Statistical arbitrage – critical view

Przemysław Jaśko[1]

Abstract

Statistical arbitrage is a trading strategy on portfolio with value driven by stationary, autoregressive process. This article briefly presents role of innovations with conditional heteroscedasticity on cointegration based statistical arbitrage ability, and their influence on cointegration testing according to frequentist approach.

Keywords: *statistical arbitrage, cointegration, conditional heteroscedasticity*

JEL Classification: C320, C580

1. Introduction – general description of statistical arbitrage problem

Statistical arbitrage developed by (Burgess, 2000), belongs to the group of quantitative trading strategies. It is based on the assumption that log-prices of related financial instruments (ex. some subgroup of index constituent stocks, term structure of interest rates) are driven by reduced number of common stochastic trends and there exists equilibrium relation between log-prices of these instruments. Moreover deviations from levels suggested by equilibrium relation, caused by idiosyncratic shocks on log-prices of particular instrument (or subgroup of instruments), are reverted by arbitrageurs and related log-prices tends towards new levels in which equilibrium relation is met. Assuming that equilibrium relation is given by linear function $\boldsymbol{\beta}'\mathbf{x}_t = 0$ of related log-prices in vector $\mathbf{x}_t$, process of deviations (also called mispricing process) defined as $\{y_t = \boldsymbol{\beta}'\mathbf{x}_t\}$ should be stationary, autoregressive process. In such a case vector elements $\boldsymbol{\beta}$ are taken as portfolio weights and value of y_t represents approximately value of such portfolio in time (approximation is derived in (Chan, 2011)). Considered portfolio is called statistical arbitrage or Beta portfolio. In statistical arbitrage theory $\{y_t\}$ which approximates portfolio value, is a stationary, autoregressive process and when y_t value deviates from 0, it is expected in the following time moments to take moves towards zero (what is signaled by the level of expected value conditional on the process past). Knowing that, one observing positive (negative) deviations could take long (short) position in statistical arbitrage portfolio and realize profit by taking opposite position when equilibrium is subsequently restored. In this article we are showing that relying on expected value of y_t

[1] Corresponding author: Cracow University of Economics, Department of Computational Systems, Rakowicka 27, 31-510 Kraków, Poland, e-mail: jaskop@uek.krakow.pl.

conditional on process past is not sufficient to precisely forecast future movement of y_t values. According to stylized facts about financial log-returns processes (and therefore log-prices as their cumulative sums) their innovations processes are characterized by conditional (typically MGARCH or MSV type) or unconditional heteroscedasticity. Because of that, the same idiosyncratic shocks (innovations), which cause deviations of y_t from equilibrium level also inflate its future conditional variances, what despite autoregressive property of y_t considerably limits ability to make precise value movements forecasts.

In statistical arbitrage problem, when we treat log-prices (for example daily closing log-prices of stocks) of related instruments as belonging to the class of integrated processes (most frequently as $I(1)$ vector processes), cointegration come in mind as a phenomenon which can describe equilibrium relations between log-prices and VECM model and its extensions, as tool for modeling dynamics of log-prices vector process, which is driven by common stochastic trends which make it $I(1)$ process and temporary component $I(0)$ shaped by error correction mechanism and short term dynamics of log-returns (first differences of log-prices).

2. Cointegration, heteroscedasticity of model innovations and statistical arbitrage problem

Before we consider cointegrated processes we need to define integrated n-dimensional (vector) processes.

We call n-dimensional process $\{\mathbf{x}_t\}$ integrated of order 0 process: $\{\mathbf{x}_t\} \sim I(0) \Leftrightarrow \mathbf{x}_t = \sum_{i=0}^{\infty} \boldsymbol{\gamma}_i L^i \boldsymbol{\varepsilon}_t$, where L is a lag operator, $\{\boldsymbol{\varepsilon}_t\} \sim WN(\mathbf{0}, \boldsymbol{\Sigma})$ (n-dimensional white noise process) and $\sum_{i=0}^{\infty} \boldsymbol{\gamma}_i \neq \mathbf{0}$. We call n-dimensional process $\{\mathbf{x}_t\}$ integrated of order d ($d \in \mathbf{Z}$) process: $\{\mathbf{x}_t\} \sim I(d) \Leftrightarrow \{\Delta^d \mathbf{z}_t\} \sim I(0)$ and $\{\Delta^{d-1} \mathbf{z}_t\} \nsim I(0)$

Let now assume n-dimensional process $\{\mathbf{x}_t\} \sim I(1)$ given by VAR(k) model $\mathbf{x}_t = \sum_{i=1}^{k} \boldsymbol{\Pi}_i \mathbf{x}_{t-i} + \boldsymbol{\varepsilon}_t, t = 1, \ldots, t$, with $\{\boldsymbol{\varepsilon}_t\} \sim iiN(\mathbf{0}, \boldsymbol{\Sigma})$, represented equivalently by $\Delta \mathbf{x}_t = \boldsymbol{\Pi} \mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i \Delta \mathbf{x}_{t-i} + \boldsymbol{\varepsilon}_t, t = 1, \ldots, T$, where $\boldsymbol{\Pi} = \sum_{i=1}^{k} \boldsymbol{\Pi}_i - \mathbf{I}_n$, $\boldsymbol{\Pi}_i = -\sum_{j=i+1}^{k} \boldsymbol{\Pi}_j$, with characteristic polynomial matrix $\mathbf{A}(\mathrm{z}) = (1-z)\mathbf{I}_n - \boldsymbol{\Pi} z - \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i (1-z) z^i$. Additionally we assume $\left|\mathbf{A}(z)\right| = 0$ 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 $\left|\mathbf{A}(1)\right| = \left|-\mathbf{\Pi}\right| = 0$, implying that $\mathbf{\Pi}$ has reduced rank: $rk(\mathbf{\Pi}) = r < n$. So we can 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 processes $\Delta\mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \sum_{i=1}^{k-1}\mathbf{\Gamma}_i\Delta\mathbf{x}_{t-i} + \boldsymbol{\varepsilon}_t$ and $\boldsymbol{\beta}'\mathbf{x}_t$ (which is r-dimensional process) to exist initial conditions such that both will be $I(0)$ processes, it is necessary and sufficient that $\left|-\boldsymbol{\alpha}_\perp'\dot{\mathbf{A}}(1)\boldsymbol{\beta}_\perp\right| = \left|\boldsymbol{\alpha}_\perp'\mathbf{\Gamma}\boldsymbol{\beta}_\perp\right| \neq 0$, 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 above mentioned conditions are met Johansen version of Granger Representation Theorem 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 $t = 1, \ldots, T$):

$$\Delta\mathbf{x}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{x}_{t-1} + \sum_{i=1}^{k-1}\mathbf{\Gamma}_i\Delta\mathbf{x}_{t-i} + \boldsymbol{\varepsilon}_t \Leftrightarrow \mathbf{x}_t = \mathbf{C}\sum_{i=1}^{t}\boldsymbol{\varepsilon}_i + \mathbf{C}_1(L)\boldsymbol{\varepsilon}_t + \mathbf{A}, \quad \text{where} \quad \mathbf{C} = \boldsymbol{\beta}_\perp\left(\boldsymbol{\alpha}_\perp'\mathbf{\Gamma}\boldsymbol{\beta}_\perp\right)^{-1}\boldsymbol{\alpha}_\perp',$$

$\mathbf{C}_1(L)\boldsymbol{\varepsilon}_t \sim I(0)$ and $\boldsymbol{\beta}'\mathbf{A} = \mathbf{0}$ ($\mathbf{A}$ is associated with initial value).

Column vectors from $\boldsymbol{\beta}$ matrix form a basis for cointegration space which is 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 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 $\{\boldsymbol{\beta}_\perp'\Delta\mathbf{x}_t\} \sim I(0)$.

When related log-prices are already identified, the central part in statistical arbitrage problem is to model and forecast deviations process. When we assume $r = 1$ (higher cointegration rank may suggest that chosen group of assets includes some mutually exclusive subgroups of related log-prices) deviations process is represented by scalar process $\{\mathrm{y}_t = \boldsymbol{\beta}'\mathbf{x}_t\}$, which is stationary, autoregressive process. Unfortunately when heteroscedastic variance in y_t is present, autoregressive property is not a sufficient condition to make precise directional forecast of y_t and take profitable positions on Beta portfolio based on them.

To show this let make another assumptions which will incorporate stylized facts about financial log-returns by extending VECM model with iiN innovations process.

For most financial log-returns, innovations processes $\{\boldsymbol{\varepsilon}_t\}$ show conditional heteroscedasticity (ex. MGARCH type) and are no longer strict white noise processes. They are composed of uncorrelated but not independent in time variables. Sometimes also unconditional heteroscedasticity is observed, caused for example by structural breaks, which permanently increases mean dispersion level from some moment in time. Mentioned types of innovations heteroscedasticity are embraced by a group of martingale difference sequence (MDS) processes.

Consider VECM-MGARCH[2] model for log-returns of related stocks with $CI(1,1)$ cointegrated n-dimensional log-prices process, with $r=1$ implying $\boldsymbol{\beta}$ composed of only one cointegrating vector. For ease of interpretation we assume that there is no short-term dynamics in model i.e. $\boldsymbol{\Gamma}_i = \mathbf{0}$, $i=1,\ldots,k-1$.

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

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

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

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

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

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t = \left(1+\boldsymbol{\beta}'\boldsymbol{\alpha}\right)\boldsymbol{\beta}'\mathbf{x}_{t-1} + \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t,$$

$$y_t = \phi y_{t-1} + \varepsilon_t^y$$

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

[2] Type of MGARCH model is not precisely specified here to make more general statements.

Deviations process $\{y_t\}$ is in fact stationary autoregressive, but let analyze its properties such as expected value and variance conditional on process past.

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

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

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

$$\Leftrightarrow V\left(y_t \mid \Psi_{t-1}\right) = \sum_{i=1}^{n} \beta_i^2 h_{ii,t} + 2\sum_{i=1}^{n}\sum_{j>i} \beta_i \beta_j h_{ij,t}.$$

Conditional variance form for y_t shows that in general conditions $\{\varepsilon_t^y\}$ is not given by univariate GARCH model. 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) conditional variances $h_{ii,t}$ of univariate constituents of $\boldsymbol{\varepsilon}_t$ from innovations process, increasing value of $V\left(y_t \mid \Psi_{t-1}\right)$. Second component which is twice $\sum_{i=1}^{n}\sum_{j>i} \beta_i \beta_j h_{ij,t}$, may take negative values (but not necessarily), and in some conditions may reduce level of conditional variance of y_t. Sign of second component depends on signs of parameters β_i, β_j and conditional covariances $h_{ij,t}$ for constituents of $\boldsymbol{\varepsilon}_t$. This findings confirm that because of increased conditional variance $V\left(y_{t+1} \mid \Psi_t\right)$, information about $E\left(y_{t+1} \mid \Psi_t\right)$ alone is not precise indicator of future y_{t+1} value movements. Moreover conditional distribution $\boldsymbol{\varepsilon}_{t+1} \mid \Psi_t$ type and parameters strongly affects conditional distribution of $y_{t+1} \mid \Psi_t$ as a linear combination of $\mathbf{x}_{t+1} \mid \Psi_t$ constituents. In such a case to make useful predictions all available information on conditional distribution of $y_{t+1} \mid \Psi_t$ should be exploited, not only on selected parameters of it. From conditional distribution of $y_{t+1} \mid \Psi_t$ we can derive quantile forecasts or asses probability of up or down movement from current y_t value. Because of complex shape of conditional distribution $y_{t+1} \mid \Psi_t$ (which may be asymmetric) and complicated relations describing its parameters such situation may occur that despite $\operatorname{sgn}\left[E\left(y_{t+1} \mid \Psi_t\right) - y_t\right]$ suggests specific direction of future movement, taking into account all available information about conditional distribution of $y_{t+1} \mid \Psi_t$ may point that opposite direction movement is more probable. In this situation expected autoregressive, reverting tendency is dominated by overdispersion and statistical arbitrage cannot be realized.

We show below $T = 1000$ length sample series simulated from VECM-CCC-GARCH ($n = 2$, $r = 1$ with 2×1 cointegration vector $\boldsymbol{\beta}$, model has no short-term dynamics) for $\mathbf{x}_t$, $\Delta\mathbf{x}_t$, $y_t = \boldsymbol{\beta}'\mathbf{x}_t$ and scatter plot for $\mathbf{x}_t = (x_{t1}, x_{t2})'$.

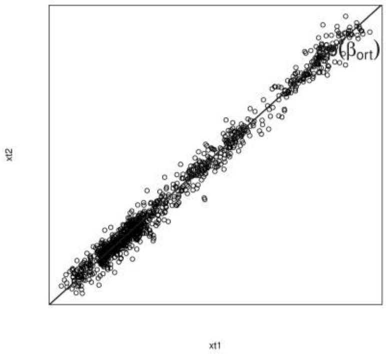

Fig. 1. Scatter plot for $(x_{t1}, x_{t2})'$ with marked attractor given by the subspace $\mathrm{sp}(\boldsymbol{\beta}_\perp)$.

One dimensional subspace spanned by $\boldsymbol{\beta}$ orthogonal complement, denoted by $\mathrm{sp}(\boldsymbol{\beta}_\perp)$ forms an attractor for considered 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 assumed model $y_t = \boldsymbol{\beta}'\mathbf{x}_t$ is driven towards 0.

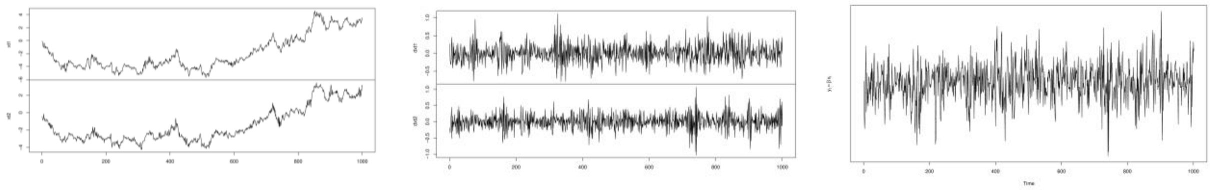

Fig. 2. Simulated time series for log-prices $\mathbf{x}_t = (x_{t1}, x_{t2})'$ (Left) and log-returns $\Delta\mathbf{x}_t = (\Delta x_{t1}, \Delta x_{t2})'$ (Middle). Simulated realization of deviations (mispricing) process $y_t = \boldsymbol{\beta}'\mathbf{x}_t$ (Right).

VECM-MGARCH may be too restrictive in its construction, because it is suggested that because of transaction costs, only higher absolute deviations from equilibrium relation are corrected by arbitrageurs. To include this fact, (Balke and Fomby, 1997) proposed extension of VECM model part called TVECM (Threshold VECM) with three regimes and one cointegrating vector, $r = 1$:

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

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

In this case, because of nonlinear dynamics, model has no representation stated by Granger Representation Theorem, so we need to extend definitions of integration and cointegration. Extended definition of $I(0)$ n-dimensional (vector) process make use of functional central limit theorem (FCLT) (more in Davidson, 1994).

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

Cointegration in this extended approach is defined without appealing to some explicit model specification (so it can embrace models with different types of short-term and error correction dynamics).

Let $\{\mathbf{x}_t\}\sim I(1)$ with respect to extended definition. Additionally we assume decomposition of invertible matrix $\tilde{\boldsymbol{\beta}} = [\boldsymbol{\beta}_\perp, \boldsymbol{\beta}]$, where $\dim(\boldsymbol{\beta}) = n\times r$, $\dim(\boldsymbol{\beta}_\perp) = n\times(n-r)$, $0<r<n$ and $\boldsymbol{\beta}'\boldsymbol{\beta}_\perp = \mathbf{0}$. We can decompose $\{\mathbf{x}_t\}$ 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 $T^{-\frac{1}{2}}\mathbf{u}_{\lfloor aT\rfloor} \xrightarrow{d} \mathbf{W}(a)\sim I(1)$ 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 transitory component (which can also be generated by nonlinear process with short-memory), additionally $\boldsymbol{\beta}$ spans r-dimensional cointegration space, on the other hand $\{\mathbf{u}_t = \boldsymbol{\beta}_\perp{}'\mathbf{x}_t\}$ is a stochastic trend component, which is "variance dominating", what means that $\{\mathbf{u}_t\}$ diverges at faster rate than $\{\mathbf{y}_t\}$.

3. Difficulties with inference on cointegration in case of heteroscedastic innovations

We briefly consider here only frequentist approach to testing cointegration under heteroscedastic innovations of specific type.

Classical Johansen cointegration rank tests (Johansen, 2006) associated with $CI(1,1)$ process VECM model with iiN innovations called maximum eigenvalue test (cointegration rank equals: $H_0 : r$ *vs.* $H_1 : r+1$) and trace test ($H_0 : r$ vs. $H_1 : n \Leftrightarrow \{\mathbf{x}_t\} \sim I(0)$) with asymptotic distributions under null hypothesis derived with use of FCLT and specified as functionals of standard Wiener process. It has been shown (Cavaliere et al., 2010) that when we attenuate assumptions about innovations process from iiN to one belonging to 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 tests for finite-length samples are regarded as quasi-likelihood ratio tests because they use likelihood function for VECM model with iiN innovations. These quasi-LR tests use asymptotic critical values what is reflected in mild to high test size distortions. In simulation study (Maki, 2013) for Johansen tests, true null hypothesis of no cointegration ($r = 0$) was more frequently rejected than assumed nominal critical level stated, under innovations with MGARCH type of conditional heteroscedasticity. To improve finite-length sample Johansen quasi-LR tests performance wild bootstrap procedure was suggested (Cavaliere et al., 2010). Wild bootstrap unlike other types of bootstrap methods (ex. VECM iid bootstrap by (Swensen, 2006)) enables to retain heteroscedasticity structure of original series. In single wild bootstrap replication QML estimated VECM model errors $\{\boldsymbol{\varepsilon}_t\}_{t=1}^{T}$ are multiplicatively distorted by univariate $iid(0,1)$ process $\{v_t\}_{t=1}^{T}$ and new series of $\Delta\mathbf{x}_t^{\ b}$ are constructed using $\Delta\mathbf{x}_t^{\ b} = \hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\beta}}'\mathbf{x}_{t-1}^{\ b} + \sum_{i=1}^{k}\hat{\boldsymbol{\Gamma}}_i\Delta\mathbf{x}_{t-i}^{\ b} + \boldsymbol{\varepsilon}_t^{\ b}$, $t = 1,\ldots,T$ where $\boldsymbol{\varepsilon}_t^{\ b} = v_t \cdot \boldsymbol{\varepsilon}_t$ with $\{v_t\}_{t=1}^{T} \sim iid(0,1)$, $\Delta\mathbf{x}_0^{\ b} = (\mathbf{x}_0, \mathbf{x}_{-1}, \ldots, \mathbf{x}_{-k+1})'$. Wild bootstrap p-value of Johansen quasi-LR test with null hypothesis of r cointegration rank, 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 a indicator function, Q_r and $Q_{r,b}$ are respectively a quasi-LR test value calculated for VECM model estimated using genuine series $\Delta\mathbf{x}_t$ and series $\Delta\mathbf{x}_t^{\ b}$ constructed in b-th replication of wild bootstrap.

Simulations (Cavaliere et al., 2008; Cavaliere et al., 2010) showed under null of no cointegration and MGARCH innovations, reduction in test size distortion for presented wild bootstrap variant in comparison to tests using asymptotic critical values, but these bootstrap tests are still associated with VECM model assuming linear error-correction and short-term dynamics. There are some propositions of cointegration tests having as an alternative hypothesis models with some specific type of nonlinear error-correction and short-term dynamics, but according to simulations (Maki, 2013) they suffer from non-acceptably large size distortions under MGARCH heteroscedastic innovations. In statistical arbitrage problem it is desired to use cointegration tests not requiring specification of model dynamics in advance. Using presented earlier in this work extended definitions of integrated and cointegrated processes Breitung suggested cointegration rank test (Breitung, 2002) which is asymptotically free of nuisance parameter of long-term covariance, which is influenced by type of short-term dynamics (linear/non-linear, number of included lags etc.), eventual conditional heteroscedasticity and parameters related to them. No prior model specification is needed to conduct Breitung cointegration test. It is a very important aspect because in statistical arbitrage problem it is not known in advance which assets have related log-prices processes. Specification of models for log-prices for many subgroups from adopted universe of assets would be problematic.

We now briefly discuss Breitung cointegration rank test construction.

Let $E_T = \sum_{t=1}^{T} \mathbf{x}_t \mathbf{'x}_t$ and $F_T = \sum_{t=1}^{T} \mathbf{X}_t \mathbf{'X}_t$ where $\mathbf{X}_t = \sum_{i=1}^{t} \mathbf{x}_i$. Breitung cointegration test uses in its construction solution of generalized eigenvalue problem: $\left|\lambda F_T - E_T\right| = 0$. For eigenvalue $\lambda_j \left(j = 1, \ldots, n \right)$ we have $\lambda_j = \frac{\mathbf{v}_j ' E_T \mathbf{v}_j}{\mathbf{v}_j ' F_T \mathbf{v}_j}$, so when $\mathbf{v}_j$ belongs to $\mathrm{sp}\left(\boldsymbol{\beta}_\perp\right)$ we have: $\mathbf{v}_j ' E_T \mathbf{v}_j = O_p\left(T^2\right)$, $\mathbf{v}_j ' F_T \mathbf{v}_j = O_p\left(T^4\right)$ and $\lambda_j = O_p\left(T^{-2}\right)$. On the other hand when $\mathbf{v}_j \in \mathrm{sp}\left(\boldsymbol{\beta}\right)$, then for $T \to \infty$: $T^2 \lambda_j \to \infty$.

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 generalized eigenproblem. Under null hypothesis test statistic has asymptotic distribution derived with use of FCLT which is a trace of a specified functional of $\left(n-r\right)$-dimensional standard Wiener process defined on $[0,1]$, free of nuisance parameter of long-term covariance. Under alternative test statistic asymptotically tends to infinity, so test has right-side critical area. According to simulations results (Maki, 2013), Breitung cointegration test using is recommended for finite length samples, when innovations are of MGARCH type and null of no cointegration (its size distortion was minimal among considered tests).

Conclusion

Cointegration between log-prices of related assets is a necessary but not sufficient condition to statistical arbitrage condition to hold. Shocks causing deviations from equilibrium relation also increase dispersion of deviations process, so autoregressive reverting tendency may be dominated by inflated conditional variance, and future movements of process can be hard to predict and even opposite to those suggested by expected value conditional on process past. Another obstacle in implementing this strategy under heteroscedastic innovations is increased chance (with respect to assumed test critical level) for most types of tests with null of no cointegration, of finding false relations of log-prices.

References

Balke, N. S., & Fomby, T. B. (1997). Threshold cointegration. *International economic review*.

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

Burgess, A. N. (2000). *A computational methodology for modelling the dynamics of statistical arbitrage* (Doctoral dissertation, University of London).

Cavaliere, G., Rahbek, A., & Robert Taylor, A. M. (2008). Testing for co-integration in vector autoregressions with non-stationary volatility. *CREATES Research Paper*.

Cavaliere, G., Rahbek, A., & Taylor, A. M. (2010). Cointegration rank testing under conditional heteroskedasticity. *Econometric Theory*, *26*(06), 1719-1760.

Chan, N. H. (2011). *Time series: applications to finance with R and S-Plus*. John Wiley & Sons.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. OUP.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. OUP.

Maki, D. (2013). The influence of heteroskedastic variances on cointegration tests: A comparison using Monte Carlo simulations. *Computational Statistics*, *28*(1), 179-198.

Swensen, A. R. (2006). Bootstrap Algorithms for Testing and Determining the Cointegration Rank in VAR Models1. *Econometrica*, *74*(6), 1699-1714.