Managerial Economics
2013, No. 14, pp. 39–60
http://dx.doi.org/10.7494/manage.2013.14.39

Henryk Gurgul*, Robert Syrek**

The structure of contemporaneous price-volume relationships in financial markets

1. Introduction

The learning of price-volume dependencies is important, because it enables to get an insight into the structure of financial markets, and into the information arrival process. In addition, one can learn how information is disseminated among market participants.

There are two competitive hypotheses: the Mixture of Distribution Hypothesis (MDH hereafter) [1, 5, 8, 26] and the Sequential Information Arrival Hypothesis [6, 13]. While MDH implies contemporaneous price-volume relationships the Sequential Information Arrival Hypothesis assumes dynamic, causal dependence price-trading volume.

Under the Mixture of Distributions Hypothesis the time series of the volatility of stock returns and trading volume are positively correlated, but the time series of stock returns and trading volume do not show correlation. Most contributions involving price-volume dependencies were based upon the Pearson linear correlation coefficient, which does not allow the testing of extreme value dependencies. Fleming and Kirby [9] found a strong correlation between innovations and trading volume and volatility in the case of 20 firms on the Major Market Index (MMI). The results suggest that trading volume can be used to obtain more precise estimates of daily volatility for cases in which high-frequency returns are unavailable. Balduzzi et al. [1] using linear regression (with trading

* AGH University of Science and Technology in Cracow, Department of Applications of Mathematics in Economics, e-mail: henryk.gurgul@gmail.com. Financial support for this paper from the National Science Centre of Poland (Research Grant DEC-2012/05/B/HS4/00810) is gratefully acknowledged.

** Jagiellonian University in Cracow, Institute of Economics and Management, e-mail: robert.syrek@uj.edu.pl

volume as a dependent variable) arrived at a low correlation between extremely low (below –4.09%) stock returns and trading volume for the American Index. Marsh and Wagner [15] tested tail relationships (the indexes under study were the AEX, CAC, DAX, HSI, FTSE, S&P500 and TPX) using extreme value theory. The authors found a lower degree of dependence in the left tail than in the right tail in the pair stock returns-trading volume.

In one of more recent studies Gurgul et al. [12] modeled the dependence structure of log-volume and volatility (calculated as absolute values of stock returns) for eight stocks from the DAX. The results indicate a significant dependence between high values of variables and a lack of dependence for low values.

Rossi and de Magistris [24] using mixtures of copulas and survival copulas (Gumbel and Clayton) found that volatility and volume are more dependent for high values than for low. The volatility was computed using high-frequency data and realized volatility estimators. Ning and Wirjanto [18] using Archimedean copulas tested the degree of dependence of stock returns and trading volume for some Asian indexes. The presented results indicate that there is no dependence between low stock returns and high (low) trading volume.

A special kind of dependence is known as long-memory. (Robinson and Yajima [3] , Phillips and Shimotsu [20, 21, 22], Shimotsu [25]). If a time series possesses long memory, there is a persistent temporal dependence between observations even considerably separated in time. The long memory property of volatility has been widely documented in empirical research. This topic was discussed in Bollerslev and Mikkelsen [4] and Ding and al. [7], among others. On the other hand, Lobato and Velasco [14], Bollerslev and Jubinski [3], Fleming and Kirby [9], Rossi and de Magistris [19] found that trading volume also exhibits long-run dependence (long memory). The interesting question is the link between long memory in volatility and in trading volume.

The central question of our paper is the examination of dependence stuctures of stock returns, volatility and trading volumes of companies included in CAC40 and FTSE100. Moreover, we aim to test the MDH hypothesis in version with long memory. We will check the equality of the long memory parameters of volatility and trading volume and fractional cointegration of these time series.

In particular we will examine the existence of essential dependence between high volatility and high trading volume. The important goal of this study is the choice of proper copulas necessary to capture contemporaneous dependence structures of returns and trading volume. In addition, we will also compare the dependence structure of times series under study based on companies included in CAC40 and FTSE100.

The structure of the paper is as follows. The methodology and main notions applied are outlined in the following section. Third section is concerned with a description of the dataset. Empirical results and their discussion are provided in fourth section. Finally, in the last section we summarize major conclusions and suggest directions for future research.

2. Methodology

2.1. Long memory

The autocorrelation function (ACF) of time series with long memory tails off hyperbolically. The short-memory property is easy to detect by the low order correlation structure of a series. This type of time series is characterized by exponentially declining autocorrelations and, in the spectral domain, demonstrates high-frequency distribution. The standard ARMA-processes do not show long memory. They can only exhibit short run (high-frequency) properties.

The presence of long memory in financial data is a source of both theoretical and empirical problems. The long memory property arises from nonlinearities in economic data. The well-known martingale models of stock prices cannot follow from arbitrage, because new information cannot be entirely arbitraged away. A second problem caused by long memory is pricing derivative securities with the martingale method. This method is usually false if the accompanying stochastic (continuous) processes exhibit long memory. The process X_t has a degree of fractional integration d (we write $I(d)$), when:

$$(1-L)^d X_t = u_t, \tag{1}$$

where L is a lag operator ($LX_t = X_{t-1}$) and u_t is a process with a short memory. The expression $1(-L)^d$ is presented in the form of the infinite series:

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(d)\Gamma(k+1)} L^k,$$

where $\Gamma(x)$ is the Gamma function. The process *ARMA* $(p.q)$ is defined as:

$$\Phi(L)(u_t-\mu) = \Theta(L)\varepsilon_t, \tag{2}$$

where $\Theta(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$ and $\Theta(L) = 1 + \sum_{j=1}^{q} \theta_j z^j$ are lag polynomials of degree p and q, respecively. The process is stationary and invertible if the roots of $\Phi(z)$ and $\Theta(L)$ lie outside the unit circle. If u_t is described by (2.2), and ε_t is white noise then the process is the Autoregressive Fractionally Integrated Moving Average process *ARFIMA*(p, d, q).

If the parameter $0 < |d| < 0{,}5$ then the process is stationary and invertible and the autocorrelation function exhibits hyperbolic decay, because for the lag k it is proportional to $\frac{\Gamma(1-d)}{\Gamma(d)}k^{2d-1}$ when $k \to \infty$. If $d \in (0; 0{,}5)$, we say that the process has a long memory and if $d \in (-0{,}5; 0)$ the process is antipersistent and has intermediate memory. For $d \in [0{,}5; 1]$ the variance of X_t is infinite, so the process is covariance nonstationary but mean-reverting.

There are many different estimators of long memory parameter d (Phillips and Shimotsu [21]). We use the exact local Whittle estimator (Phillips and Shimotsu [20, 22], Shimotsu [25]). Following (2.1) we get:

$$X_t = (1-L)^{-d} u_t = \sum_{k=0}^{t-1} \frac{\Gamma(d+k)}{\Gamma(d)\mathrm{k}!} u_{t-k}, \quad t = 0, \pm 1, \pm 2 \ldots$$

Discrete Fourier transformations and periodogram of α_t are defined as:

$$w_a(\lambda_j) = (2\pi n)^{-1/2} \sum_{t=1}^{n} a_t e^{it\lambda_j}, \text{ where } \lambda_j = \frac{2\pi j}{n}, j = 1, \ldots, n,$$

$$I_a(\lambda_j) = \left| w_a(\lambda_j) \right|^2.$$

Supposing that process X_t is covariance stationary and spectral density function $f(\lambda)$ fulfills the condition $f(\lambda) \sim G\lambda^{-2d}$, if $\lambda \to 0_+$, Phillips and Shimotsu [20] minimize the function:

$$Q_m(G,d) = \frac{1}{m} \sum_{j=1}^{m} \left[log\left(G\lambda_j^{-2d}\right) + \frac{1}{G} I_{\Delta^d x}(\lambda_j) \right]$$

The ELW estimator of long memory parameter d is then:

$$\hat{d}_{ELW} = \arg \min_{d \in [\Delta_1, \Delta_2]} R(d),$$

and

$$R(d) = log\hat{G}(d) - 2d \frac{1}{m} \sum_{j=1}^{m} log\lambda_j, \hat{G}(d) = \frac{1}{m} \sum_{j=1}^{m} I_{\Delta^d x}(\lambda_j).$$

If d_0 is value of the true parameter of long memory parameter d then if $\Delta_2 - \Delta_1 \leq \frac{9}{2}$ and the assumed m is such that $\frac{m}{n} + \frac{1}{m} \to 0$, if $n \to \infty$, then the ELW estimator is consistent and it holds true that:

$$\sqrt{m}(\hat{d}_{ELW} - d_0) \xrightarrow{d} N\left(0, \frac{1}{4}\right).$$

2.2. Fractional cointegration

Stationarity is a crucial precondition for standard linear Granger causality tests. Nonstationarity of the time series under study may lead to false conclusions by a traditional linear causality test. This phenomenon has been investigated in previous empirical (Granger and Newbold [11]) and theoretical (Phillips [19]) deliberations which led to a cointegration analysis.

A cointegration analysis (based on the estimation of a VEC model) may be performed for variables which are integrated in the same order. As shown by Granger the existence of cointegration implies long run Granger causality in at least one direction (Granger [11]). To establish the direction of this causal link one should estimate a suitable VEC model and check (using a *t*–test) the statistical significance of the error correction terms. Testing the joint significance (using an *F*–test) of lagged differences provides a basis for short run causality investigations.

The classical definition of cointegration can be generalized as for any d and d_e two $I(d)$ processes are fractionally cointegrated, if there exists a linear combination of these processes that is $I(d_e)$ with $d_e < d$. In this case there exists long-run dependence and a common stochastic trend. Assume that $z_t = (x_t, y_t)$ with $x_t \in I(d)$ and $y_t \in I(d)$. If there exists $\beta \neq 0$ such that there is the linear combination $y_t - \beta x_t \in I(d_e)$, where $0 \leq d_e < d$, then x_t and y_t are fractionally cointegrated. We write $z_t \in CI(d, b)$, for $b = d - d_e$. Robinson i Yajima [23] consider the case of stationary variables, whereas Nielsen and Shimotsu [17] analyse the case of covariance nonstationary variables too. The model under consideration is given by (Shimotsu [25]):

$$\begin{cases} (1-L)^{d_e}(y_t - \beta x_t) = u_{1t} \\ (1-L)^{d} x_t = u_{2t} \end{cases} \tag{3}$$

where $u_t = (u_{1t}\, u_{2t})' = C(L)\varepsilon_t$ is a bidimensional stationary vector with spectral density $f_{\mathrm{u}}(\lambda)$. In matrix notations (2.3) has the form:

$$Bz_t = \begin{pmatrix} (1-L)^{-d_e} & 0 \\ 0 & (1-L)^{-d} \end{pmatrix} u_t,\ B = \begin{pmatrix} 1 & -\beta \\ 0 & 1 \end{pmatrix},\ z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix}.$$

The rank of the matrix $C(1)$ determines whether the processes y_t and x_t are cointegrated. Denoting as r the number of cointegration vectors, the rank of $C(1)$ is equal to $2 - r \leq 2$. If the variables are cointegrated, then $C(1)$ does not have full rank.

The fractional cointegration can be tested as follows. Firstly using Whittle estimators long memory parameters are estimated, and then a test of their equality is performed. Let d_* be the common value of the long memory parameters of series x_t and y_t (with parameters d_1 and d_2 , respectively). When testing:

$$H_0 : d_i = d_*, i = 1,2,$$

test statistics of Robinson and Yajima [23] has the form:

$$\hat{T}_0 = m(S\hat{d})'\left(S\frac{1}{4}\hat{D}^{-1}\left(\hat{G}\odot\hat{G}\right)\hat{D}^{-1}S' + h(n)^2\right)^{-1}(S\hat{d}),$$

where $S = (1\,-1)'$, $h(n)$ is the function which is convergent to 0, $D = diag\ (G_{11}, G_{22})$, whereas $\hat{G}$ is expressed as:

$$\hat{G}\left(\hat{d}\right) = \frac{1}{m}\sum_{j=1}^{m} Re[I_{\Delta(\mathrm{L};d_*)[x,y]}(\lambda_j)],$$

where $I_{\Delta(\mathrm{L};d_*)[x,y]}$ is the periodogram of $\left((1-L)^{d_1} x_t, (1-L)^{d_2} y_t\right)'$. If the variables under study are not cointegrated (cointegration rank $r = 0$ then $\hat{T}_0 \to^d \chi_1^2$. Otherwise $\hat{T}_0 \to 0$, which means that $r = 1$. If $H_0 : d_i = d_*, i = 1,2$ cannot be rejected then one can estimate the cointegration rank using the eigenvalues of matrix $\hat{G}$. If $\hat{\delta}_i$ is $i-th$ eigenvalue, then the rank of cointegration is equal to

$$\hat{r} = \underset{u=0,1}{\operatorname{argmin}}\, L(u),$$

where

$$L(u) = v(n)(2-u) - \sum_{i=1}^{2-u}\hat{\delta}_i,$$

and $v(n)$ is a function with $(n) + \frac{1}{\sqrt{m_1}v(n)} \to 0$, for $n \to \infty$. The value $\hat{G}(d_4)$ is estimated as:

$$\hat{G}(d_*) = \frac{1}{m_1}\sum_{j=1}^{m_1} Re[I_{\Delta(\mathrm{L};d_*)[x,y]}(\lambda_j),$$

where $I_{\Delta(\mathrm{L};d_*)[x,y]}$ is the periodogram of $\left((1-L)^{d_*} x_t, (1-L)^{d_*} y_t\right)'$ whereas m_1 is the function of n. The value of d_* is unknown, so it is computed as the mean of the estimated long memory parameter values of x_t and y_t. Finally $\hat{G}\left(\overline{d_*}\right)$ is computed.

2.3. Dependence between volatility and trading volume

In this subsection the methods of the dependence structure analysis of volatility and trading volume is described. Using copulas we can model the degree of dependence in the tails. i.e. for extreme values.

Having estimated long memory parameters to filter the time series we can use FIVAR models (Rossi and de Magistris [24]). We should transform the series using formulas:

$$(1-L)^{d_{R_t^2}} R_t^2 = \widetilde{R_t^2},$$

$$(1-L)^{d_{\ln V_t}} \ln V_t = \widetilde{\ln V_t}.$$

As a result we obtain stationary time series I(0). Then, we apply a *VAR* (*k*) model to capture linear dependencies. This model for vector $\boldsymbol{P_t} = (X_t\, Y_t)'$ can be described as :

$$\boldsymbol{P_t} = \Phi_0 + \sum_{i=1}^{k} \Phi_i \boldsymbol{P_{t-i}} + \boldsymbol{\varepsilon_t},$$

where Φ_0 is the vector of intercepts. $\Phi_i = \begin{pmatrix} \phi_{11.i} & \phi_{12.i} \\ \phi_{21.i} & \phi_{22.i} \end{pmatrix}$ is the matrix of parameters (for $i = 1 \ldots k$) and ε_t is the vector of error terms. Optimal lags k are chosen using information criteria and likelihood ratio tests. For vector $\boldsymbol{P_t} = \left(\widetilde{R_t^2}\, \widetilde{\ln V_t}\right)'$ in most cases $k <= 3$. We estimate the variance-covariance matrix of parameters with heteroscedasticity correction. In most cases this correction is enough to get homoscedastic errors. If not, ARCH-type models are used. We standardize the residuals and fit different distribution functions: NIG (abbreviated from normal inverse Gaussian), hyperbolic, *t* location-scale.
Probability density functions are given by:

– NIG:

$$f_{NIG}\left(x;\alpha;\beta;\delta;\mu\right) = \frac{\alpha\delta}{\pi}\exp\left(\delta\gamma + \beta\left(x-\mu\right)\right)\frac{K_1\left(\alpha\sqrt{\delta^2 + \left(x-\mu\right)^2}\right)}{\sqrt{\delta^2 + \left(x-\mu\right)^2}}.$$

where $x \in \mathbf{R}, \alpha \in \left(0.\infty\right), \beta \in \left(-\alpha.\alpha\right), \delta \in \left(0.\infty\right), \gamma = \sqrt{\alpha^2 - \beta^2}$ and $K_1\left(\cdot\right)$ is a modified Bessel function of the third kind with an index one of the form:

$$K_1\left(z\right) = \frac{1}{2}\int_0^{\infty} exp\left(-\frac{1}{2}\left(z\left(t + t^{-1}\right)\right)\right)dt;$$

– hyperbolic:

$$f_{HYP}\left(x;\alpha;\beta;\delta;\mu\right) = \frac{\gamma}{2\alpha\delta K_1\left(\delta\gamma\right)}\exp\left(-\alpha\sqrt{\delta^2 + \left(x-\mu\right)^2}\right) + \beta\left(x-\mu\right);$$

$$f_{Skal-t}(x;\mu;\sigma;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{\nu+\left(\frac{x-\mu}{\sigma}\right)^2}{\nu}\right)^{-(\nu+1)/2}.$$

All of the distribution functions presented above are special cases of generalized hyperbolic distributions.

2.4. Copulas

Copulas reflect the dependence structures among financial variables. We use in empirical part Gaussian copula, Archimedean (Clayton and Gumbel) copulas, survival copulas and their convex combination (Nelsen, 1999).

The Gaussian copula (or normal copula) is given by:

$$C_\rho^{Ga}(u_1,u_2) = \Phi_\rho\left(\Phi^{-1}(u_1)\Phi^{-1}(u_2)\right) =$$

$$= \int_{-\infty}^{\Phi^{-1}(u_1)}\int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\rho^2)^{1/2}} exp\left(\frac{-\left(s_1^2-2\rho s_1 s_2+s_2^2\right)}{2(1-\rho^2)}\right) ds_1 ds_2,$$

where Φ_ρ is bivariate normal distribution with correlation coefficient $|\rho|< 1$ and Φ denotes standard univariate normal distribution function.

The Clayton copula is given by:

$$C(u_1,u_2;\theta) = \max(\left[u_1^{-\theta}+u_2^{-\theta}-1\right]^{-\frac{1}{\theta}},0).$$

with $\theta \in [-1.\infty)\setminus\{0\}$. If parameter θ is positive then

$$C(u_1,u_2;\theta) = (u_1^{-\theta}+u_2^{-\theta}-1)^{-\frac{1}{\theta}}.$$

The Gumbel copula is given by:

$$C(u_1,u_2;\theta) = \exp(-\left[(-lnu_1)^\theta+(-lnu_2)^\theta\right]^{\frac{1}{\theta}}),$$

for $\theta \in [-1.\infty)$.

The Gumbel and survival Clayton copulas describe asymptotic dependence in the right tail, and Clayton and survival Gumbel in the left tail. To model the dependence in both tails simultaneously one can use mixtures of copulas.

We consider the following copulas:

1. $\omega C_{Gum} + (1-\omega) C_{sGum}$;
2. $\omega C_{Gum} + (1-\omega) C_{Cl}$;
3. $\omega C_{sCl} + (1-\omega) C_{sGum}$;
4. $\omega C_{sCl} + (1-\omega) C_{Cl}$;
5. $\omega C_{Gum} + (1-\omega) C_{Gauss}$;
6. $\omega C_{sCl} + (1-\omega) C_{Gauss}$;

and one-parameter copulas:

7. ωC_{sCl}
8. ωC_{Gum}

The copulas that fit the best are chosen using information criterion. The correctness of the copula specification are validated by an Anderson-Darling test applied to the first derivative of copulas: $C(u|v) = \frac{dC}{du}$ and $C(v|u) = \frac{dC}{dv}$.

The classical Archimedean copulas (and survival copulas) defined above (volatility-trading volume pair) can be applied only to modeling dependence in the top-right corner (high returns-high volume). To model relationships in the top-left corner we can use rotated (anticlockwise) copulas by 90° degrees (Gumbel copula) and 270° (Clayton copula). For any copula C it holds true that:

$$C^{(90)}(u_1,u_2) = u_2 - C(1-u_1,u_2),$$

$$C^{(180)}(u_1,u_2) = u_1 + u_2 - 1 + C(1-u_1,1-u_2),$$

$$C^{(270)}(u_1,u_2) = u_1 - C(u_1,1-u_2),$$

The copula $C^{(180)}$ is of course the survival copula for C. The domain of copula parameters ($C^{(90)}$ and $C^{(270)}$) are symmetrical in respect to zero so the parameters are negative. As formerly, mixtures of copulas can be used to model dependence in both top corners simultaneously.

- $\omega C_{Gum} + (1-\omega) C_{Gum}^{(90)}$;
- $\omega C_{Gum} + (1-\omega) C_{Cl}^{(270)}$;
- $\omega C_{Cl}^{(180)} + (1-\omega) C_{Cl}^{(270)}$;
- $\omega C_{Cl}^{(180)} + (1-\omega) C_{Gum}^{(90)}$.

Using the reviewed methods we will check in different aspects links between returns and trading volume. In the next section we will show the dataset.

3. Data description

We consider the prices and trading volumes of stocks from the French (CAC40) and the English (FTSE100) indexes from 1 October 2002 to 1 October 2012. The dataset comes from Thomson Reuters data services and covers a period of 2610 trading days. Throughout the paper stock returns were approximated by log-returns.

3.1. Descriptive statistics

Using daily prices at close we computed logarithmic stock returns and multiplied them by 100. The series of trading volumes are mostly leptokurtic and positively skeweed so we apply a logarithmic transformation. As a result, the returned series are close to normal. The Tables 1 and 2 present the descriptive statistics of the log-returns, volatilities (square of log- returns) and log-volumes.

Table 1

Descriptive statistics of companies listed on CAC40

log-returns				
statistics	**mean**	**standard dev.**	**skewness**	**kurtosis**
minimum	–0.050	1.434	–2.039	5.418
1st quartile	–0.014	1.862	–0.122	7.285
median	0.012	2.183	0.084	8.750
3rd quartile	0.032	2.574	0.267	10.527
maximum	0.074	3.730	0.970	53.052
log–volume				
statistics	**mean**	**standard dev.**	**skewness**	**kurtosis**
minimum	5.150	0.438	–1.003	2.737
1st quartile	6.746	0.486	–0.317	4.077
median	7.361	0.531	–0.188	4.391
3rd quartile	8.111	0.618	0.058	4.830
maximum	9.763	1.155	0.356	7.642
volatility				
statistics	**mean**	**standard dev.**	**skewness**	**kurtosis**
minimum	2.057	5.438	5.760	51.531
1st quartile	3.467	9.342	7.292	78.266
median	4.763	12.646	9.168	127.220
3rd quartile	6.624	19.192	12.896	270.836
maximum	13.910	100.382	42.043	1991.226

Source: own elaboration based on Reuters data basis

Table 2

Descriptive statistics of companies listed on FTSE100

log-returns				
statistics	**mean**	**standard dev.**	**skewness**	**kurtosis**
minimum	–0.126	0.990	–10.355	3.620
1st quartile	0.012	1.699	–0.282	7.524
median	0.032	1.963	–0.077	9.248
3rd quartile	0.057	2.516	0.081	13.482
maximum	0.113	4.179	1.409	316.662
log-volume				
statistics	**mean**	**standard dev.**	**skewness**	**kurtosis**
minimum	5.100	0.489	–1.877	2.868
1st quartile	7.434	0.604	–0.353	3.643
median	8.162	0.667	–0.192	4.121
3rd quartile	9.144	0.745	–0.048	4.830
maximum	12.124	1.690	0.612	10.228
volatility				
statistics	**mean**	**standard dev.**	**skewness**	**kurtosis**
minimum	0.981	4.139	3.185	15.753
1st quartile	2.885	8.240	8.002	97.386
median	3.857	11.588	10.329	156.916
3rd quartile	6.330	19.931	15.339	358.252
maximum	17.456	243.561	50.070	2539.289

Source: own elaboration based on Reuters data basis

For all stocks under consideration we observe significant skewness and excess kurtosis in stock returns. The null hypothesis about normality by the Jarque-Bera test is rejected in all cases. Some of the log-volume series have a kurtosis close to 3, but the non-zero skewness causes a departure from normality in the series. The null hypothesis about lack of autocorrelation by the Ljung-Box test is also rejected. Using regression we may remove, if necessary, any deterministic trend from the series of log-volumes to achieve trend-stationary time series. Additionally, we use dummy variables in order to describe calendar effects i.e. the effect of the month in the year and the day in the week in the log-volume series. The time series of volatility are far from normal because of high values of the kurtosis and skewness (positive in all cases).

4. Empirical results

4.1. Results of long memory and fractional cointegration estimation

Based upon the methodology presented above we computed the long memory parameters of the time series (Robinson and Yajima [23], Phillips and Shimotsu [20, 21, 22], Shimotsu [25]). The long memory parameters of return volatility and log-volume are denoted by $d_{R_t^2}$ and $d_{\ln V_t}$, respectively. To test the equality of long memory parameters we use (Robinson and Yajima [23]):

$$h_1(n) = 1/\ln n,$$
$$h_2(n) = 1/\ln^2 n,$$
$$m = n^{0,6}.$$

In the Tables 3 and 4 we present the results of the estimation of long memory parameters:

Table 3

Long memory parameters

	CAC40		FTSE100	
statistics	$d_{R_t^2}$	$d_{\ln V_t}$	$d_{R_t^2}$	$d_{\ln V_t}$
minimum	0.189	0.154	-0.054	-0.005
1st quartile	0.354	0.262	0.280	0.174
median	0.417	0.300	0.390	0.244
3rd quartile	0.456	0.348	0.488	0.285
maximum	0.679	0.495	0.717	0.427

Source: own elaboration based on Reuters data basis

All parameters of long memory are significant for French stocks. In eight cases the long memory parameters of R_t^2 are less than of lnV_t. The long memory parameters of R_t^2 are greater than 0.5 in seven cases. This indicates that the time series are covariance non-stationary. Taking into account that critical values are $\chi_1^2 = 2.706$, $\chi_1^2 = 3.841$, $\chi_1^2 = 6.635$, at significance levels of 10%, 5% and 1%, respectively, in twelve cases there is no reason to reject the null hypothesis of the equality of estimated long memory parameters.

In the case of English stocks $d_{R_t^2} > d_{\ln V_t}$ for 69 stocks. 96 long memory parameters of volatility are significant (at 0.1 significance level). The same conclusion is valid for 83 parameters for log-volumes. Some of the parameters are negative and close to zero. There is no reason to reject the null that they equal to zero. The null hypothesis of parameter equality is rejected for about 70% of stocks. Based upon the results above we analyzed the problem of the fractional cointegration of volatility and trading volume. We estimated the eigenvalues δ_1 and δ_2 (multiplied by 10000) of matrix $\hat{G}$ and computed the values of function $L(u)$ for $m_1 = n^{0,55}$ and $v(n) = m_1^{-0,45}$. In the tables below we present the results of the estimation of long memory parameters in detail and the fractional cointegration tests.

The descriptions of the columns of Tables 4 and 5 below refers to notations described in the section Methodology (Fractional Cointegration).

Table 4

Fractional cointegration (CAC40)

Company	$d_{R_t^2}$	$d_{\ln V_t}$	$T_0(b_1)$	$T_0(b_2)$	δ_1	δ_2	$L(0)$	$L(1)$
ACCOR	0.425	0.350	0.953	1.198	5.243	0.020	–1.713	–1.267
BNP PARIBAS	0.319	0.288	0.130	0.168	30.514	0.030	–1.713	–1.394
CARREFOUR	0.351	0.285	0.934	1.194	5.341	0.030	–1.713	–1.340
CREDIT AGRICOLE	0.373	0.320	0.308	0.386	22.102	0.036	–1.713	–1.251
EADS	0.305	0.375	0.816	1.016	25.224	0.027	–1.713	–1.208
ESSILOR INTL.	0.354	0.337	0.173	0.217	2.095	0.023	–1.713	–1.247
SAFRAN	0.353	0.386	0.363	0.449	5.975	0.029	–1.713	–1.171
SANOFI	0.299	0.348	0.443	0.563	4.471	0.021	–1.713	–1.319
SOCIETE GENERALE	0.357	0.353	0.008	0.010	32.197	0.026	–1.713	–1.348
SOLVAY	0.358	0.331	0.029	0.037	2.798	0.028	–1.713	–1.303
TECHNIP	0.448	0.365	1.535	1.941	11.138	0.026	–1.713	–1.295
VEOLIA ENVIRONNEMENT	0.274	0.306	0.280	0.355	30.874	0.026	–1.713	–1.307

Source: own elaboration based on Reuters data basis

The estimated rank of cointegration is equal to 0 for all stocks under consideration. Despite the equality of long memory parameters fractional cointegration does not exist. The same is observed when using $v(n) = m_1^{-0,35}$ and $v(n) = m_1^{-0,25}$. It is worth mentioning that for parameters $m_1 = n^{0,55}$ and $m_1 = n^{0,45}$ the conclusions are analogous.

Table 5

Fractional cointegration (FTSE100)

Company	$d_{R_t^2}$	$d_{\ln V_t}$	$T_0(b_1)$	$T_0(b_2)$	δ_1	δ_2	$L(0)$	$L(1)$
ABERDEEN ASSET MAN.	0.403	0.411	0.035	0.043	72.322	0.072	–1.713	–1.165
AGGREKO	0.348	0.304	0.882	1.096	9.377	0.058	–1.713	–1.200
ASTRAZENECA	0.252	0.250	0.002	0.002	4.978	0.028	–1.713	–1.350
BABCOCK INTL.	0.210	0.152	0.453	0.557	11.075	0.221	–1.713	–1.116
BAE SYSTEMS	0.172	0.109	0.896	1.122	44.409	0.072	–1.713	–1.250
BARCLAYS	0.237	0.231	0.001	0.001	557.495	0.056	–1.713	–1.148
BRITISH SKY BCAST. GROUP	0.262	0.203	0.676	0.842	12.532	0.078	–1.713	–1.211
BURBERRY GROUP	0.326	0.313	0.020	0.025	14.540	0.068	–1.713	–1.239
CAPITAL SHOPCTS. GROUP	0.426	0.320	1.894	2.329	3.703	0.037	–1.713	–1.116
COMPASS GROUP	0.192	0.198	0.006	0.008	30.611	0.082	–1.713	–1.199
CRODA INTERNATIONAL	0.395	0.303	1.210	1.482	2.475	0.083	–1.713	–1.071
G4S	0.154	0.113	0.255	0.316	34.623	0.150	–1.713	–1.181
GLAXOSMITHKLINE	0.261	0.244	0.109	0.137	2.425	0.038	–1.713	–1.271
HARGREAVES LANSDOWN	0.277	0.305	0.057	0.073	9.415	0.078	–1.665	–1.263
INTL.CONS.AIRL.GP.(CDI)	0.277	0.123	1.064	1.376	10.683	0.037	–1.554	–1.118
LLOYDS BANKING GROUP	0.340	0.266	0.948	1.160	323.025	0.061	–1.713	–1.047
MORRISON(WM)SPMKTS.	0.237	0.264	0.054	0.068	6.362	0.053	–1.713	–1.257
NATIONAL GRID	0.218	0.267	0.505	0.626	8.365	0.049	–1.713	–1.184
PENNON GROUP	0.321	0.261	0.688	0.853	2.148	0.061	–1.713	–1.180
RANDGOLD RESOURCES	0.438	0.427	0.065	0.079	8.820	0.042	–1.708	–1.019
RIO TINTO	0.321	0.269	0.495	0.624	119.727	0.040	–1.713	–1.288
SAINSBURY (J)	0.202	0.261	0.769	0.963	17.965	0.060	–1.713	–1.249
SCHRODERS	0.284	0.342	0.655	0.796	59.326	0.044	–1.713	–0.897
SEVERN TRENT	0.221	0.173	0.369	0.461	5.346	0.068	–1.713	–1.233
SHIRE	0.207	0.241	0.187	0.239	8.458	0.045	–1.713	–1.330
SMITHS GROUP	0.314	0.285	0.143	0.177	5.677	0.051	–1.713	–1.167
SSE	0.317	0.218	1.752	2.153	3.693	0.055	–1.713	–1.105
TESCO	0.278	0.222	0.643	0.807	4.407	0.044	–1.713	–1.262
TULLOW OIL	0.391	0.343	0.376	0.467	16.349	0.053	–1.713	–1.197
UNITED UTILITIES GROUP	0.218	0.269	0.554	0.683	5.164	0.044	–1.713	–1.130
WOLSELEY	0.244	0.227	0.030	0.037	71.620	0.069	–1.713	–1.242

Source: own elaboration based on Reuters data basis

4.2. Results of estimation of dependence between volatility and trading volume

The best fitted distributions are chosen using goodness of fit tests and information criteria. In most cases the distributions that fit best are NIG and *t*-location-scale distributions (hyperbolic distribution was fitted for only a few log-volume series of English stocks). Next, using selected distributions we transform the series to get uniformly distributed variables (comp. Fig. 1).

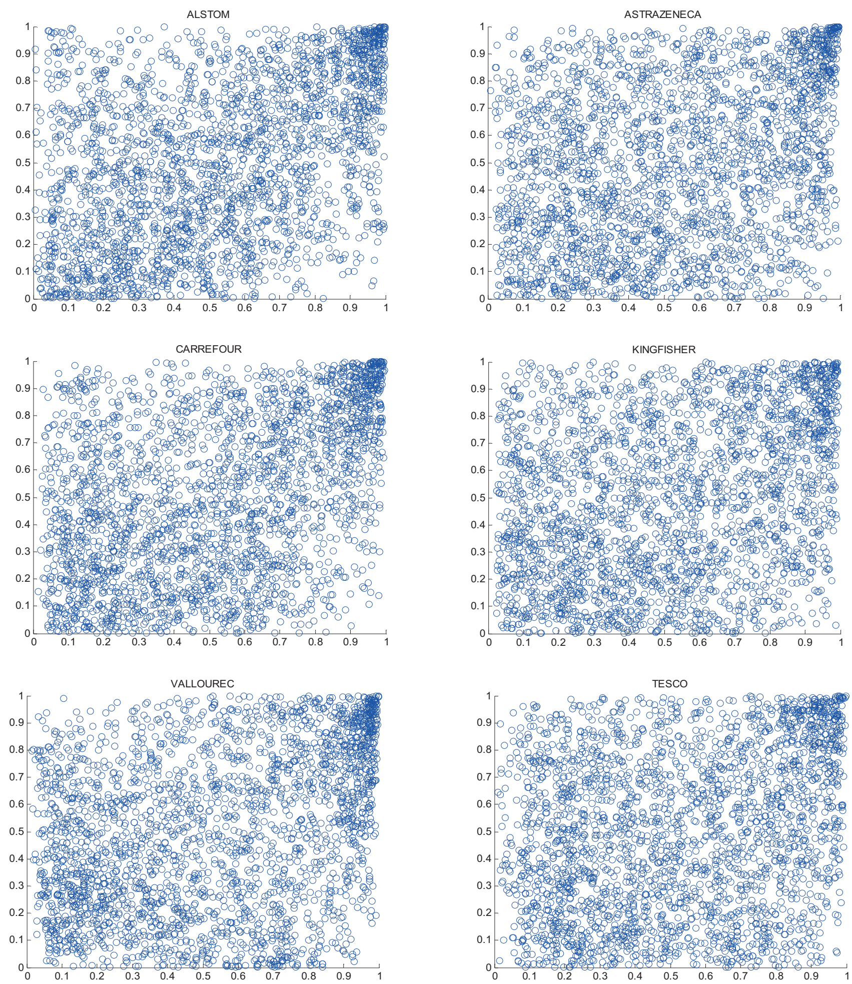

Figure 1. Dependence structure of volatility and trading volume

Source: own elaboration based on Reuters data basis

Because of the large number of companies under investigation, the figure below presents only chosen (but typical) examples of dependence structures of volatility and trading volume that are modeled using copulas. The left column contains examples of stocks from the CAC40 (ALSTOM, CARREFOUR, VINCI), the right from the FTSE100 (ASTRAZENECA, KINGFISHER, TESCO).

There are concentrations of points in the bottom-left and top-right corners i.e. extremely low and extremely high values of volatility and trading volume occur together. To describe these patterns we apply a Gaussian copula, Archimedean copulas, survival copulas and their convex combination describe above.

The Tables 6 and 7 contain the results of the estimation and dependence measures. α_1 and α_2 are the parameters of copulas used in mixtures, first and second, respectively. We compute the Kendall correlation coefficient τ using convex combinations of copulas. Tail dependence coefficients, denoted by λ_U (upper) and λ_L (lower) are scaled with a mixture parameter ω. The symbols of copulas used refers to these from section Methodology (Copulas).

Table 6

Estimation results of dependence for pair $\widetilde{R_t^2} - \widetilde{\ln V_t}$

CAC40							
company	**copula**	α_1	α_2	ω	τ	λ_U	λ_L
ALSTOM	5	1.31	0.66	0.73	0.30	0.22	0.00
CARREFOUR	6	0.73	1.50	0.96	0.27	0.37	0.02
VINCI	7	0.60	—	—	0.23	0.31	0.00
FTSE100							
company	**copula**	α_1	α_2	ω	τ	λ_U	λ_L
ASTRAZENECA	7	0.53	—	—	0.21	0.27	0.00
KINGFISHER	5	0.50	0.26	0.80	0.19	0.20	0.00
TESCO	7	0.43	—	—	0.18	0.20	0.00

Source: own elaboration based on Reuters data basis

For the most part the mixture $\omega C_{sCl} + (1-\omega) C_{Gauss}$ fits the data best for stocks traded on the CAC40. In some mixtures the estimated parameters were on a boundary, so these copulas were simplified and one-parameter copulas were used instead. On the whole, for English stocks the survival Clayton copula best fits the dataset. In the table below we present the statistics of the dependence measure of all stocks under consideration.

Table 7

Dependence measures for pair $\widetilde{R_t^2} - \widetilde{\ln V_t}$ (CAC40)

CAC40			
statistics	τ	λ_U	λ_L
minimum	0.16	0.12	0.00
1st quartile	0.21	0.22	0.00
median	0.23	0.27	0.00
3rd quartile	0.25	0.31	0.00
maximum	0.30	0.38	0.08
FTSE100			
statistics	τ	λ_U	λ_L
minimum	0.08	0.01	0.00
1st quartile	0.12	0.08	0.00
median	0.15	0.12	0.00
3rd quartile	0.17	0.16	0.00
maximum	0.21	0.27	0.04

Source: own elaboration based on Reuters data basis

The dependence in the right tail (for extremely high values) is stronger than in the left tail (extremely low values). This is because of the high values of mixture parameter omega. So dependence in the right tail is dominant. The conclusions drawn for English stocks are analogous. Dependence in the right tail is stronger.

4.3. Analysis of dependence between stock returns and trading volume

We use VAR models applied to stock returns r_t and trading volumes $\widetilde{logV_t}$ (long memory was removed from the series). To describe the heteroscedasticity observed we use a GARCH type model. As in the previous section we fitted some distributions for the residuals of the VAR models. Additionally, we considered GED and skewed *t* distributions. For the residuals of the equation for stock returns GED and skewed *t* distributions were generally. As with the results of the VAR models, for the pair volatility-volume, generally NIG distributions and the *t*-location scale were fitted for trading volumes.

Figure 2 presents typical examples of dependence structures of stock returns and trading volumes.

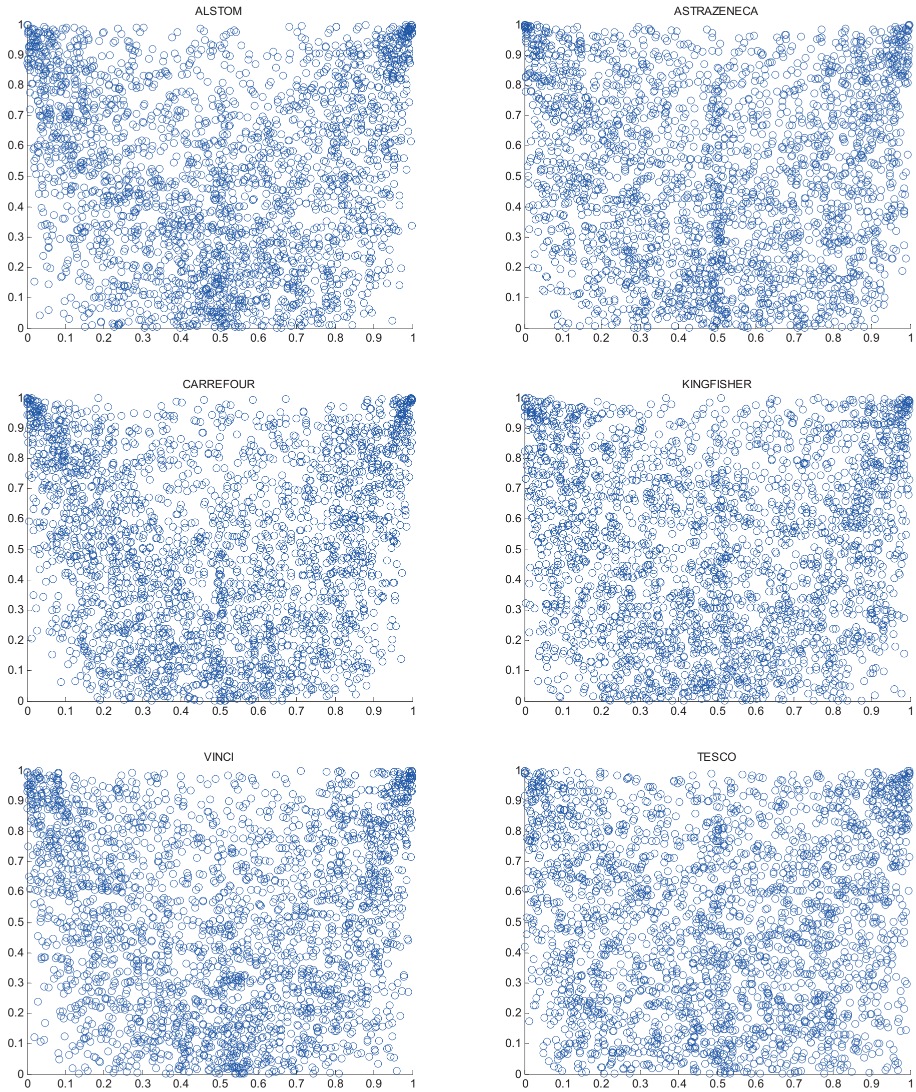

Figure 2. Dependence structure of stock returns and trading volume

Source: own elaboration based on Reuters data basis

There is a clustering of points in the top corners, which means that extremely a high trading volume is interrelated with high stock returns (positive and negative). The concentration of points for $u_1 \approx 0.5$ and $u_2 < 0.5$ is a sign of low trading volume linked with low volatility (stock returns close to zero).

The computation results corroborate the observation made above. We computed Kendall correlation coefficients for the whole sample and in all corners (for quantiles 0.01 and 0.99). When using whole samples, the correlation between stock returns and trading volumes of companies are close to zero. For some companies, in spite of their significance the computed values are small. For all companies under study, there is no correlation for the pairs low stock returns-low trading volumes and high stock returns-low trading volumes. The correlation coefficients $\tau_{u_1>0.99, u_2>0.99}$ and $\tau_{u_1<0.01, u_2>0.99}$ are significant for the majority of the sample and greater than 0.1 To sum up, even using rank correlation coefficients it is impossible to model dependence structures. One can model relationships for negative and positive returns separately but it is not then obvious what is the ratio of the correlations. Moreover, the correlations presented above are not equivalent to tail dependence coefficients.

The Tables 8 and 9 contain the results of the estimation of the parameters of the mixtures (absolute values of parameters are given) along with dependence measures upon copulas.

The coefficient λ_{HH} describes the asymptotic dependence between extremely high positive stock returns and extremely high volume, whereas λ_{LH} is related to extremely low stock returns. These coefficients are computed using mixing parameters. As in the previous section, α_1 and α_2 are the parameters of copulas used in mixtures, τ is the Kendall correlation coefficient and ω is the mixture parameter.

Table 8

Estimation results and dependence measures for pair $\tilde{r}_t - \widetilde{\ln V_t}$

company	α_1	α_2	ω	τ	λ_{HH}	λ_{LH}
ALSTOM	0.74	0.72	0.49	0.27	0.19	0.20
CARREFOUR	0.74	0.71	0.51	0.27	0.20	0.19
VINCI	0.52	0.68	0.51	0.23	0.13	0.18
company	α_1	α_2	ω	τ	λ_{HH}	λ_{LH}
ASTRAZENECA	0.68	0.69	0.45	0.26	0.16	0.20
KINGFISHER	0.50	0.58	0.55	0.21	0.14	0.14
TESCO	0.57	0.60	0.55	0.23	0.16	0.14

Source: own elaboration based on Reuters data basis

In almost all cases the mixture $\omega C_{Cl}^{(180)} + (1-\omega) C_{Cl}^{(270)}$ fits the dataset best. The only exception is the English stock Evraz (mixture of $\omega C_{Gum} + (1-\omega) C_{Cl}^{(270)}$). The table below presents the rank statistics of dependence measures for all stocks under study.

Table 9

Dependence measures for pair $\tilde{r}_t - \widetilde{\ln V_t}$

	CAC40			FTSE100		
statistics	τ	λ_{HH}	λ_{LH}	τ	λ_{HH}	λ_{LH}
minimum	0.13	0.05	0.05	0.10	0.03	0.00
1st quartile	0.21	0.13	0.16	0.15	0.08	0.07
median	0.23	0.15	0.17	0.18	0.10	0.10
3rd quartile	0.25	0.18	0.19	0.20	0.13	0.14
maximum	0.29	0.22	0.22	0.26	0.18	0.21

Source: own elaboration based on Reuters data basis

The dependence structures in the analyzed corners are not unique. For 15 out of CAC40 companies the dependence between the pair high returns-high trading volume is stronger than that between low returns-high volume. English stocks are characterized mostly (in 59 cases) by the strongest high returns-high volume dependence.

5. Conclusions

We analyzed the dependence stuctures of stock returns, volatility and trading volumes of companies listed on the CAC40 and FTSE100. Additionally, we tested the MDH with long memory i.e. the equality of the long memory parameters of volatility and trading volume and fractional cointegration of these series. With some exceptions the estimation results of long memory parameters show that the series under study are stationary.

Moreover, taking into account the lack of fractional cointegration, the extended hypothesis is rejected in all cases. This means that a common long-run dependence does not exist. In other words, the series are not driven by a common information arrival process with long memory.

The correlation between volatility and trading volume is present for almost all stocks of companies under investigation. There exists a significant dependence between high volatility and high trading volume. In general dependence is stronger for the French than for the English stocks.

It was noted that the classical correlation coefficient (even rank correlation) does not allow the capture of the specific dependence structures of returns and trading volume. Using mixtures of rotated copulas and a Kendall correlation coefficient based upon them, extreme return-volume dependence was investigated. In the case of CAC40 companies we can conclude that high trading volume is not correlated as frequently with high stock returns as with low stock returns. For companies listed on the FTSE100 high stock returns are mostly related with high trading volume.

References

[1] Andersen T.G., *Return volatility and trading volume: An information flow interpretation of stochastic volatility*, "Journal of Finance" 1996, vol. 51, pp. 169–204.

[2] Balduzzi P., Kallal H., Longin F., *Minimal Returns and the Breakdown of the Price-Volume Relation*, "Economic Letters" 1996, vol. 50, pp. 265–269.

[3] Bollerslev T., Jubinski D., *Equity trading volume volatility: latent information arrivals and common long-run dependencies*, "Journal of Business & Economic Statistics" 1999, vol. 17, pp. 9–21.

[4] Bollerslev T., Mikkelsen H.O., *Modeling and pricing long memory in stock market volatility*, "Journal of Econometrics" 1996, vol. 73, pp. 151–184.

[5] Clark P.K., *A subordinated stochastic process model with finite variance for speculative prices* "Econometrica" 1973, vol. 41, pp. 135–155.

[6] Copeland T., *A model of asset trading under the assumption of sequential information arrival*, "Journal of Finance" 1976, vol. 31, pp. 135–155.

[7] Ding Z., Granger C.W.J., Engle R.F., *A Long Memory Property of Stock Market Returns and a New Model*, "Journal of Empirical Finance" 1993, vol. 1, pp. 83–106.

[8] Epps T., Epps M., *The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distribution hypothesis*, "International Review of Financial Analysis" 1976, vol. 44, pp. 305–321.

[9] Fleming J., Kirby C., *Long memory in volatility and trading volume*, "Journal of Banking & Finance" 2011, Elsevier, vol. 35(7), pp. 1714–1726.

[10] Granger C.W.J., *Some recent development in a concept of causality*, "Journal of Econometrics" 1988, vol. 39, pp. 199–211.

[11] Granger C.W.J., Newbold P., *Spurious regressions in econometrics*, "Journal of Econometrics" 1974, vol. 2, pp. 111–120.

[12] Gurgul H., Mestel R., Syrek R., *Kopule i przyczynowość w badaniach związków pomiędzy zmiennymi finansowymi wybranych spółek z DAX*, "Ekonomia Menedżerska" 2008, vol. 3, pp. 7–20.

[13] Jennings R., Starks L., Fellingham J., *An Equilibrium model of asset trading-with sequential information arrival*, "Journal of Finance" 1981, vol. 36, pp. 143–161.

[14] Lobato I.N., Velasco C., *Long memory in stock-market trading volume*, "Journal of Business & Economic Statistics" 2000, vol. 18(4), pp. 410–427.

[15] Marsh T. A., Wagner N., *Return-Volume Dependence and Extremes in International Equity Markets*, Working Paper, Haas School of Business 2000.

[16] Nelsen R., *An Introduction to Copulas.* Springer, New York 1999.

[17] Nielsen M.Ø., Shimotsu, K., *Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach*. "Journal of Econometrics" 2007, vol. 141, pp. 574–96.

[18] Ning C., Wirjanto T., *Extreme return-volume dependence in East-Asian stock markets: A copula approach*, "Finance Research Letters" 2009, vol. 6(4), pp. 202–209.

[19] Phillips P.C.B., *Understanding spurious regressions in econometrics*, "Journal of Econometrics" 1986, vol. 33, pp. 311–340.

[20] Phillips P.C.B., Shimotsu K., *Local Whittle estimation in nonstationary and unit root cases*, "Annals of Statistics" 2004, vol. 34(2), pp. 656–692.

[21] Phillips P.C.B., Shimotsu K., *Exact local Whittle estimation of fractional integration*, "Annals of Statistics" 2005, vol. 33(4), pp. 1890–1933.

[22] Phillips P.C.B., Shimotsu K., *Local Whittle estimation of fractional integration and some of its variants*, "Journal of Econometrics" 2006, vol. 103(2), pp. 209–233.

[23] Robinson P. M., Yajima Y., *Determination of cointegrating rank in fractional systems*, "Journal of Econometrics" 2002, vol. 106 (2), pp. 217–241.

[24] Rossi E., de Magistris P.S., *Long memory and tail dependence in trading volume and volatility*, "CREATES Research Paper" 2010, No. 2009-30.

[25] Shimotsu K., *Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend*, "Econometric Theory", Cambridge University Press 2010, vol. 26(02), pp. 501–540.

[26] Tauchen G.E., Pitts M., *The price variability – volume relationship on speculative markets*, "Econometrica" 1983, vol. 51, pp. 485–505.