some cointegration tests assume in their alternative hypotheses models with a specific type of nonlinear error-correction and short-term dynamics, but according to simulations they suffer from unacceptably large test-size distortions under MGARCH heteroscedastic innovations (Maki 2013). It is of more benefit in the statistical arbitrage problem to use cointegration tests that do not require advance specification of the model dynamics.

The extended definitions of integrated and cointegrated processes presented earlier in this paper can be referred to the Breitung cointegration rank test (Breitung 2002), which is asymptotically free of the nuisance parameter of long-term covariance, influenced by short-term dynamics

[4] Most frequently for process variables we assume Rademacher, standard normal or some discrete asymmetric distribution with $E(\omega_t) = 0$ and $E(\omega_t^2) = 1$.

(linear/nonlinear, number of lags included) and by potential conditional heteroscedasticity and parameters related to them. The Breitung cointegration test can be conducted without advance specification of a model. This is a very important aspect because in the statistical arbitrage problem it is not known in advance which assets have related log-prices processes. Specifying log-price models for numerous subgroups from an adopted universe of assets would be problematic. Instead, subgroups of cointegrated log-prices need to be identified by automatic searching, and Breitung test p-values (with a null hypothesis of no cointegration) can be applied to measure the strength of the relationships. This is a combinatorial optimization problem, which can be solved using a genetic algorithm with binary coding of solutions (with 1 when the asset log-price is included in the relationship) and a fitness function defined, for example, as 1 – p-value of a test with a null hypothesis of no cointegration.

There follows a short discussion of the Breitung cointegration rank test.

Let $\mathbf{E}_T = \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}'_t$ and $\mathbf{F}_T = \sum_{t=1}^{T} \mathbf{X}_t \mathbf{X}'_t$, where $\mathbf{X}_t = \sum_{i=1}^{t} \mathbf{x}_i$. The Breitung cointegration test incorporates the solution of a generalized eigenvalue problem:

$$\left| \lambda \mathbf{F}_T - \mathbf{E}_T \right| = 0.$$

For eigenvalue λ_j (j = 1, …, n) we have:

$$\lambda_j = \frac{\mathbf{v}_j' \mathbf{E}_T \mathbf{v}_j}{\mathbf{v}_j' \mathbf{F}_T \mathbf{v}_j},$$

so when $\mathbf{v}_j$ belongs to sp($\boldsymbol{\beta}_\perp$) we have[5]: $\mathbf{v}_j' \mathbf{E}_T \mathbf{v}_j = O_p\left(T^2\right)$, $\mathbf{v}_j' \mathbf{F}_T \mathbf{v}_j = O_p\left(T^4\right)$ and $\lambda_j = O_p(T^{-2})$. On the other hand, when $\mathbf{v}_j \in sp(\boldsymbol{\beta})$ then for $T \to \infty$: $T^2 \lambda_j \to \infty$.

The Breitung test considers hypothesis H_0: $n - r$ common stochastic trends (r cointegration rank) against H_1: $< n - r$ common stochastic trends ($> r$ cointegration rank) and employs statistic:

$$\Lambda_{n-r} = T^2 \sum_{j=1}^{n-r} \lambda_j,$$

where $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ are eigenvalues from the solution of a generalized eigenproblem.

Under the null hypothesis the test statistic has an asymptotic distribution derived using FCLT, which is a trace of a specified functional of (n – r)-‑dimensional standard Wiener process defined on [0, 1]. This distribution

[5] Derivations can be found in Breitung (2002).

is free of the nuisance parameter of long-term covariance. Under the alternative hypothesis, test statistic tends asymptotically to infinity, which means that the test has a right-side critical area.

According to the results of simulations (Maki 2013), the use of the Breitung cointegration test is recommended for samples of finite length, when the innovations are characterized by MGARCH conditional heteroscedasticity and a null hypothesis assumes no cointegration (the Breitung test has minimal size distortion among considered tests).

It must not be forgotten that when conditional or unconditional heteroscedasticity of innovations exerts a strong influence, the cointegration results returned by the tests can be spurious.

5. Conclusion

Cointegration between the log-prices of related assets is a necessary, but not a sufficient condition for the statistical arbitrage opportunity to hold. Idiosyncratic shocks that cause deviations from the equilibrium relation also increase the dispersion of the mispricing process. In this way the autoregressive tendency of the mispricing process (whose values approximate the value of the statistical arbitrage portfolio over time) can be masked by inflated conditional variance. Future movements of the mispricing process can be hard to predict and also opposite to those suggested by the expected value conditional on the process past. Another difficulty in implementing a strategy of statistical arbitrage under heteroscedastic innovations is the increased chance (with respect to the critical level assumed in the test) of finding false log-price relations in many types of tests with a null hypothesis of no cointegration.

Bibliography

Balke, N. S. and Fomby, T. B. (1997) "Threshold Cointegration". *International Economic Review* 38(3): 627–45, https://doi.org/10.2307/2527284.

Breitung, J. (2002) "Nonparametric Tests for Unit Roots and Cointegration". *Journal of Econometrics* 108 (2): 343–63, https://doi.org/10.1016/s0304-4076(01)00139-7.

Burgess, A. N. (2000) *A Computational Methodology for Modelling the Dynamics of Statistical Arbitrage*. PhD dissertation. London: University of London.

Cavaliere, G., Rahbek, A. and Taylor, A. M. R. (2008) *Testing for Co-integration in Vector Autoregressions with Non-stationary Volatility*. CREATES Research Paper No. 2008–50, Copenhagen: University of Copenhagen.

Cavaliere, G., Rahbek, A. and Taylor, A. M. R. (2010) "Cointegration Rank Testing under Conditional Heteroskedasticity". *Econometric Theory* 26 (6): 1719–60, https://doi.org/10.1017/s0266466609990776.

Chan, N. H. (2011) *Time Series: Applications to Finance with R and S-Plus*. New York: John Wiley & Sons.
Davidson, J. (1994) *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford: Oxford University Press.
Jarrow, R., Teo, M., Tse, Y. K. and Warachka, M. (2012) "An Improved Test for Statistical Arbitrage". *Journal of Financial Markets* 15 (1): 47–80, https://doi.org/10.1016/j.finmar.2011.08.003.
Johansen, S. (1995) *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford–New York: Oxford University Press.
Maki, D. (2013) "The Influence of Heteroskedastic Variances on Cointegration Tests: A Comparison Using Monte Carlo Simulations". *Computational Statistics* 28 (1): 179–98, https://doi.org/10.1007/s00180-011-0293-x.
Swensen, A. R. (2006) "Bootstrap Algorithms for Testing and Determining the Cointegration Rank in VAR Models". *Econometrica* 74 (6): 1699–1714, https://doi.org/10.1111/j.1468-0262.2006.00723.x.

Abstract

Arbitraż statystyczny – ujęcie krytyczne

Rozpatrywana w ramach strategii arbitrażu statystycznego dynamika procesu odchyleń od równowagi (*mispricing process*) ma charakter autoregresyjnego procesu stacjonarnego. Proces ten reprezentuje w przybliżeniu wartość w czasie portfela z wagami odpowiadającymi elementom wektora kointegracyjnego dla procesów logarytmów cen powiązanych instrumentów. Strategia polega na zajmowaniu długich bądź krótkich pozycji na wspomnianym portfelu na podstawie prognoz dotyczących kształtowania się procesu odchyleń od równowagi. W artykule przeprowadzono na gruncie teoretycznym analizę dotyczącą testowania kointegracji w przypadku warunkowej heteroskedastyczności procesów innowacji. Testy kointegracji wykorzystywane są w procedurze poszukiwania powiązanych procesów logarytmów cen instrumentów, które będą tworzyć portfel arbitrażu statystycznego. W pracy rozważano także charakter dynamiki procesu odchyleń od równowagi, będącego liniową kombinacją (elementy wektora kointegracji są jej parametrami) powiązanych procesów logarytmów cen, dla których zakłada się, że są generowane przez klasę modeli (T)VECM-GARCH. Przy takich założeniach dotyczących modelu procesów stawianie precyzyjnych prognoz dotyczących dynamiki procesu odchyleń od równowagi na podstawie przeszłych realizacji jest utrudnione. Praca może być punktem wyjścia do analiz empirycznych dotyczących konstrukcji portfela arbitrażu statystycznego. Wykorzystując rozważania teoretyczne, wskazuje się problemy, które można napotkać w badaniach empirycznych dotyczących konstrukcji opartej na kointegracji strategii arbitrażu statystycznego oraz modelowania i prognozowania procesu odchyleń od równowagi w przypadku warunkowej heteroskedastyczności procesu innowacji.

Słowa kluczowe: arbitraż statystyczny, kointegracja, warunkowa heteroskedastyczność, VECM-MGARCH, test kointegracji Breitunga.