Cena 15,00 zł
(VAT 5%)

Indeks 371262
ISSN 0033-2372

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

PRZEGLĄD STATYSTYCZNY

STATISTICAL REVIEW

TOM 65

2

2018

100 lat GUS

Information for Authors

1. *Statistical Review* publishes scientific articles in the field of statistics, econometrics and other disciplines using quantitative methods to study economic phenomena. The submitted works should contain significant theoretical contributions or interesting empirical applications. Articles related to research conducted as part of research projects are especially required. Also, book reviews, reports on scientific activity of statisticians and econometricians, as well as studies containing original proposals in the field of didactics of statistics and econometrics are published in *Statistical Review*.
2. Articles are published in Polish and English. In the latter case, the author should send the English text carefully prepared and revised in terms of language.
3. A typescript with a maximum of 20 pages (including tables and diagrams, written using the Times New Roman 12-point font, with 1.5 line spacing, with margins of 2.5 cm) should be edited through the journal's editorial platform: http://www.editorialsystem.com/pst/
4. The literature that is cited should be listed in alphabetical order (in titles in English the first letter of the word should be capitalised), for example:
   Bauwens L., Laurent S., Rombouts J. V. K., (2006), Multivariate GARCH Models: A Survey, *Journal of Applied Econometrics*, 21 (1), 79–110.
   Brockwell P. J., Davis R. A., (1996), *Introduction to Time Series and Forecasting*, Springer-Verlag, New York.
   The text should use the Harvard referencing style, for example: This problem was discussed in the works of Granger (1969) and Brockwell and Davis (1996). This issue has been presented in many works (see *e.g.*, Granger, 1969; Brockwell and Davis, 1996; Bauwens *et al*., 2006).
5. If the work is divided into sections, the sections should be numbered with Arabic numerals. It is necessary to use continuous numbering for tables and figures (marked as Table 1, Table 2, *etc*., Figure 1, Figure 2, *etc.*).
6. Submitted papers should include (at the end of the paper) the title, a summary of the work (not exceeding 1/2 page of the typescript) and key words, all in Polish and English.
7. Please, also provide the information on the affiliation of the authors (name of the institution, faculty, department, postal address) and the e-mail address of the author responsible for correspondence in the footnote referring to the name. If the paper is an effect of a research project, this fact should be mentioned in the footnote referring to the title of the study, giving the number and title of the project.
8. Please be advised that in the process of reviewing the submitted papers, double anonymity will be maintained. Therefore, the paper should be sent in an anonymous version, and all identification information should be deleted.
9. Authors of papers accepted for publication are required to send scanned statements about the originality of the paper and the contribution of individual authors to the preparation of the publication as well as the transfer of the author's property rights.
10. Submission of an paper to *Statistical Review* and the fact of its acceptance for publication also means the author's consent for its publication on the website of the journal and in the databases of journals in which *Statistical Review* is included.
11. Papers that do not meet the requirements given will not be considered.

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

PRZEGLĄD STATYSTYCZNY

STATISTICAL REVIEW

TOM 65

2

2018

WARSZAWA 2018

Strona www „Przegląd Statystyczny”:
http://www.przegladstatystyczny.pan.pl

Informacje w sprawie sprzedaży czasopisma tel.: 22 608 32 10, 22 608 38 10

ISSN 0033-2372

Indeks 371262

ZAKŁAD WYDAWNICTW STATYSTYCZNYCH
al. Niepodległości 208, 00-925 Warszawa, tel. 22 608 31 45.
Zbigniew Karpiński (redaktor techniczny), Katarzyna Szymańska (skład i łamanie)

Zam. 496/2018 — nakład 220 egz.

CONTENTS

Anna Staszewska-Bystrova — Refined Bonferroni prediction bands for autoregressive models **155**

Piotr Kębłowski — A Monte Carlo comparison of LCCA- and ML-based cointegration tests for panel var process with cross-sectional cointegrating vectors **173**

Krzysztof Piasecki, Joanna Siwek — Multi-asset portfolio with trapezoidal fuzzy present values **183**

Robert Szóstakowski — The use of the Hurst exponent to investigate the quality of forecasting methods of ultra-high-frequency data of exchange rates **200**

Iwona Bąk, Katarzyna Cheba — Study of spatial uniformity of sustainable development of the European Union before, during and after the economic crisis **224**

DOI: 10.5604/01.3001.0014.0532

PRZEGLĄD STATYSTYCZNY
TOM LXV – ZESZYT 2 – 2018

Anna STASZEWSKA-BYSTROVA[1]

Refined Bonferroni prediction bands for autoregressive models[2]

1. INTRODUCTION

A number of papers address the problem of constructing prediction bands for multivariate and univariate autoregressive (AR) models (see e.g., Jordà, Marcellino, 2010; Staszewska-Bystrova, 2011, 2013; Staszewska-Bystrova, Winker, 2013; Wolf, Wunderli, 2015). Joint prediction bands are designed to contain the future trajectory of a predicted variable with probability given by the coverage level and therefore provide valuable information on the predictive uncertainty. The most successful methods of band construction use the bootstrap (Efron, 1979) to derive the relevant predictive distributions. Bootstrap methods are also commonly used for this class of models for forming prediction intervals (see i.a. Thombs, Schucany, 1990; Masarotto, 1990; Breidt et al., 1995; Grigoletto, 1998; Kim, 2001; Clements, Kim, 2007).

The methods of building joint bands, which have been proposed, lead to obtaining prediction regions which differ with respect to the estimated coverage levels and widths. Simulation studies reported by Lütkepohl et al. (2015*a*, 2015*b*) in the context of constructing confidence bands for impulse responses show that the conservative bootstrap Bonferroni bands are quite successful in terms of maintaining the nominal coverage probability. However, the estimated coverage rates are often larger than the nominal values for these bands. Excessive coverage is, in turn, associated with unnecessarily large width of the bands.

The aim of this paper is to refine the basic bootstrap Bonferroni bands in two ways: first, by applying higher order Bonferroni-type inequalities (Hoover, 1990, see also Glaz, Ravishanker, 1991) and second, by considering imbalanced Bonferroni bands found through optimization. Both refinements should lead to reductions in the width of the bands. The working of the methods is compared to the performance of the sup-*t* procedure described by Wolf, Wunderli (2015).

[1] University of Lodz, Faculty of Economics and Sociology, Chair of Econometric Models and Forecasts, 41 Rewolucji 1905r. St., 90–214 Lodz, Poland, e-mail: anna.bystrova@uni.lodz.pl.

[2] Financial support provided by the National Science Center, Poland (NCN) through HARMONIA 6: UMO-2014/14/M/HS4/00901 is gratefully acknowledged.

The proposed methods are applied to persistent AR models containing a linear trend. The parameters of such models can be estimated using various methods. Some standard estimators including the ordinary least squares (OLS) or the Yule-Walker estimators are, however, not recommended due to their small sample bias (see e.g. Andrews, Chen, 1994). Alternative estimation methods for univariate AR models with time trend have been proposed for instance by Andrews, Chen (1994), Kilian (1998) and Roy, Fuller (2001). Clements, Kim (2007) report that bootstrap prediction intervals for the AR model based on the approximately median unbiased Roy-Fuller estimator have the best small sample properties. This estimator is therefore applied in the study reported below.

The structure of the paper is as follows. The next section presents the AR framework and the estimation method used. In section 3 the bootstrap algorithm for obtaining predictive distributions is described and in section 4 the standard Bonferroni bands, the proposed refinements and the benchmark sup-*t* bands are discussed. Section 5 presents the Monte Carlo comparison of the methods while section 6 concludes.

2. THE MODEL

The model considered in this paper is an AR(*p*) with intercept and a linear time trend (see Box, Jenkins, 1970; Lütkepohl, Krätzing, 2004):

$$y_t = \mu + \beta t + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon_t, \tag{1}$$

where $\varepsilon_t \sim \mathrm{iid}(0, \sigma^2)$.

The model can be reparametrized either as

$$y_t = \mu + \beta t + \gamma_1 y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t, \tag{2}$$

where $\Delta y_t = y_t - y_{t-1}$, $\gamma_1 = \sum_{i=1}^{p} \alpha_i$, $\alpha_1 = \gamma_1 + \delta_1$, $\alpha_i = \delta_i - \delta_{i-1}$ for $2 \leq i \leq p-1$ and $\alpha_p = -\delta_{p-1}$ or as

$$y_t = \mu + \beta t + \gamma_{-1} y_{t-1} + \theta_1 S y_{t-1} + \cdots + \theta_{p-1} S y_{t-p+1} + \varepsilon_t, \tag{3}$$

where $S y_t = y_t + y_{t-1}$, $\gamma_{-1} = \sum_{i=1}^{p} (-1)^{i+1} \alpha_i$, $\alpha_1 = \gamma_{-1} + \theta_1$, $\alpha_i = \theta_i + \theta_{i-1}$ for $2 \leq i \leq p-1$ and $\alpha_p = \theta_{p-1}$. The parameter γ_1 in (2) describes the persistence of the AR process. In what follows it is assumed to belong to the interval $(-1,1)$.

Given the pre-sample values $y_{-p+1}, \ldots, y_0$ and the sample values $y_1, \ldots, y_T$, the parameters of model (2) can be estimated using the method proposed by Roy,

Fuller (2001). The Roy-Fuller (RF) estimator of γ_1 is approximately unbiased and its mean squared error is smaller than that of the ordinary least squares estimator for time series with a root near 1. The estimator is defined as

$$\hat{\gamma}_1^{RF} = \min(\tilde{\gamma}_1, 1), \tag{4}$$

where $\tilde{\gamma}_1 = \dot{\gamma}_1 + [C_p(\dot{\tau}_1) + C_{-p}(\dot{\tau}_{-1})]\dot{\sigma}_1$, $\dot{\gamma}_1$ is the least squares estimator of the parameter of $\hat{y}_{t-1}$ in the model where $\hat{y}_t$ is regressed on $\hat{y}_{t-1}, \Delta\hat{y}_{t-1}, \ldots, \Delta\hat{y}_{t-p+1}$, $\hat{y}_t$ denotes the least squares residual from the regression of y_t on the constant and linear trend t and $\dot{\sigma}_1$ is the standard error of $\dot{\gamma}_1$. The functions $C_p(\dot{\tau}_1)$ and $C_{-p}(\dot{\tau}_{-1})$ are based respectively on the unit root statistic

$$\dot{\tau}_1 = \frac{\dot{\gamma}_1 - 1}{\dot{\sigma}_1}, \tag{5}$$

and the negative unit-root statistic

$$\dot{\tau}_{-1} = \frac{\dot{\gamma}_{-1} + 1}{\dot{\sigma}_{-1}}, \tag{6}$$

where $\dot{\gamma}_{-1}$ and $\dot{\sigma}_{-1}$ are the least squares estimator of the coefficient of $\hat{y}_{t-1}$ in the model where $\hat{y}_t$ is regressed on $\hat{y}_{t-1}, S\hat{y}_{t-1}, \ldots, S\hat{y}_{t-p+1}$ and the standard error of $\dot{\gamma}_{-1}$. $C_p(\dot{\tau}_1)$ has the form[3]

$$\begin{aligned}
C_p(\dot{\tau}_1) &= -\tau_{med} + d_T(\dot{\tau}_1 - \tau_{med}), && \dot{\tau}_1 > \tau_{med},\\
&= I_p(T^{-1}\dot{\tau}_1) - 3[\dot{\tau}_1 + k(\dot{\tau}_1 - K)]^{-1}, && K < \dot{\tau}_1 \le \tau_{med},\\
&= I_p(T^{-1}\dot{\tau}_1) - 3[\dot{\tau}_1]^{-1}, && -\sqrt{\frac{3T}{I_p}} < \dot{\tau}_1 \le K,\\
&= 0, && \dot{\tau}_1 \le -\sqrt{\frac{3T}{I_p}},
\end{aligned}$$

where I_p stands for the integer part of $\frac{1}{2}(p+1)$, τ_{med} is the median of the limiting distribution of $\dot{\tau}_1$ under the null, $k = \left[3T - \tau_{med}^2\left(I_p + T\right)\right][\tau_{med}(\tau_{med} - K)(I_p + +T)]^{-1}$. The constants *K* and d_T are set, following Roy, Fuller (2001), to –5 and 0.29, respectively. The expression for $C_{-p}(\dot{\tau}_{-1})$ is as follows

[3] Apart from the paper by Roy, Fuller (2001) see also the errata available at Anindya Roy's webpage.

$$\begin{aligned} C_{-p}(\dot{t}_{-1}) &= 0, && \dot{t}_{-1} \geq \sqrt{k_{-1}}, \\ &= \left(\left\lfloor \frac{p+1}{2} \right\rfloor + I_{2p}\right) T^{-1}\dot{t}_{-1} - \dot{t}_{-1}^{-1}, && K \leq \dot{t}_{-1} < \sqrt{k_{-1}}, \\ &= a_T + b_T(\dot{t}_{-1} + K), && \dot{t}_{-1} < K, \end{aligned}$$

where I_{2p} is equal to 0 if p is even and is given by 3 if p is odd, $\lfloor q \rfloor$ stands for the greatest integer less than or equal to q, $k_{-1} = \left(\left\lfloor \frac{p+1}{2} \right\rfloor + I_{2p}\right)^{-1} T$, $a_T = C_{-p}^{*}(-K)$, $b_T = C_{-p}^{*\prime}(-K)$, and $C_{-p}^{*}(\dot{t}_{-1}) = \left(\lfloor (p+1)/2 \rfloor + I_{2p}\right) T^{-1}\dot{t}_{-1} - \dot{t}_{-1}^{-1}$.

Given an estimate $\hat{\gamma}_1^{RF}$, the parameters $\mu, \beta, \delta_1, \ldots, \delta_{p-1}$ can be estimated from the regression of $y_t - \hat{\gamma}_1^{RF} y_{t-1}$ on the constant, trend and lagged differences $\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$, producing $\hat{\mu}^{RF}, \hat{\beta}^{RF}, \hat{\delta}_1^{RF}, \ldots, \hat{\delta}_{p-1}^{RF}$. The only exception arises if $\hat{\gamma}_1^{RF} = 1$ when the parameter on trend is restricted to 0. In the next step, estimates of $\alpha_1, \ldots, \alpha_p$, denoted by $\hat{\alpha}_1^{RF}, \ldots, \hat{\alpha}_p^{RF}$, can be obtained. The variance of the random error ε_t can be estimated using

$$\hat{\sigma}^2 = \frac{1}{T-l} \sum_{t=1}^{T} \hat{\varepsilon}_t^2, \tag{7}$$

where l stands for the number of estimated coefficients and

$$\hat{\varepsilon}_t = y_t - \hat{\mu}^{RF} - \hat{\beta}^{RF} t - \hat{\alpha}_1^{RF} y_{t-1} - \cdots - \hat{\alpha}_p^{RF} y_{t-p}. \tag{8}$$

The point forecasts $\hat{y}(h)$ for $1, \ldots, H$ and the corresponding prediction standard errors $\hat{\sigma}(h)$ may be calculated according to

$$\hat{y}(h) = \hat{\mu}^{RF} + \hat{\beta}^{RF}(T+h) + \hat{\alpha}_1^{RF}\hat{y}(h-1) + \cdots + \hat{\alpha}_p^{RF}\hat{y}(h-p), \tag{9}$$

where $\hat{y}(j) = y_{T+j}$ for $j \leq 0$,

and

$$\hat{\sigma}(h) = \hat{\sigma}^2 \sqrt{\hat{\theta}_0^2 + \cdots + \hat{\theta}_{h-1}^2}, \tag{10}$$

where $\hat{\theta}_j = \sum_{i=1}^{j} \hat{\theta}_{j-i}\, \hat{\alpha}_i^{RF}$ for $j = 1,2, \ldots, h-1$ with $\hat{\alpha}_i^{RF} = 0$ for $i > p$ and $\hat{\theta}_0 = 1$.

In what follows the main interest lies in using the Roy-Fuller estimator and the bootstrap method for constructing prediction bands which should cover the future *H*-dimensional path of realizations $y(H) = (y_{T+1}, \ldots, y_{T+H})'$ which a preassigned probability.

3. THE BOOTSTRAP ALGORITHM

Predictive distributions and estimates of standardized prediction errors are obtained using the residual bootstrap procedure. Calculations involve a number of steps:

1) The parameters of model (1) are estimated using the Roy-Fuller method and the corresponding residuals are computed. The residuals from (8) are inflated using a factor of $\sqrt{\frac{T}{T-l}}$ (see e.g. Stine, 1987) and denoted by $\hat{\varepsilon}_t^*$.
2) A sample of pseudo-data of size T is generated from the bootstrap data generating process of the form (see Clements, Kim, 2007; Fresoli et al., 2015):

$$y_t^* = \hat{\mu}^{RF} + \hat{\beta}^{RF} t + \hat{\alpha}_1^{RF} y_{t-1}^* + \cdots + \hat{\alpha}_p^{RF} y_{t-p}^* + \varepsilon_t^*, \tag{11}$$

where actual observations $y_{-p+1}, \ldots, y_0$ are used as pre-sample values $y_{-p+1}^*, \ldots, y_0^*$ and ε_t^* is drawn randomly from the rescaled residual series $\hat{\varepsilon}_t^*$.

3) The pseudo-data set is used to re-estimate the parameters of model (1) producing $\hat{\mu}^{RF*}, \hat{\beta}^{RF*}, \hat{\alpha}_1^{RF*}, \ldots, \hat{\alpha}_p^{RF*}$ and also to compute forecasts $\hat{y}^*(h)$ and prediction standard errors $\hat{\sigma}^*(h)$ as in (9) and (10) but with $\hat{\mu}^{RF}, \hat{\beta}^{RF}, \hat{\alpha}_1^{RF}, \ldots, \hat{\alpha}_p^{RF}$ replaced by $\hat{\mu}^{RF*}, \hat{\beta}^{RF*}, \hat{\alpha}_1^{RF*}, \ldots, \hat{\alpha}_p^{RF*}$.
4) Bootstrap future trajectory for horizon H, $(y^*(1), \ldots, y^*(H))'$ is generated using:

$$\begin{aligned} y^*(h) = \hat{\mu}^{RF*} + \hat{\beta}^{RF*}(T+h) + \hat{\alpha}_1^{RF*} y^*(h-1) + \cdots + \\ + \hat{\alpha}_p^{RF*} y^*(h-p) + \varepsilon_h^*, \end{aligned} \tag{12}$$

where $h = 1, \ldots, H$, $y^*(i) = y_{T+i}$ for $i \leq 0$ and ε_h^* is drawn randomly from the series $\hat{\varepsilon}_t^*$.

5) Bootstrap vector of standardized prediction errors $\hat{S}^*(H) = (\hat{s}^*(1), \ldots, \hat{s}^*(H))'$ is evaluated by generating y_t^* for $T+1 \leq t \leq T+H$ analogously as in (11) but with initial values given by $y_{T-p+1}, \ldots, y_T$ and calculating, for $h = 1, \ldots, H$:

$$\hat{s}^*(h) = \frac{\hat{y}^*(h) - y^*(h)}{\hat{\sigma}^*(h)}. \tag{13}$$

The procedure in steps (2)–(5) is repeated *N* times (where *N* denotes the number of iterations in the bootstrap loop), providing *N* bootstrap replicates of the future trajectory of y and the same number of replicates of the vector of standardized prediction errors. These values can be used to construct various bootstrap prediction bands.

4. REFINED BONFERRONI BANDS

Prediction bands can be constructed using Bonferroni's method. Suppose the objective is to construct a $(1-\gamma)\times 100\%$ prediction band *B* for the elements of an *H*-dimensional vector $y(H)=(y_{T+1},\dots,y_{T+H})'$. The Bonferroni inequality

$$P(y(H)\in B)\geq \sum_{h=1}^{H}(1-\gamma_h)-(H-1), \tag{14}$$

where $1-\gamma_h$ is the coverage probability of the band at horizon h (i.e. computed with respect to y_{T+h}), indicates that in order to achieve at least the desired coverage of the band, it can be assumed that $\sum_{h=1}^{H}\gamma_h=\gamma$. The most common approach is to set each γ_h to the same value $\frac{\gamma}{H}$. The resulting band is constructed from $\left(1-\frac{\gamma}{H}\right)\times 100\%$ prediction intervals for each element y_{T+h} separately:

$$B=[B_{L1},B_{U1}]\times\ [B_{L2},B_{U2}]\times\dots\times[B_{LH},B_{UH}], \tag{15}$$

where B_{Lh} and B_{Uh} denote respectively, the $\frac{\gamma}{2H}$ and $1-\frac{\gamma}{2H}$ quantiles of the predictive distribution of y_{T+h}.

Given that the actual coverage of the Bonferroni band may easily exceed the desired level and the band may, in effect, be excessively wide (see, e.g. Lütkepohl et al., 2015*b*) it makes sense to try to refine the Bonferroni bands in such a way that the actual coverage becomes closer to the nominal level and the width of the bands is reduced.

The first refinement uses higher order Bonferroni-type inequalities of Hoover (1990). Glaz, Ravishanker (1991) apply these inequalities to construct prediction bands for ARIMA models using the properties of the multivariate normal distribution. In this paper, the bootstrap distribution of the predictor is considered. The condition implied by the Bonferroni-type inequality of order *k*, for $1<k\leq H-1$ for the band B^k has the form:

$$P(y(H)\in B^k)\geq \sum_{h=1}^{H+1-k}\left(1-\gamma_{h,h+k-1}\right)-\sum_{h=1}^{H-k}\left(1-\gamma_{h+1,h+k-1}\right), \tag{16}$$

where for $1\leq m\leq n\leq H$, $(1-\gamma_{m,n})$ is the probability that realizations of the predicted variable observed from horizon *m* to horizon *n* is covered by the corresponding stretch of the band, $\gamma_{m,m}=\gamma_m$ and $\gamma_{m+1,m}\equiv 1$.

The band with equal values of γ_h for $h=1,\dots,H$, could be constructed in an iterative manner by starting from the band derived from Bonferroni's inequality

and reducing the coverage of the intervals in each step by 1/*N*, where *N* is the number of bootstrap replications (making γ_h larger in each step by 1/*N*). In each iteration it would be checked whether the relevant higher order Bonferroni inequality was observed. This can be achieved by computing the bootstrap coverage at single horizons and multiple horizons and evaluating the Bonferroni-type inequality of order k of interest. The final values of γ_h would be the largest values for which the inequality was met. The larger the value of k, the less conservative the resulting band.

The second refinement of the basic Bonferroni band aims at finding such γ_h for $h = 1, \dots, H$, $\sum_{h=1}^{H} \gamma_h = \gamma$ that the width of the band is as small as possible. The resulting Bonferroni band can be described as imbalanced (see e.g. Wolf, Wunderli, 2015).

Optimization is done using threshold accepting (TA) belonging to a class of refined local search methods. The procedure was proposed by Dueck, Scheuer (1990) and applied to the problem of constructing prediction bands e.g. by Staszewska-Bystrova, Winker (2013) and Grabowski et al. (2017).

The objective function which is minimized is the width of the band:

$$W(B^{TA}) = \sum_{h=1}^{H} (B_{Uh}^{TA} - B_{Lh}^{TA}), \tag{17}$$

where B^{TA} is the Bonferroni band obtained using threshold accepting. The detailed steps of the optimization procedure are presented in algorithm 1.

The algorithm is initialized (step 1) by considering as the starting solution B^c, the basic Bonferroni band *B* and evaluating the objective function for this band. The number of search steps (n_{iter}) and the threshold sequence ($t_1, \dots, t_{n_{iter}}$) of the corresponding length are also set. The threshold values should be positive and decreasing. Then n_{iter} iterations are performed. In each iteration i, a new solution (prediction band) belonging to the neighborhood of the current solution is considered. The neighboring band is created randomly by modifying the width of the currently considered band in two points in such a way that the constraint $\sum_{h=1}^{H} \gamma_h = \gamma$ is not violated. To achieve this, two values h_1 and h_2 are randomly selected from the set $\{1, \dots, H\}$ (step 2) and the corresponding values γ_{h_1} and γ_{h_2} are changed (steps 3–5). First, γ_{h_1} is made smaller by subtracting a random fraction of its current value (f). Second, γ_{h_2} is enlarged by f. Then, the width of the new band B^n is computed and compared to the width of the current band (step 6). If the difference is smaller than the threshold value for iteration i, then the solution is accepted as the current solution (step 7). The algorithm continues until n_{iter} search steps are completed. The band with the smallest value of the objective function found throughout the search steps, denoted by B^{TA} is presented as the final solution.

Algorithm 1.

Threshold accepting procedure

1. Obtain initial solution B^c and compute $W(B^c)$. Set the value of n_{iter} and $t_1, \dots, t_{n_{iter}}$
2. for $i = 1, \dots, n_{iter}$ do
3. Randomly select two integers: h_1 and h_2 from the set $\{1, \dots, H\}$
4. Randomly select *p* from the interval (0,1) and compute $f = p\gamma_{h_1}^c$
5. Obtain new solution B^n by setting $\gamma_{h_1}^n = \gamma_{h_1}^c - f$ and $\gamma_{h_2}^n = \gamma_{h_2}^c + f$
6. Calculate $\Delta = W(B^n) - W(B^c)$
7. if $\Delta < t_i$ then $B^c = B^n$
8. end for

The proposed Bonferroni-type bands are compared to the sup-*t* method described by Wolf, Wunderli (2015). The benchmark procedure has some optimality properties (e.g smaller width) as compared to the traditional Bonferroni algorithm in large samples (see Montiel Olea, Plagborg-Møller, 2017). The sup-*t* bands are computed by finding the largest value in each of the N vectors $|\hat{S}^*(H)|$ and obtaining $d_{1-\alpha}$ equal to the $1-\alpha$ quantile of these maxima. In the next step the band is formed as

$$[\hat{y}(1) \pm d_{1-\alpha}\hat{\sigma}(1)] \times \dots \times [\hat{y}(H) \pm d_{1-\alpha}\hat{\sigma}(H)]. \tag{18}$$

5. A SIMULATION STUDY

Small-sample properties of the bands were studied using Monte Carlo simulations. A number of data generating processes (DGPs) were investigated. The first set of DGPs (denoted by DGP.A), considered also by Clements, Kim (2007), had the form:

$$y_t = 1 + (1-\alpha)t + \alpha y_{t-1} + \varepsilon_t, \tag{19}$$

where $\alpha \in \{0.5, 0.9, 0.95\}$. Higher values of α correspond to larger degree of persistence of the AR process. Three different distributions of the errors, ε_t, were considered for DGP.A: a standard normal distribution ($N(0,1)$), a chi-square distribution with 4 degrees of freedom, centered to have mean 0 and standardized to have variance 1 and a *t*-distribution with 4 degrees of freedom standardized to have variance equal to 1.

More complex DGPs (DGP.B), corresponding to AR(2) were given by:

$$y_t = 1 + (1.85 - \alpha_1)t + \alpha_1 y_{t-1} - 0.85 y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim N(0,1), \tag{20}$$

where $\alpha_1 \in \{1.35, 1.75, 1.8\}$. For these processes $\gamma_1 = \alpha_1 - 0.85$ and so the persistence also grows as the value of α_1 increases.

The other settings of the Monte Carlo experiments were as follows. The number of Monte Carlo replications *M* was set to 1000 and the number of iterations in the bootstrap procedure *N* was equal to 2000.[4] In each Monte Carlo iteration, parameters of an AR model with a constant and trend were estimated. The lag order was selected using Akaike's information criterion (AIC) allowing for up to 8 lags. The same number of lags was used for the models estimated in the bootstrap procedure. The sample size *T* and the forecast horizon *H* belonged respectively to the following sets: $T \in \{100, 400\}$ and $H \in \{4, 8, 12\}$. The nominal coverage rate of the bands was given by 0.9.

Further parameter settings corresponded to specific methods of constructing bands. The value of k for the procedure based on higher order Bonferroni's inequality was equal to 2, 3 or 4. Since the combination $k = 4$ and $H = 4$ is not feasible due to the condition $k \leq H - 1$, the results were not obtained for these cases and the corresponding entries in tables are given as NA. TA optimization was performed for $n_{iter} = 500000$ and the threshold sequence defined as $t_i = \frac{n_{iter}-i}{n_{iter}} \times 0.05$ for $i = 1, \ldots, n_{inter}$.

Two properties of prediction bands were evaluated in the simulations: mean coverage rates and average width. In order to evaluate the coverage probabilities, 1000 future trajectories of length *H* (each computed conditionally on the last *p* values from the generated sample) were obtained from the DGPs. Then, in every Monte Carlo replication the proportions of trajectories lying entirely within the alternative prediction bands were computed. Mean coverage rates were obtained as averages of these proportions over *M* replications. To provide a measure of width of the bands, sum of differences between the upper and lower bounds were calculated for $h = 1, \ldots, H$ and divided by *H*. Average values for *M* Monte Carlo iterations are reported.

The results of all experiments are presented in tables 1–6 and tables A1–A6 from the Appendix. Tables 1–3 and 4–6 contain results obtained respectively for DGP.A with normal errors and DGP.B with alternative parameter values. In tables A1–A3 and A4–A6 estimated coverage rates and width measures are presented for DGP.A with chi-square distributed and *t*-distributed errors. Quantities without parentheses correspond to the estimated coverage probabilities, while values in parentheses indicate average width of the bands.

Table 1. RESULTS FOR DGP.A WITH $\alpha = 0.5$ AND NORMAL ERRORS

H	*B*	sup-*t*	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 100}$			
4	89.16 (5.18)	88.55 (5.07)	87.83 (5.04)	87.59 (5.02)	NA	88.58 (5.12)

[4] Some experiments showed that using 5000 Monte Carlo and 5000 bootstrap iterations did not change the conclusions significantly.

Table 1. RESULTS FOR DGP.A WITH $\alpha = 0.5$ AND NORMAL ERRORS (cont.)

H	*B*	sup-*t*	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 100}$ **(dok.)**			
8	88.67 (6.01)	88.27 (5.87)	86.93 (5.82)	86.68 (5.80)	86.55 (5.79)	87.90 (5.93)
12	88.06 (6.45)	88.09 (6.31)	86.22 (6.26)	85.92 (6.23)	85.79 (6.22)	87.20 (6.36)
			$\boldsymbol{T = 400}$			
4	90.51 (5.00)	89.68 (4.89)	89.35 (4.89)	89.15 (4.87)	NA	89.91 (4.95)
8	90.40 (5.72)	89.68 (5.61)	88.86 (5.58)	88.68 (5.56)	88.59 (5.55)	89.53 (5.66)
12	90.07 (6.10)	89.59 (5.99)	88.42 (5.95)	88.29 (5.94)	88.21 (5.93)	89.00 (6.02)

The Monte Carlo results for the proposed versions of the Bonferroni band can be summarized as follows. While all the refinements work in the expected way and bring down the average width of the prediction bands as compared to the Bonferroni band, the size of the reduction differs between methods and depends on the specific features of the DGP and the sample size. It has also varying impact on the estimated coverage probabilities.

A general observation is that B^{TA} is almost always wider than the widest of the B^k bands, i.e. B^2. The three versions of the bands based on higher order Bonferroni-type inequalities do not differ much in terms of width. Given that B^2 is considerably more aggressive than the Bonferroni band and that further small reductions in width as implied by B^3 and B^4 tend to impair the coverage probabilities of these bands, B^2 might be preferred over the remaining B^k methods. The length of the forecast horizon has an expected impact on the width of the prediction bands for all the methods, i.e. the width grows as H increases, however it does not influence the relative ordering of the Bonferroni-type procedures for band construction.

Table 2. RESULTS FOR DGP.A WITH $\alpha = 0.9$ AND NORMAL ERRORS

H	*B*	sup-*t*	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 100}$			
4	91.36 (6.93)	89.03 (6.46)	88.57 (6.43)	88.15 (6.37)	NA	90.91 (6.80)
8	91.77 (9.90)	88.59 (8.96)	88.13 (8.94)	87.46 (8.81)	87.18 (8.76)	91.46 (9.69)
12	91.74 (12.16)	88.08 (10.85)	87.84 (10.84)	12.20 (10.66)	86.72 (10.58)	91.51 (11.90)

Table 2. RESULTS FOR DGP.A WITH $\alpha = 0.9$ AND NORMAL ERRORS (cont.)

H	*B*	sup-*t*	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 400}$			
4	92.32 (6.43)	89.77 (6.02)	89.71 (6.05)	89.34 (6.00)	NA	91.88 (6.34)
8	93.28 (8.71)	89.81 (7.97)	89.92 (8.06)	89.30 (7.97)	89.07 (7.93)	92.88 (8.57)
12	93.66 (10.20)	89.81 (9.21)	90.08 (9.40)	89.41 (9.28)	89.10 (9.23)	93.26 (10.03)

Table 3. RESULTS FOR DGP.A WITH $\alpha = 0.95$ AND NORMAL ERRORS

H	*B*	sup-*t*	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 100}$			
4	91.54 (7.14)	89.04 (6.62)	88.62 (6.59)	88.18 (6.52)	NA	91.05 (7.00)
8	92.06 (10.57)	88.47 (9.47)	88.19 (9.42)	87.48 (9.28)	87.19 (9.22)	91.72 (10.33)
12	91.99 (13.32)	87.70 (11.76)	87.82 (11.69)	87.00 (11.48)	86.61 (11.39)	91.73 (13.01)
			$\boldsymbol{T = 400}$			
4	92.63 (6.73)	89.81 (6.24)	89.85 (6.29)	89.44 (6.23)	NA	92.24 (6.63)
8	93.84 (9.59)	89.90 (8.62)	90.21 (8.78)	89.58 (8.66)	89.30 (8.62)	93.55 (9.42)
12	94.41 (11.73)	89.93 (10.33)	90.55 (10.64)	89.81 (10.48)	89.47 (10.42)	94.16 (11.51)

A more detailed assessment of all the methods leads to the following observations. Results obtained for DGP.A with normal errors (tables 1–3) indicate that performance of the procedures depends on the persistence of the process that generated the data and the available number of observations. These features influence the working of the basic Bonferroni band which becomes more conservative for more persistent processes and larger samples which in turn brings about the need for refinement and also the sup-*t* method which tends to undercover for the smaller sample size, especially if the forecast horizon is long. For $\alpha = 0.5$ (table 1) the B band is overall best for both sample sizes in terms of the estimated coverage probabilities, however for bigger values of α (tables 2–3) these probabilities become too large as compared to the nominal rate of 0.9. This effect can be observed for both smaller sample size and larger sample size, where as expected, it is even more pronounced. In these cases, selected alter-

native methods may be preferable. In particular, if the number of observations is small, B^{TA} bands could be used as they maintain the nominal coverage rate, while for larger sample sizes B^2 and sup-*t* bands have the best coverage properties.[5]

Table 4. RESULTS FOR DGP.B WITH $\alpha_1 = 1.35$

H	*B*	sup-*t*	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 100}$			
4	90.41 (7.68)	88.62 (7.29)	87.27 (7.15)	87.19 (7.14)	NA	89.78 (7.54)
8	89.55 (10.00)	88.55 (9.72)	86.44 (9.44)	86.41 (9.44)	86.11 (9.39)	88.72 (9.82)
12	89.12 (11.58)	88.57 (11.34)	86.23 (11.00)	86.19 (11.00)	85.77 (10.93)	88.28 (11.37)
			$\boldsymbol{T = 400}$			
4	92.13 (7.45)	89.69 (6.99)	89.18 (6.95)	89.10 (6.94)	NA	91.63 (7.33)
8	92.16 (9.83)	89.74 (9.30)	89.37 (9.29)	89.34 (9.29)	89.00 (9.23)	91.41 (9.67)
12	92.10 (11.39)	89.74 (10.80)	89.50 (10.84)	89.47 (10.83)	89.10 (10.76)	91.30 (11.20)

Table 5. RESULTS FOR DGP.B WITH $\alpha_1 = 1.75$

H	*B*	sup-*t*	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 100}$			
4	92.39 (11.55)	88.85 (10.35)	87.61 (10.18)	87.42 (10.14)	NA	91.50 (11.17)
8	92.53 (20.13)	88.21 (17.63)	86.09 (17.23)	85.87 (17.15)	85.78 (17.13)	91.98 (19.51)
12	92.20 (25.08)	87.94 (21.92)	85.28 (21.44)	85.07 (21.36)	84.99 (21.33)	91.67 (24.41)
			$\boldsymbol{T = 400}$			
4	94.02 (11.15)	89.80 (9.85)	89.49 (9.83)	89.31 (9.79)	NA	93.45 (10.83)
8	94.85 (18.87)	89.70 (16.29)	89.09 (16.24)	88.90 (16.18)	88.83 (16.16)	94.49 (18.37)
12	94.82 (22.90)	89.62 (19.91)	88.85 (19.87)	88.70 (19.81)	88.64 (19.79)	94.35 (22.38)

[5] Some additional simulation results obtained for DGP.A with $\alpha = 0.9$ indicate that B^{TA} bands might be considered as superior to B^2 and sup-*t* bands for sample sizes smaller than 130, where the mean coverage of the latter bands falls below 0.89 for longer forecast horizons.

Table 6. RESULTS FOR DGP.B WITH $\alpha_1 = 1.8$

H	*B*	sup-*t*	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 100}$			
4	92.63 (12.28)	88.82 (10.89)	87.65 (10.73)	87.46 (10.69)	NA	91.64 (11.84)
8	93.19 (24.11)	87.97 (20.55)	86.26 (20.19)	85.97 (20.07)	85.85 (20.03)	92.54 (23.23)
12	92.96 (33.44)	87.22 (28.06)	85.04 (27.58)	84.76 (27.44)	84.65 (27.38)	92.43 (32.35)
			$\boldsymbol{T = 400}$			
4	94.19 (11.80)	89.81 (10.35)	89.52 (10.35)	89.33 (10.30)	NA	93.56 (11.43)
8	95.34 (22.44)	89.75 (18.94)	89.26 (18.96)	89.01 (18.86)	88.92 (18.83)	95.03 (21.74)
12	95.60 (29.94)	89.66 (25.06)	89.03 (25.13)	88.80 (25.02)	88.71 (24.97)	95.29 (29.11)

Conclusions from the results for DGP.B (tables 3–6) are similar as those for DGP.A. Refined and the sup-*t* methods are most useful in larger samples and for predicting persistent processes for which the Bonferroni band tends to be too wide and have excessive coverage probability. The largest gains can be obtained for larger samples for the sup-*t* and B^2 methods, for which the reduction in band width can be considerable (e.g. more than 15% for $\alpha_1 = 1.8, T = 400$ and $H = 12$). At the same time the coverage probabilities for these procedures are quite close to 0.9, especially in the case of the sup-*t* method. For prediction based on persistent processes and smaller data sets B^{TA} method could be selected.

As follows from the analysis of tables A1–A6, the findings for normal DGPs are to some extent, robust with respect to the distribution of the random errors. For more persistent processes with either chi-square errors or *t*-distributed errors considered in tables A2–A3 and A5–A6, as previously, the B^{TA} procedure could be considered as most robust for all values of *H* for the smaller sample size, while sup-*t* or B^2 bands, which have very similar properties, would be the natural choice for larger samples. Some new effects can be observed for DGP.A with $\alpha = 0.5$, however. For the process with chi-square innovations (table A1), the Bonferroni bands still have the best coverage properties for $T = 100$, however there is a new winner for $T = 400$ given by the B^{TA} procedure. The B^{TA} bands do not undercover for any value of *H* and are narrower as compared to the Bonferroni bands (by construction) and to the sup-*t* bands. As table A4 reveals, a different method, namely the sup-*t* procedure is most reliable for this less persistent process with fat-tailed distribution of the error terms for both $T = 100$ and $T = 400$.

6. CONCLUSIONS

Joint prediction bands are needed for forming expectations concerning the future trajectory of a variable. The construction of such bands is usually based on the bootstrap predictive distribution. One of the classic approaches to build prediction bands rests on the Bonferroni inequality. The drawback of this method is that the bands can be too wide and exhibit larger probability content than the nominal coverage rate.

In this study two refinements of the Bonferroni band were considered in the context of predicting persistent univariate autoregressive processes. The first refinement used higher order Bonferroni-type inequalities, while the second consisted in constructing the band from intervals with unequal coverage rates. The proposed bands were compared to the Bonferroni bands and the benchmark given by the sup-*t* procedure in a Monte Carlo study.

Simulation results indicated that the refined bands were superior to the basic Bonferroni bands in a number of scenarios involving quite persistent processes. In particular, bands based on the second-order Bonferroni-type inequality worked well for relatively large samples where they exhibited similar properties as the sup-*t* bands, while the imbalanced Bonferroni method was preferable for smaller sample sizes.

REFERENCES

Andrews D. W. K., Chen H.-Y., (1994), Approximate Median-Unbiased Estimation of Autoregressive Models, *Journal of Business & Economic Statistics*, 12, 187–204.

Box G. E. P., Jenkins G. M., (1970), *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco.

Breidt F. J., Davis R. A., Dunsmuir W., (1995), Improved Bootstrap Prediction Intervals for Autoregressions, *Journal of Time Series Analysis*, 16, 177–200.

Clements M. P., Kim J. H., (2007), Bootstrap Prediction Intervals for Autoregressive Time Series, *Computational Statistics & Data Analysis*, 51, 3580-3594.

Dueck, G., Scheuer T., (1990), Threshold Accepting: A General Purpose Algorithm Appearing Superior to Simulated Annealing, *Journal of Computational Physics*, 90, 161–175.

Efron B., (1979), Bootstrap Methods: Another Look at the Jackknife, *The Annals of Statistics*, 7, 1–26.

Fresoli D., Ruiz E., Pascual L., (2015), Bootstrap Multi-Step Forecasts of Non-Gaussian VAR Models, *International Journal of Forecasting*, 31, 834–848.

Glaz J., Ravishanker N., (1991), Simultaneous Prediction Intervals for Multiple Forecasts Based on Bonferroni and Product-Type Inequalities, *Statistics & Probability Letters*, 12, 57–63.

Grabowski D. Staszewska-Bystrova A., Winker P., (2017), Generating Prediction Bands for Path Forecasts from SETAR Models, *Studies in Nonlinear Dynamics & Econometrics*, 21, 5.

Grigoletto M., (1998), Bootstrap Prediction Intervals for Autoregressions: Some Alternatives, *International Journal of Forecasting*, 14, 447–456.

Hoover D. R., (1990), Subset Complement Addition Upper Bounds – An Improved Inclusion-Exclusion Method, *Journal of Statistical Planning and Inference*, 24, 195–202.

Jordà Ò., Marcellino M., (2010), Path Forecast Evaluation, *Journal of Applied Econometrics*, 25, 635–662.

Kilian L., (1998), Small-Sample Confidence Intervals for Impulse Response Functions, *Review of Economics and Statistics*, 80, 218–230.

Kim J. H., (2001), Bootstrap-after-Bootstrap Prediction Intervals for Autoregressive Models, *Journal of Business & Economic Statistics*, 19, 117–128.

Lütkepohl H., Krätzig M., (2004), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge.

Lütkepohl H., Staszewska-Bystrova A., Winker P., (2015*a*), Comparison of Methods for Constructing Joint Confidence Bands for Impulse Response Functions, *International Journal of Forecasting*, 31, 782–798.

Lütkepohl H., Staszewska-Bystrova A., Winker P., (2015*b*), Confidence Bands for Impulse Responses: Bonferroni versus Wald, *Oxford Bulletin of Economics and Statistics*, 77, 800–821.

Masarotto G. (1990), Bootstrap Prediction Intervals for Autoregressions, *International Journal of Forecasting* 6, 229–239.

Montiel Olea J. L., Plagborg-Møller M., (2017), Simultaneous Confidence Bands: Theoretical Comparisons and Suggestions for Practice, *Working Paper*.

Roy A., Fuller W. A., (2001), Estimation for Autoregressive Time Series with a Root Near One, *Journal of Business & Economic Statistics*, 19, 482–493.

Staszewska-Bystrova A., (2011), Bootstrap Prediction Bands for Forecast Paths from Vector Autoregressive Models, *Journal of Forecasting*, 30, 721–735.

Staszewska-Bystrova A., (2013), Modified Scheffè's Prediction Bands, *Jahrbücher für Nationalökonomie und Statistik*, 233, 680–690.

Staszewska-Bystrova A., Winker P., (2013), Constructing Narrowest Pathwise Bootstrap Prediction Bands using Threshold Accepting, *International Journal of Forecasting*, 29, 221–233.

Stine R. A., (1987), Estimating Properties of Autoregressive Forecasts, *Journal of the American Statistical Association,* 82, 1072–1078.

Thombs L. A., Schucany W. R., (1990), Bootstrap Prediction Intervals for Autoregression, *Journal of the American Statistical Association*, 85,. 486–492.

Wolf M., Wunderli D., (2015), Bootstrap Joint Prediction Regions, *Journal of Time Series Analysis*, 36, 352–376.

APPENDIX

A1.

Table RESULTS FOR DGP.A WITH $\alpha = 0.5$ AND X^2-DISTRIBUTED ERRORS

H	*B*	sup-*t*	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 100}$			
4	90.54 (5.28)	88.85 (5.34)	89.16 (5.08)	88.93 (5.05)	NA	89.83 (5.18)
8	90.09 (6.23)	88.07 (6.63)	88.43 (6.01)	88.18 (5.98)	88.05 (5.96)	88.99 (6.10)
12	89.14 (6.67)	87.45 (7.40)	87.47 (6.46)	87.22 (6.43)	87.09 (6.41)	88.09 (6.56)

A1. (cont.)

Table RESULTS FOR DGP.A WITH $\alpha = 0.5$ AND X^2-DISTRIBUTED ERRORS (cont.)

H	*B*	sup-*t*	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 400}$			
4	91.45 (5.01)	89.73 (5.15)	90.19 (4.86)	89.97 (4.84)	NA	90.86 (4.93)
8	90.06 (6.82)	89.78 (6.48)	88.47 (6.43)	88.35 (6.40)	88.28 (6.39)	90.67 (5.75)
12	91.40 (6.32)	89.43 (7.14)	89.65 (6.10)	89.51 (6.08)	89.41 (6.07)	90.32 (6.18)

A2.

Table RESULTS FOR DGP.A WITH $\alpha = 0.9$ AND X^2-DISTRIBUTED ERRORS

H	*B*	sup-*t*	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 100}$			
4	92.01 (6.97)	89.43 (6.55)	89.25 (6.39)	88.84 (6.32)	NA	91.37 (6.83)
8	92.14 (10.02)	88.98 (9.34)	88.65 (8.95)	88.01 (8.81)	87.74 (8.75)	91.60 (9.79)
12	91.87 (12.27)	88.57 (11.37)	88.18 (10.87)	87.49 (10.68)	87.15 (10.59)	91.44 (12.01)
			$\boldsymbol{T = 400}$			
4	92.92 (6.42)	89.95 (5.94)	90.26 (5.98)	89.87 (5.92)	NA	92.45 (6.30)
8	93.89 (8.78)	89.93 (8.15)	90.48 (8.05)	89.92 (7.95)	89.68 (7.91)	93.39 (8.62)
12	94.17 (10.30)	89.92 (9.58)	90.59 (9.41)	89.95 (9.29)	89.65 (9.24)	93.66 (10.11)

A3.

Table RESULTS FOR DGP.A WITH $\alpha = 0.95$ AND X^2-DISTRIBUTED ERRORS

H	*B*	sup-*t*	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 100}$			
4	92.25 (7.15)	89.66 (6.66)	89.38 (6.52)	88.96 (6.45)	NA	91.63 (6.99)
8	92.46 (10.62)	89.03 (9.76)	88.69 (9.38)	88.04 (9.23)	87.75 (9.17)	91.94 (10.36)
12	92.17 (13.31)	88.24 (12.15)	88.16 (11.64)	87.38 (11.42)	87.03 (11.32)	91.78 (13.00)
			$\boldsymbol{T = 400}$			
4	93.25 (6.71)	89.98 (6.13)	90.37 (6.21)	89.95 (6.15)	NA	92.77 (6.59)
8	94.46 (9.66)	89.99 (8.75)	90.79 (8.74)	90.15 (8.63)	89.86 (8.57)	94.06 (9.47)
12	94.91 (11.81)	90.01 (10.62)	90.98 (10.62)	90.24 (10.45)	89.91 (10.38)	94.52 (11.58)

A4.

Table RESULTS FOR DGP.A WITH $\alpha = 0.5$ AND *t*-DISTRIBUTED ERRORS

H	*B*	sup-*t*	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 100}$			
4	88.69 (5.71)	88.87 (5.43)	87.51 (5.44)	87.31 (5.40)	NA	88.04 (5.56)
8	86.77 (7.03)	88.00 (6.76)	85.58 (6.82)	85.43 (6.79)	85.34 (6.78)	85.58 (6.78)
12	84.40 (7.52)	87.16 (7.68)	83.13 (7.33)	82.97 (7.31)	82.88 (7.29)	83.32 (7.34)
			$\boldsymbol{T = 400}$			
4	90.68 (5.47)	90.03 (5.25)	89.38 (5.22)	89.21 (5.19)	NA	90.12 (5.37)
8	90.06 (6.82)	89.78 (6.48)	88.47 (6.43)	88.35 (6.40)	88.28 (6.39)	89.25 (6.63)
12	89.36 (7.71)	89.56 (7.26)	87.85 (7.25)	87.71 (7.22)	87.64 (7.20)	88.35 (7.41)

A5.

Table RESULTS FOR DGP.A WITH $\alpha = 0.9$ AND *t*-DISTRIBUTED ERRORS

H	*B*	sup-*t*	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 100}$			
4	91.29 (7.39)	89.62 (6.62)	88.42 (6.60)	88.06 (6.53)	NA	90.67 (7.20)
8	91.31 (10.43)	89.14 (9.35)	87.71 (9.23)	87.13 (9.09)	86.89 (9.03)	90.72 (10.16)
12	90.81 (12.58)	88.56 (11.35)	87.08 (11.10)	86.39 (10.92)	86.09 (10.83)	90.42 (12.29)
			$\boldsymbol{T = 400}$			
4	92.93 (6.97)	90.10 (6.12)	89.74 (6.14)	89.42 (6.07)	NA	92.53 (6.83)
8	93.67 (9.93)	90.19 (8.35)	89.82 (8.48)	89.32 (8.35)	89.11 (8.30)	93.10 (9.65)
12	93.73 (11.64)	90.15 (9.80)	89.93 (10.06)	89.36 (9.89)	89.10 (9.82)	93.11 (11.35)

A6.

Table RESULTS FOR DGP.A WITH $\alpha = 0.95$ AND *t*-DISTRIBUTED ERRORS

H	*B*	sup-*t*	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 100}$			
4	91.61 (7.62)	89.77 (6.77)	88.62 (6.75)	88.24 (6.67)	NA	90.98 (7.42)
8	91.75 (11.10)	89.21 (9.86)	87.86 (9.68)	87.26 (9.52)	87.01 (9.46)	91.19 (10.80)
12	91.30 (13.71)	88.41 (12.25)	87.24 (11.91)	86.46 (11.68)	86.11 (11.58)	90.90 (13.38)

A6. (cont.)

Table RESULTS FOR DGP.A WITH $\alpha = 0.95$ AND t-DISTRIBUTED ERRORS (cont.)

H	B	sup-t	B^2	B^3	B^4	B^{TA}
			$\boldsymbol{T = 400}$			
4	93.25 (7.29)	90.11 (6.32)	89.81 (6.35)	89.47 (6.28)	NA	92.88 (7.14)
8	94.26 (10.86)	90.21 (8.94)	90.04 (9.12)	89.48 (8.97)	89.23 (8.90)	93.75 (10.56)
12	94.56 (13.19)	90.22 (10.83)	90.34 (11.19)	89.66 (10.97)	89.36 (10.88)	94.02 (12.86)

ZMODYFIKOWANE PASMA PREDYJKCYJNE BONFERRONIEGO DLA MODELI AUTOREGRESYJNYCH

Streszczenie

*Pasma predykcyjne konstruuje się często z użyciem nierówności Bonferroniego. Wadą takich pasm może być ich duża rozpiętość i zawyżone prawdopodobieństwo zawierania przyszłej trajektorii prognozowanej zmiennej. W artykule zaproponowano dwie poprawki dla metody konstrukcji bootstrapowych pasm predykcyjnych Bonferroniego wykorzystujące nierówności wyższego rzędu i procedurę minimalizacji szerokości pasma. Metody zastosowano do prognozowania jednowymiarowych procesów autoregresyjnych. Ich właściwości zbadano za pomocą eksperymentów Monte Carlo. Wykazano, że zaproponowane procedury prowadzą, w wielku przypadkach, do uzyskania stosunkowo wąskich pasm predykcyjnych o odpo*wiednich prawdopodobieństwach zawierania przyszłej trajektorii zmiennej.

Słowa kluczowe: pasmo predykcyjne, proces autoregresyjny, nierówność Bonferroniego

REFINED BONFERRONI PREDICTION BANDS FOR AUTOREGRESSIVE MODELS

Abstract

Joint prediction bands are often constructed using Bonferroni's inequality. The drawback of such bands may be their large width and excessive coverage probability. The paper proposes two refinements to the basic Bonferroni method of constructing bootstrap prediction bands. These are based on higher order inequalities and optimization of the width of the band. The procedures are applied to the problem of predicting univariate autoregressive processes. Their properties are studied by means of Monte Carlo experiments. It is shown that the proposed methods lead, in many scenarios, to obtaining relatively narrow prediction bands with desired coverage probabilities.

Keywords: prediction band, autoregressive process, Bonferroni's inequality

DOI: 10.5604/01.3001.0014.0533

PRZEGLĄD STATYSTYCZNY
TOM LXV – ZESZYT 2 – 2018

Piotr KĘBŁOWSKI[1]

A Monte Carlo comparison of LCCA- and ML-based cointegration tests for panel var process with cross-sectional cointegrating vectors[2]

1. INTRODUCTION

Cointegration analysis of non-stationary panel data is typically based on the univariate framework, see Phillips, Moon (1999), where the OLS-based cointegration tests developed firstly by Kao (1999), Pedroni (1999) and McCoskey, Kao (1998) are considered and the DOLS or the FMOLS estimator is employed. Assuming lack of any long-run cross-sectional dependencies, this approach shall work reasonably well for small homogenous systems with one cointegrating vector. However, this will not be valid when the long-run cross-sectional dependencies occurs, for example due to cross-sectional cointegrating vector, see Banerjee et al. (2004) and Jacobson et al. (2008), as well as in case of medium- or large-sized systems with a number of cointegration vectors. Clearly, the panel VAR (PVAR) model proposed by Larsson, Lyhagen (2007) should be considered here, even though in case of lack of long-run cross-sectional dependencies also the global VAR (GVAR) model can be applied, see Pesaran et al. (2004).

The multivariate cointegration analysis of panel data was considered first by Groen, Kleibergen (2003) and Larsson, Lyhagen (2007), who advocated use of VAR models for analysis of non-stationary panel data (see also Larsson, Lyhagen, 2000; Larsson et al., 2001 and Jacobson et al., 2008). Moreover, in the case of panel VAR framework, Anderson et al. (2006) suggested the use of levels canonical correlation analysis (LCCA), as proposed by Box, Tiao (1977) and Bewley et al. (1994), instead of maximum likelihood (ML) estimation proposed by Johansen (1988). Anderson et al. (2006) highlight that they were able to find an additional cross-sectional cointegrating vector in empirical data using Box

[1] University of Lodz, Faculty of Economics and Sociology, Chair of Econometric Models and Forecasts, 41 Rewolucji 1905 r. St, 90–214, Lodz, Poland, email:piotr.keblowski@uni.lodz.pl.

[2] Financial support from the National Science Centre in Poland through the research grant D/HS4/01767 is gratefully acknowledged.

and Tiao approach and the LCCA-based cointegration rank test, as opposed to the results suggested by Johansen's tests.

The aforementioned finding of Anderson et al. (2006) provides a rationale for investigating small sample properties of cointegration rank tests for both Box and Tiao as well as Johansen approach in the framework of panel VAR process with cross-sectional cointegrating vectors. To this end, performance of the min-root and the trace test of Yang, Bewley (1996) are compared with the widely used ML-counterparts derived by Johansen (1988) – the max-root and the trace test.

Since the cointegration rank tests suffer from severe size distortions in small samples, see Johansen (2002), the bootstrap tests are considered. The algorithm of the bootstrap cointegration rank test is similar to algorithms proposed by van Giersbergen (1996) and Svensen (2006), and the asymptotic theory of the bootstrap method was given by Svensen (2006) and Cavaliere et al. (2012). Note however, that even when the bootstrap cointegration rank test or size-corrected test like the test with Bartlett correction is used, some significant size distortions can still occur, especially for high dimensional process such as panel VAR.

The comparative investigations based on the purely time-series context are available in Bewley et al. (1994) and Bewley, Yang (1995). Moreover, performance of canonical correlation estimators of cointegrating vectors for panel VAR models were investigated by Kębłowski (2016).

2. PANEL VAR MODEL, CANONICAL CORRELATION ANALYSIS AND COINTEGRATION TESTS

Consider two main strands of model's specification for the panel VAR/VEC process. The first one would be the straightforward panel augmentation of the VEC model for the double-indexed processes

$$\Delta \mathbf{y}_{it} = \mathbf{\Pi}_i \mathbf{y}_{i,t-1} + \sum_{k=1}^{K-1} \mathbf{\Gamma}_{ki} \Delta \mathbf{y}_{i,t-k} + \mathbf{\Phi}_i \mathbf{d}_t + \boldsymbol{\varepsilon}_{it}, \tag{1}$$

where $\mathbf{y}_{it} = [y_{1it}\ y_{2it}\ \ldots\ y_{Pit}]'$ is a *P*-dimensional vector of observations for given cross-section i and period t, $\mathbf{\Pi}_i$, and $\mathbf{\Gamma}_{ki}$ are $P \times P$ matrices of coefficient, $\mathbf{d}_t$ and $\mathbf{\Phi}_i$ denote a *N*-dimensional vector of (common) deterministic components and $P \times N$ matrix of their coefficients and $\boldsymbol{\varepsilon}_{it}$ is a *P*-dimensional independently and identically distributed error term with mean equal to zero and covariance matrix $\mathbf{\Omega}_i$ for cross-section i.

Consider next the following VEC model for the panel VAR process

$$\Delta \mathbf{y}_t = \mathbf{\Pi} \mathbf{y}_{t-1} + \sum_{k=1}^{K-1} \mathbf{\Gamma}_k \Delta \mathbf{y}_{t-k} + \mathbf{\Phi} \mathbf{d}_t + \boldsymbol{\varepsilon}_t, \tag{2}$$

where $\mathbf{y}_t = [\mathbf{y}_{1t}' \ \mathbf{y}_{2t}' \ \ldots \ \mathbf{y}_{It}']'$ is a *IP*-dimensional vector of observations for period t, $\mathbf{\Pi}$ and $\mathbf{\Gamma}_k$ are $IP \times IP$ matrices of coefficient, $\mathbf{\Phi}$ denotes $IP \times N$ matrix of deterministic term coefficients and $\boldsymbol{\varepsilon}_t$ is a *IP*-dimensional independently and identically distributed error term with mean equal to zero and covariance matrix $\mathbf{\Omega}$.

In case of the cointegrated panel VAR process matrix $\mathbf{\Pi}$ can be decomposed into $IP \times IR$ full rank matrices $\mathbf{A}$ and $\mathbf{B}$, and the panel VEC model is

$$\Delta \mathbf{y}_i = \mathbf{A}\mathbf{B}' \mathbf{y}_{t-1} + \sum_{k=1}^{K-1} \mathbf{\Gamma}_k \Delta \mathbf{y}_{t-k} + \mathbf{\Phi} \mathbf{d}_t + \boldsymbol{\varepsilon}_t. \tag{3}$$

Clearly, model (1) assumes lack of any cross-sectional dependencies, whereas model (2) allows for both short- and long-run cross-sectional dependencies, since $\mathbf{A}$, $\mathbf{B}$, $\mathbf{\Gamma}_k$ and $\mathbf{\Omega}$ matrices are not assumed to be block-diagonal (even though in practice $\mathbf{B}$ is most often assumed to be block-diagonal, see Larsson, Lyhagen, 2007). This allows for four different sources of cross-sectional dependence: in the error term, in the short-run dynamics, in the adjustments to the long-run equilibrium and in the cointegration space, as opposed to the assumptions of model (1), and is the main rationale for using model (2) instead of model (1). The only, but significant, disadvantage of using model (2) is the potential dimensionality effect that can limit its application for small samples in case of a large number of cross-sections and variables simultaneously.

The levels canonical correlation analysis of Box and Tiao is performed as follows. At first, the short-run effects are concentrated out and the concentrated regression is

$$\tilde{\mathbf{y}}_t = \mathbf{\Theta} \tilde{\mathbf{y}}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{4}$$

where $\tilde{\mathbf{y}}_{t-1} = \mathbf{y}_t - \mathbf{y}_t \mathbf{z}_t'(\mathbf{z}_t \mathbf{z}_t')^{-1} \mathbf{z}_t$, $\tilde{\mathbf{y}}_{t-1} = \mathbf{y}_{t-1} - \mathbf{y}_{t-1} \mathbf{z}_t'(\mathbf{z}_t \mathbf{z}_t')^{-1} \mathbf{z}_t$
and $\boldsymbol{z}_t = [\Delta \boldsymbol{y}_{t-1}' \ \ldots \ \Delta \mathbf{y}_{t-K+1}' \mathbf{d}_t']'$.

Next, the canonical transformation is achieved by solving the eigenvalue problem

$$\left| \lambda \tilde{\mathbf{Y}} \tilde{\mathbf{Y}}' - \left(\tilde{\mathbf{Y}} \tilde{\mathbf{Y}}_{-1} \right) \left(\tilde{\mathbf{Y}}_{-1} \tilde{\mathbf{Y}}_{-1} \right)^{-1} \left(\tilde{\mathbf{Y}}_{-1} \tilde{\mathbf{Y}} \right) \right| = 0 \tag{5}$$

for the eigenvalues $0 < \hat{\lambda}_1 < \cdots < \hat{\lambda}_{IP} < 1$ and eigenvectors $\hat{\mathbf{V}} = [\hat{\mathbf{v}}_1 \ \ldots \ \hat{\mathbf{v}}_{IP}]$, of which the first IR constitute the cointegration subspace.

To compute the cointegration tests of Yang, Bewley (1996), the following series are calculated

$$\hat{\mathbf{y}}_{t-1} = \widehat{\mathbf{A}}'_{\perp}\hat{\mathbf{y}}_{t-1} \tag{6a}$$

and

$$\hat{\mathbf{y}}_t = \hat{\mathbf{y}}_{t-1} + \widehat{\mathbf{A}}'_{\perp}\hat{\mathbf{C}}\mathbf{e}_t, \tag{6b}$$

where $\mathbf{e}_t$ denotes residuals from concentrating out the short-run effects in (4) and $\hat{\mathbf{C}}$ is the impact matrix from the moving average representation, $\hat{\mathbf{C}} = \widehat{\mathbf{B}}'_{\perp}\left(\widehat{\mathbf{A}}_{\perp}'\hat{\mathbf{\Gamma}}\widehat{\mathbf{B}}_{\perp}\right)^{-1}\widehat{\mathbf{A}}_{\perp}$ and $\hat{\mathbf{\Gamma}} = \mathbf{I} - \sum_{k=1}^{K-1}\widehat{\mathbf{\Gamma}}_k$.

Then the following eigenvalue problem is solved

$$\left|\mu\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}' - \left(\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}_{-1}'\right)\left(\widetilde{\mathbf{Y}}_{-1}\widetilde{\mathbf{Y}}_{-1}'\right)^{-1}\left(\widetilde{\mathbf{Y}}_{-1}\widetilde{\mathbf{Y}}'\right)\right| = 0 \tag{7}$$

for the eigenvalues $0 < \hat{\mu}_1 < \cdots < \hat{\mu}_{I(P-R)} < 1$. The LCCA-based cointegration rank tests of the null $H_0\colon \mathrm{rank}\,(\mathbf{\Pi}) = IR$ vs. the alternative $H_1\colon \mathrm{rank}\,(\mathbf{\Pi}) = IR$ are calculated as the minimum-root test $minroot^{LCCA} = T(1-\hat{\mu}_1)$ and the trace test $trace^{LCCA} = T\sum_{j=1}^{I(P-R)}\left(1-\hat{\mu}_j\right)$. Both tests diverge under the alternative, see Yang, Bewley (1996).

Similarly, the canonical correlation analysis of differences and lagged levels, derived by Johansen (1988), is calculated by concentrating out at first the short-run effects, thus the concentrated regression is

$$\Delta\hat{\mathbf{y}}_t = \mathbf{A}\mathbf{B}'\tilde{\mathbf{y}}_{t-1}. \tag{8}$$

Then the canonical transformation is performed by solving

$$\left|\lambda\widetilde{\mathbf{Y}}_{-1}\widetilde{\mathbf{Y}}_{-1}{}' - \left(\widetilde{\mathbf{Y}}_{-1}\Delta\widetilde{\mathbf{Y}}'\right)\left(\Delta\widetilde{\mathbf{Y}}\Delta\widetilde{\mathbf{Y}}'\right)^{-1}\left(\Delta\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}_{-1}'\right)\right| = 0 \tag{9}$$

for the eigenvalues $1 > \hat{\lambda}_1 < \cdots < \hat{\lambda}_{IP} < 0$ and $\widehat{\mathbf{B}} = [\hat{\mathbf{v}}_1 \ldots \hat{\mathbf{v}}_{IR}]$. The cointegration tests of the same hypotheses as in Yang, Bewley (1996) are calculated as the maximum-root test $maxroot^{LCCA} = -T\ln\left(1-\hat{\lambda}_{IR+1}\right)$ and the trace test $race = -T\sum_{j=IR+1}^{IP}\ln\left(1-\hat{\lambda}_j\right)$.

3. DESIGN OF EXPERIMENT AND RESULTS

Monte Carlo simulation is used to compare performance of the LCCA-based and the ML-based cointegration rank tests within the framework of second-order panel VEC model (3) with five variables for each cross section $(P = 5)$ and a constant restricted to the cointegration space

$$\Delta\hat{\mathbf{y}}_t = \mathbf{AB}'[\mathbf{y}'_{t-1}\mathbf{j}']' + \mathbf{\Gamma}_1\Delta\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{10}$$

where short- and long-run cross-sectional dependencies are allowed and the error term $\boldsymbol{\varepsilon}_t$ comes from the multivariate normal distribution, $\boldsymbol{\varepsilon}_t \backsim N_{IP}(\mathbf{0};\mathbf{\Omega})$, with covariance matrix from the inverse Wishart distribution – $\backsim W_{IP}^{-1}(\mathbf{I};100)$.

The results for the case with an unrestricted constant are not reported here (available upon request), since they do not alter the conclusions drawn from the former case. Two cointegrating vectors for each cross-section are imposed plus an additional cross-sectional cointegrating vector describing a homogeneous long-run relationship between the fifth variable of each cross-section

$$\mathbf{B} = \begin{bmatrix} \begin{matrix}1 & -1 & 0 & 0\\ 0 & 0 & 1 & -1\end{matrix} & \mathbf{0} & \cdots & \mathbf{0} & \begin{matrix}1\\1\end{matrix} \\ \mathbf{0} & \begin{matrix}1 & -1 & 0 & 0\\ 0 & 0 & 1 & -1\end{matrix} & \cdots & \mathbf{0} & \begin{matrix}1\\1\end{matrix} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \begin{matrix}1 & -1 & 0 & 0\\ 0 & 0 & 1 & -1\end{matrix} & \begin{matrix}1\\1\end{matrix} \\ \begin{matrix}0 & 0 & 0 & 0 & 1\end{matrix} & \begin{matrix}0 & 0 & 0 & 0 & \beta\end{matrix} & \cdots & \begin{matrix}0 & 0 & 0 & 0 & \beta\end{matrix} & 1 \end{bmatrix}, \tag{11}$$

where $\beta = -\big(1/(I-1)\big)$. The loadings matrix allows for the long-run cross--sectional adjustments

$$\mathbf{A} = \begin{bmatrix} -0.5 & 0 & 0 & 0 & 0 & \ldots & -0.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.5 & 0 & 0 & & 0 & 0 & -0.1 & 0 & 0 \\ & & \vdots & & & & & & \vdots & & \\ -0.1 & 0 & 0 & 0 & 0 & \ldots & -0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.1 & 0 & 0 & & 0 & 0 & -0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.5 & \ldots & 0 & 0 & 0 & 0 & -0.5 \end{bmatrix}. \tag{12}$$

The cross-sectional dependence in the short-run adjustments is allowed, since $\underset{i=j}{\forall}\gamma_{ij,1} = 0.5$ and $\underset{i\neq j}{\forall}\gamma_{ij,1}\sim U(-0.1,0.1)$. The roots of the autoregressive polynomial are computed in order to exclude explosive roots in the DGP. The initial values comes from the multivariate normal distribution and the first hundred observations are removed. The number of cross-sections varies from two to six and the sample size is $T \in \{100, 200, 400, 800\}$. The number of replications for the Monte Carlo simulation is 10000.

The algorithm of the bootstrap method for cointegration rank tests is similar to algorithms proposed by van Giersbergen (1996) and Svensen (2006), and it is as follows:

Step 1. Estimate (2) by means of LCCA, for $minroot^{LCCA}$ and $trace^{LCCA}$ tests, or ML for $maxroot$ and $trace$ tests, and calculate a realization of the test statistic.

Step 2. Estimate the parameters under the null – (3) and check whether the roots of the autoregressive polynomial are equal to 1 or lie outside the unit circle.

Step 3. Using (3) compute recursively bootstrap sample using sampled residuals drawn with replacement from the estimated residuals.

Step 4. Calculate bootstrap realization of the test statistic.

Step 5. Repeat steps 3 to 4 a large number of times.

Step 6. Reject the null $H_0: \operatorname{rank}(\mathbf{\Pi}) = IR$ if the test statistic from step (1) is larger than the critical value from the bootstrap distribution.

The number of replications for the bootstrap method is 1000.

The relative frequencies of rejecting the false null $H_0: \operatorname{rank}(\mathbf{\Pi}) = IR$ and the true null $H_0: \operatorname{rank}(\mathbf{\Pi}) = IR + 1$ of bootstrap cointegration tests for PVAR model (10) are contained in tables 1–2 respectively. Comparison of the empirical size of the tests reveals that the ML-based cointegration rank tests $maxroot$ and $trace$ are undersized in small samples, as compared to 5% nominal size, whereas the LCCA-based cointegration rank tests $minroot^{LCCA}$ and $trace^{LCCA}$ are oversized. The size distortion becomes negligible in general only for a quite long samples of at least 400 observations within time dimension, even though for PVAR models with moderate dimension $(IP \leq 20)$ and 200 observations it seems to be still under control.

Table 1. EMPIRICAL SIZE OF BOOTSTRAP COINTEGRATION TESTS FOR PANEL VAR PROCESS WITH CROSS-SECTIONAL COINTEGRATING VECTOR

I	*T*	ML		LCCA	
		maxroot	*trace*	*minroot*	*trace*
2	100	0.025	0.045	0.124	0.134
	200	0.058	0.081	0.065	0.057
	400	0.057	0.088	0.049	0.037
	800	0.056	0.053	0.043	0.033
3	100	0.000	0.008	0.306	0.308
	200	0.046	0.050	0.108	0.082
	400	0.057	0.056	0.052	0.039
	800	0.052	0.064	0.036	0.020
4	100	0.000	0.001	0.240	0.777
	200	0.021	0.036	0.090	0.079
	400	0.040	0.057	0.061	0.039
	800	0.048	0.048	0.029	0.015
5	100	0.000	0.000	0.334	0.959
	200	0.000	0.012	0.217	0.226
	400	0.052	0.064	0.066	0.052
	800	0.066	0.073	0.054	0.030
6	100	—	—	—	—
	200	0.000	0.002	0.225	0.557
	400	0.049	0.065	0.075	0.065
	800	0.048	0.063	0.046	0.026

Table 2. PERFORMANCE OF BOOTSTRAP COINTEGRATION TESTS FOR PANEL VAR PROCESS WITH CROSS-SECTIONAL COINTEGRATING VECTOR

I	*T*	ML		LCCA	
		maxroot	*trace*	*minroot*	*trace*
2	100	0.540	0.432	0.280	0.320
	200	1.000	0.999	0.619	0.495
	400	1.000	1.000	1.000	1.000
	800	1.000	1.000	1.000	1.000
3	100	0.002	0.026	0.447	0.540
	200	0.948	0.735	0.431	0.486
	400	1.000	1.000	0.987	0.965
	800	1.000	1.000	0.997	0.987
4	100	0.000	0.002	0.354	0.875
	200	0.298	0.293	0.605	0.559
	400	1.000	0.999	0.989	0.922
	800	1.000	1.000	0.992	0.957
5	100	0.000	0.000	0.386	0.969
	200	0.004	0.074	0.798	0.740
	400	1.000	0.915	0.990	0.899
	800	1.000	1.000	0.998	0.955
6	100	—	—	—	—
	200	0.000	0.017	0.546	0.871
	400	0.990	0.638	0.992	0.880
	800	1.000	1.000	0.993	0.931

With respect to performance of both group of cointegration rank tests it can be seen that both the ML-based approach as well as the LCCA-based approach perform very poorly in short samples, as compared to the actual size of each test 0.05. According to the results, it is very likely that in a very short samples of 100 observations in time dimension and less, the LCCA-based cointegration rank tests will indicate at an additional cross-sectional cointegrating vector, the ML-based counterparts will lead to an opposite conclusion, all this regardless of the actual existence of the cross-sectional cointegrating vector.

Moreover, even though the performance of the tests is not size-adjusted, it can be easily noted that the ML-based cointegration rank tests perform in general better than their LCCA-based counterparts in detecting an additional cross-sectional cointegrating vector within a PVAR framework, see the results for $T = 200$ and $I = 2, 3$ for example. As expected, in the case of long panel data with 400 observations within time dimension and more, performance of both approaches is close to unity and thus comparable.

The results of Monte Carlo investigation on small sample properties of bootstrap cointegration rank tests for the PVAR model clearly suggest that the tests based on levels canonical correlation analysis are in general outperformed by the maximum likelihood cointegration rank tests, if the dynamic properties of the underlying process are properly specified. Moreover, the higher the number of

cross-sections is, the more observations within time dimension is needed in order to unambiguously infer on the cointegration rank for the PVAR model. If the dimension of the PVAR model IP exceeds 15 then as many as 400 observations can be needed in order to properly identify an additional cross-sectional cointegrating vector using the ML-based approach.

Clearly, the results show that the application of the bootstrap panel cointegrating rank test for the PVAR model given by (2) is in practice limited to the panels with few cross-sections, which makes the application of the tests for macroeconomic panels with larger cross-sectional dimension impossible in fact. This is an important drawback of the approach based on the PVAR model.

4. CONCLUSIONS

In this paper we have examined small sample properties of the bootstrap cointegration rank tests for the unrestricted panel VAR model when short- and long-run cross-sectional dependencies occur in the underlying process generating non-stationary panel data. Two basic frameworks were employed: levels canonical correlation analysis and maximum likelihood estimation. The results shows that the bootstrap cointegration rank tests for the panel VAR model suffer from severe size distortions in small samples, with downward bias of the ML-based tests and the upward bias of the LCCA-based tests. Weak performance of both approaches is observed for a very short sample of about 100 observations within time dimension. As a result, the LCCA-based cointegration rank tests will easily indicate at an additional cross-sectional cointegrating vector, the ML-based counterparts can lead to an opposite conclusion, all this regardless of the actual existence of the additional cointegrating vector. Moreover, it was found that the bootstrap cointegration rank tests for panel VAR model based on levels canonical correlation analysis are in general outperformed by the maximum likelihood cointegration rank tests.

The results of the investigation indicate that the bootstrap ML-based cointegration rank tests perform quite well with respect to size distortion and power for small- and medium-sized PVAR models $(IP < 20)$, if there are at least 200 observations within time dimension. Whereas in case of large-sized PVAR models, say $(IP \geq 20)$, long panel data with about 400 observations within time dimension are necessary.

REFERENCES

Anderson R., Qian H., Rasche R., (2006), Analysis of Panel Vector Error Correction Models Using Maximum Likelihood, the Bootstrap, and Canonical-Correlation Estimators, Federal Reserve Bank of St. Louis Working Paper Series, Working Paper, 2006-050A.

Banerjee A., Marcelino M., Osbat C., (2004), Some Cautions on the Use of Panel Methods for Integrated Series of Macroeconomic Data, *Econometrics Journal*, 7, 322–340.

Bewley R., Orden D., Yang M., Fisher L. A., (1994), Comparison of Box-Tiao and Johansen Canonical Estimators of Cointegrating Vectors in VEC(1) Models, *Journal of Econometrics*, 64, 3–27.

Bewley R., Yang M., (1995), Tests for Cointegration Based on Canonical Correlation Analysis, *Journal of the American Statistical Association*, 90, 990–996.

Box G. E. P., Tiao G. C., (1977), A Canonical Analysis of Multiple Time Series, *Biometrika*, 64, 355––365.

Cavaliere G., Rahbek A., Taylor M. R., (2012), Bootstrap Determination of the Co-integration Rank in Vector Autoregressive Models, *Econometrica*, 80, 1721–1740.

Groen J. J. J., Kleibergen F., (2003), Likelihood-Based Cointegration Analysis in Panels of Vector Error Correction Models, *Journal of Business and Economics Statistics*, 21, 295–318.

Jacobson T., Lyhagen J., Larsson R., Nessén M., (2008), Inflation, Exchange Rates and PPP in a Multivariate Panel Cointegration Model, *Econometrics Journal*, 11, 58–79.

Johansen S., (1988), Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control*, 12, 231–254.

Johansen S., (2002), A Small Sample Correction for the Test of Cointegrating Rank in the Vector Autoregressive Model, *Econometrica*, 70, 1929–1961.

Kao C., (1999), Spurious Regression and Residual-based Tests for Cointegration in Panel Data, *Journal of Econometrics*, 90, 1–44.

Kębłowski P., (2016), Canonical Correlation Analysis in Panel Vector Error Correction Model. Performance Comparison, *Central European Journal of Economic Modelling and Econometrics*, 4, 203–217.

Larsson R., Lyhagen J., (2000), Testing for Common Cointegrating Rank in Dynamic Panels, Stockholm School of Economics Working Paper Series in Economics and Finance, 378.

Larsson R., Lyhagen J., (2007), Inference in Panel Cointegration Models with Long Panels, *Journal of Business and Economic Statistics*, 25, 473–483.

Larsson R., Lyhagen J., Löthgren M., (2001), Likelihood-based Cointegration Tests in Heterogeneous Panels, *Econometrics Journal*, 4, 109–142.

McCoskey S., Kao C., (1998), A Residual-Based Test of the Null of Cointegration in Panel Data, *Econometric Reviews*, 17, 57–84.

Pedroni P., (1999), Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors, *Oxford Bulletin of Economics and Statistics*, 61, 653–670.

Pesaran M. H., Schuermann T., Weiner S.M., (2004), Modeling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model, *Journal of Business and Economic Statistics*, 22, 129–162.

Phillips P. C. B., Moon H. R., (1999), Linear Regression Limit Theory for Nonstationary Panel Data, *Econometrica*, 67, 1057–1111.

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

van Giersbergen N. P. A., (1996), Bootstraping the Trace Statistic in VAR Models: Monte Carlo Results and Applications, *Oxford Bulletin of Economics and Statistics*, 58, 391–408.

Yang M., Bewley R., (1996), On Cointegration Tests for VAR Models with Drift, *Economics Letters*, 51, 45–50.

ANALIZA MONTE CARLO WŁASNOŚCI TESTÓW KOINTEGRACJI DLA PANELOWEGO PROCESU VAR Z MIĘDZYPRZEKROJOWYMI WEKTORAMI KOINTEGRUJĄCYMI

Streszczenie

W artykule przedstawiono wyniki badania własności bootstrapowych testów kointegracji dla panelowego procesu VAR z międzyprzekrojowymi wektorami kointegrującymi. Wyniki badania wskazują, że bootstrapowe testy kointegracji dla modelu PVAR, które oparte są na analizie korelacji kanonicznej poziomów, cechują się przeszacowaniem rozmiaru testu, z kolei bootstrapowe testy kointegracji dla modelu PVAR wywiedzione z metody największej wiarygodności charakteryzują się zwykle niedoszacowaniem rozmiaru testu. Wykazano również, że bootstrapowe testy kointegracji dla modelu PVAR wywiedzione z metody największej wiarygodności cechują się zwykle lepszymi własnościami ze względu na moc testu. Wyniki badania wskazują, że własności bootstrapowych testów kointegracji dla modelu PVAR wywiedzionych z metody największej wiarygodności cechują się satysfakcjonującymi własnościami małopróbkowymi dla małowymiarowych modeli PVAR z ograniczoną liczbą przekroi.

Słowa kluczowe: międzyprzekrojowe wektory kointegrujące, analiza korelacji kanonicznej, testy kointegracji, panelowy model VAR, procedura Boxa i Tiao

A MONTE CARLO COMPARISON OF LCCA- AND ML-BASED COINTEGRATION TESTS FOR PANEL VAR PROCESS WITH CROSS-SECTIONAL COINTEGRATING VECTORS

Abstract

Small-sample properties of bootstrap cointegration rank tests for unrestricted panel VAR process are considered when long-run cross-sectional dependencies occur. It is shown that the bootstrap cointegration rank tests for the panel VAR model based on levels canonical correlation analysis are oversized, whereas the bootstrap cointegration rank tests based on maximum likelihood framework are undersized. Moreover, the former tests are in general outperformed by the latter in terms of performance. The results of the investigation indicate that the ML-based bootstrap cointegration rank tests perform well in small samples for small-sized panel VAR models with a few cross-sections.

Keywords: cross-sectional cointegrating vectors, canonical correlation analysis, cointegration tests, panel VAR model, Box and Tiao approach

DOI: 10.5604/01.3001.0014.0535

PRZEGLĄD STATYSTYCZNY
TOM LXV – ZESZYT 1 – 2018

Krzysztof PIASECKI[1]
Joanna SIWEK[2]

Multi-asset portfolio with trapezoidal fuzzy present values[3]

1. INTRODUCTION

By a financial asset we understand an authorization to receive a future financial revenue, payable to a certain maturity. The value of this revenue is interpreted as anticipated future value (FV) of the asset. According to the uncertainty theory (von Mises, 1962; Kaplan, Barish, 1967), any unknown future state is uncertain. This uncertainty stems from of our lack of knowledge about the future. Yet, in the researched case, we can point out this particular time in the future, in which the considered income value will be already known to the observer. After Kolmogorov (1933, 1956), von Mises (1957), Lambalgen (1996), Sadowski (1976, 1980), Czerwiński (1960, 1969), Caplan (2001) we will accept this as a sufficient condition for modelling the uncertainty with probability. All this leads to a conclusion that FV is a random variable.

The main focus of following research is present value (PV), defined as a present equivalent of a payment available in a given time in the future. PV of future cash flows is widely accepted to be an approximate value, with fuzzy numbers being one of the main tools of its modelling. Ward (1985) defined fuzzy PV as a discounted fuzzy forecast of a future cash flow's value. Fuzzy numbers were introduced into financial arithmetic by Buckley (1987). As a result, Ward's definition was then further generalized by Greenhut et. al (1995), Sheen (2005) and Huang (2007), who expands Ward's definition to the case of a future cash flow given as a fuzzy variable. More general definition of fuzzy PV was proposed by

[1] Poznan University of Economics, Faculty of Management, Department of Investment and Real Estate, 10 al Niepodległości St., 61–875 Poznań, Poland, corresponding author – e-mail: krzysztof.piasecki@ue.poznan.pl.

[2] Adam Mickiewicz University, Faculty of Mathematics and Computer Science, Department of Imprecise Information Processing Methods, 87 Umultowska St., 61–614 Poznań, Poland.

[3] This research was financed by National Science Centre Poland, granted by the decision No. DEC-2015/17/N/HS4/00206.

Tsao (2005), who assumes that future cash flow can be treated as a fuzzy probabilistic set. All those authors depict PV as a discounted, imprecisely estimated future cash flow. A different approach was given by Piasecki (2011, 2014), where fuzzy PV was estimated by a current market value of the financial asset.

Piasecki (2011) showed that if the PV of an asset is a fuzzy real number, then its return rate is a fuzzy probabilistic set (Hirota, 1981). Works of Buckley (1987), Gutierrez (1989), Kuchta (2000) and Lesage (2001) have previously proved the sensibility of using triangular or trapezoidal fuzzy numbers as a fuzzy financial arithmetic tool.

In Siwek (2015) two cases of a simple two-asset portfolio with fuzzy triangular and trapezoidal present value were researched. After Markowitz (1952), both articles assume normal distribution of simple return rates, with fuzzy expected return rate being the main tool of an asset assessment. Nonetheless, the results of performed research were highly complicated in forms of energy and entropy measures for the portfolio expected return rate, which made it difficult to continue researching the topic.

In Piasecki, Siwek (2017) an alternative approach is suggested to solve the problem researched in Siwek (2015). The fuzzy discount factor is used for appraising the financial instrument with triangular fuzzy PV. Regretfully, entropy measure of an arbitrary triangular fuzzy number is constant, which makes it difficult to analyze the impact of the diversification on imprecision of a portfolio assessment. On the other hand, trapezoidal fuzzy numbers do not have this disadvantage. Thus, the main purpose of presented article is to generalize these results to the case where PVs of portfolio assets are given by trapezoidal fuzzy numbers. In addition, in our considerations, the two-asset portfolio will be replaced with a more general multiple asset portfolio. This way we can extend the possibility of managing the risk burdening a multi-asset portfolio, constructed with use of an imprecise information stemming from present value of component assets.

2. ELEMENTS OF FUZZY NUMBER THEORY

By $\mathcal{F}(\mathbb{R})$ we denote a family of all fuzzy subsets of a real line $\mathbb{R}$. Dubois, Prade (1979) define a fuzzy number as a fuzzy subset $K \in \mathcal{F}(\mathbb{R})$, represented by membership function $\mu_K \in [0;1]^{\mathbb{R}}$ which[4] satisfies following conditions

$$\exists_{x\in\mathbb{R}}\, \mu_K(x) = 1, \tag{1}$$

[4] Symbol $[0;1]^{\mathbb{R}}$ denotes the family of all function from the real line $\mathbb{R}$ into the interval $[0,1]$.

$$\forall_{(x,y,z)\in\mathbb{R}^3}\colon\ x \le y \le z \Longrightarrow \mu_K(y) \ge \min\{\mu_K(x), \mu_K(z)\}. \tag{2}$$

The set of all fuzzy numbers will be denoted as $\mathbb{F}$. Arithmetic operations on fuzzy numbers were defined by Dubois, Prade (1978). By $\circledcirc$ we denote the ordinary arithmetic operation $\circ$ generalized to the case of fuzzy numbers. According to the Zadeh's Extension Principle (Zadeh, 1965), a sum of fuzzy numbers $K, L \in \mathbb{F}$ represented by their corresponding membership functions $\mu_K, \mu_L \in [0;1]^{\mathbb{R}}$ is a fuzzy subset

$$M = K \oplus L \tag{3}$$

described by its membership function $\mu_M \in [0;1]^{\mathbb{R}}$

$$\mu_M(z) = \sup\{\mu_K(x) \wedge \mu_L(z-x)\colon x \in \mathbb{R}\}. \tag{4}$$

The sum of the sequence $\{K_i\}_{i=1}^n \subset \mathbb{F}$ we denote by

$$\oplus_{i=1}^n K_i = K_1 \oplus K_2 \oplus \ldots \oplus K_n. \tag{5}$$

Analogously, the multiplication of a real number $y \in \mathbb{R}^+$ and a fuzzy number $K \in \mathbb{F}$ represented by membership function $\mu_K \in [0;1]^{\mathbb{R}}$ is a fuzzy subset

$$N = y \odot K \tag{6}$$

described by its membership function $\mu_N \in [0;1]^{\mathbb{R}}$

$$\mu_N(z) = \mu_K\left(\frac{z}{y}\right). \tag{7}$$

Moreover, if $y = 0$, then the multiplication (6) is equal to zero. The class of fuzzy real numbers is closed under the operations (3) and (6). Our further research will be limited to the case of fuzzy numbers with bounded support.

Fuzzy numbers are widely used for modeling assessments or estimations of a parameter given imprecisely. After Klir (1993) we understand imprecision as a superposition of ambiguity and indistinctness of information. Ambiguity can be interpreted as a lack of a clear recommendation between one alternative among various others. Indistinctness is understood as a lack of explicit distinction between recommended and not recommended alternatives. An increase in information imprecision makes it less useful and therefore it is logical to consider the problem of imprecision assessment.

We measure the ambiguity of a fuzzy number by applying the Khalili's measure (1979) to the energy measure $d: \mathcal{F}(\mathbb{R}) \to \mathbb{R}_0^+$ defined by de Luca, Termini (1979). For an arbitrary fuzzy number $K \in \mathbb{F}$, with membership function $\mu_K \in [0;1]^{\mathbb{R}}$ we have

$$d(K) = \int_{-\infty}^{+\infty} \mu_K(x)dx. \tag{8}$$

The indistinctness of an arbitrary fuzzy number can be measured by its entropy $e: \mathcal{F}(\mathbb{R}) \to \mathbb{R}_0^+$, also defined by de Luca, Termini (1972) and in form given by Kosko (1986). For an arbitrary fuzzy number $K \in \mathbb{F}$ we have

$$e(K) = \frac{d(K \cap K^C)}{d\big((K \cup K^C) \cap \mathbb{S}(K)\big)}, \tag{9}$$

where $\mathbb{S}(K) = \{x \in \mathbb{R}: \mu_K(x) > 0\}$.

The main focus of this study is a trapezoidal fuzzy number. The fuzzy number $Tr(r,s,t,u)$, defined for a non-decreasing sequence $\{r,s,t,u\} \subset \mathbb{R}$ with a membership function $\mu(\cdot\,|r,s,t,u) \in [0,1]^{\mathbb{R}}$ given by the formula

$$\mu(x|r,s,t,u) = \begin{cases} 0 & \text{for} \quad x < r, \\ \dfrac{x-r}{s-r} & \text{for} \quad r \le x < s, \\ 1 & \text{for} \quad s \le x \le t, \\ \dfrac{x-u}{t-u} & \text{for} \quad t < x \le u, \\ 0 & \text{for} \quad x > u, \end{cases} \tag{10}$$

is a trapezoidal fuzzy number.

For any arbitrary pair of trapezoidal fuzzy numbers, $Tr(r_1,s_1,t_1,u_1)$ and $Tr(r_2,s_2,t_2,u_2)$ and $a,b \in \mathbb{R}_0^+$ we have:

$$\begin{gathered} Tr(ar_1 + br_2, as_1 + bs_2, at_1 + bt_2, au_1 + bu_2) = \\ = \big(a \odot Tr(r_1,s_1,t_1,u_1)\big) \oplus \big(b \odot Tr(r_2,s_2,t_2,u_2)\big), \end{gathered} \tag{11}$$

$$d\big(Tr(r_1,s_1,t_1,u_1)\big) = \frac{1}{2}(u_1 + t_1 - r_1 - s_1), \tag{12}$$

$$e\big(Tr(r_1,s_1,t_1,u_1)\big) = \frac{s_1 - r_1 + u_1 - t_1}{-s_1 - 3r_1 + 3u_1 + t_1}. \tag{13}$$

3. RETURN RATE FROM A FINANCIAL ASSET

All calculations in this article are performed for a fixed time $t > 0$. We use simple return rates r_t defined as

$$r_t = \frac{V_t - V_0}{V_0}, \tag{14}$$

where:

— V_t is a *FV* described by random variable $\tilde{V}_t: \Omega \to \mathbb{R}$;
— V_0 is a *PV* assessed precisely or approximately.

Variable *FV* is described by a relation

$$\tilde{V}_t(\omega) = \check{C} \cdot \big(1 + \tilde{r}_t(\omega)\big), \tag{15}$$

where the simple return rate $\tilde{r}_t: \Omega \to \mathbb{R}$ is determined for *PV* equal to the market price $\check{C}$. After Markowitz (1952) we assume that $\tilde{r}$ rate has a normal probability distribution $N(\bar{r}, \sigma)$.

Moreover, in the researched case we assume that the *PV* is estimated by a trapezoidal fuzzy number $Tr\big(\check{C}_{min}, \check{C}_*, \check{C}^*, \check{C}_{max}\big)$, determined by membership function $\mu \in [0,1]^{\mathbb{R}}$ described by (10). This condition was initially introduced by Kuchta (2000) and applied in Siwek (2015). Parameters of the trapezoidal fuzzy number $Tr\big(\check{C}_{min}, \check{C}_*, \check{C}^*, \check{C}_{max}\big)$ are interpreted as follows:

— $\check{C}$ is the market price,
— $\check{C}_{min} \in \left]0, \check{C}\right]$ is the maximal lower bound of *PV*,
— $\check{C}_{max}$ $\left[\check{C}, +\infty\right[$ is the minimal upper bound of *PV,*
— $\check{C}_* \in \left[\check{C}_{min}, \check{C}\right]$ is the minimal upper assessment of prices visibly lower than the market price $\check{C}$,
— $\check{C}^* \in \left[\check{C}, \check{C}_{max}\right]$ is the maximal lower assessment of prices visibly higher than the market price $\check{C}$.

Methods of determining parameters $\check{C}_{min}, \check{C}_{max}$ are given in Piasecki, Siwek (2015). These parameters are non-negative.

According to the Zadeh's Extension Principle, a simple return rate for *PV* given as a trapezoidal fuzzy number is a fuzzy probabilistic set with membership function $\tilde{\rho} \in [0; 1]^{\mathbb{R}\times\Omega}$

$$\tilde{\rho}(r, \omega) = \sup\left\{\mu\big(x\big|\check{C}_{min}; \check{C}_*; \check{C}^*; \check{C}_{max}\big): x = \frac{\tilde{V}_t(\omega)}{1 + r}, x \in \mathbb{R}\right\} = \tag{16}$$

$$= \mu\left(\frac{\tilde{V}_t(\omega)}{1+r}\middle|\check{C}_{min};\check{C}_*;\check{C}^*;\check{C}_{max}\right) = \mu\left(\check{C}\,\frac{1+\tilde{r}_t(\omega)}{1+r}\middle|\check{C}_{min};\check{C}_*;\check{C}^*;\check{C}_{max}\right).$$

According to (10), formula (16) can be transformed into

$$\rho(r,\omega) = \begin{cases} \dfrac{\check{C}\,\dfrac{1+\tilde{r}_t(\omega)}{1+r} - \check{C}_{min}}{\check{C}_* - \check{C}_{min}}, & \text{for } \check{C}_{min} \le \check{C}\,\dfrac{1+\tilde{r}_t(\omega)}{1+r} < \check{C}_*, \\ 1, & \text{for } \check{C}_* \le \check{C}\,\dfrac{1+\tilde{r}_t(\omega)}{1+r} < \check{C}^*, \\ \dfrac{\check{C}\,\dfrac{1+\tilde{r}_t(\omega)}{1+r} - \check{C}_{max}}{\check{C}^* - \check{C}_{max}}, & \text{for } \check{C}^* < \check{C}\,\dfrac{1+\tilde{r}_t(\omega)}{1+r} \le \check{C}_{max}, \\ 0, & \text{for } \check{C}\,\dfrac{1+\tilde{r}_t(\omega)}{1+r} > \check{C}_{max}, \check{C}\,\dfrac{1+\tilde{r}_t(\omega)}{1+r} < \check{C}_{min}. \end{cases} \tag{17}$$

Thus, the expected return rate $R \in \mathbb{F}$ is a fuzzy number with membership function $\rho \in [0,1]^{\mathbb{R}}$:

$$\rho(r) = \begin{cases} \dfrac{\check{C}\,\dfrac{1+\bar{r}}{1+r} - \check{C}_{min}}{\check{C}_* - \check{C}_{min}}, & \text{for } \check{C}_{min} \le \check{C}\,\dfrac{1+\bar{r}}{1+r} < \check{C}_*, \\ 1, & \text{for } \check{C}_* \le \check{C}\,\dfrac{1+\bar{r}}{1+r} < \check{C}^*, \\ \dfrac{\check{C}\,\dfrac{1+\bar{r}}{1+r} - \check{C}_{max}}{\check{C}^* - \check{C}_{max}}, & \text{for } \check{C}^* < \check{C}\,\dfrac{1+\bar{r}}{1+r} \le \check{C}_{max}, \\ 0, & \text{for } \check{C}\,\dfrac{1+\bar{r}}{1+r} > \check{C}_{max}, \check{C}\,\dfrac{1+\bar{r}}{1+r} < \check{C}_{min}. \end{cases} \tag{18}$$

The expected discount factor $\bar{v}$ calculated using the return rate $\bar{r}$ is given by identity

$$\bar{v} = \frac{1}{1+\bar{r}} \tag{19}$$

which is its definition. Therefore, the function $\delta \in [0;1]^{\mathbb{R}}$ described by

$$\delta(v) = \delta\left(\frac{1}{1+r}\right) = \rho(r) \tag{20}$$

is a membership function of the expected discount factor $D \in \mathbb{F}$ calculated using the expected return rate $R \in \mathbb{F}$. Combining both (18) and (20) we get

$$\delta(v)=\begin{cases}\dfrac{\check{C}v-\bar{v}\check{C}_{min}}{\bar{v}\check{C}_{*}-\bar{v}\check{C}_{min}} & \text{for} \quad \dfrac{\bar{v}}{\check{C}}\check{C}_{min}\le v\le\dfrac{\bar{v}}{\check{C}}\check{C}_{*},\\ 1 & \text{for} \quad \dfrac{\bar{v}}{\check{C}}\check{C}_{*}<v<\dfrac{\bar{v}}{\check{C}}\check{C}^{*},\\ \dfrac{\check{C}v-\bar{v}\check{C}_{max}}{\bar{v}\check{C}^{*}-\bar{v}\check{C}_{max}} & \text{for} \quad \dfrac{\bar{v}}{\check{C}}\check{C}^{*}<v\le\dfrac{\bar{v}}{\check{C}}\check{C}_{max},\\ 0 & \text{for} \quad \dfrac{\bar{v}}{\check{C}}\check{C}_{max}<v, \qquad v<\dfrac{\bar{v}}{\check{C}}\check{C}_{min}.\end{cases} \tag{21}$$

One can see that the expected fuzzy discount factor stated above is also a trapezoidal fuzzy number $Tr\left(\check{C}_{min}\frac{\bar{v}}{\check{C}},\check{C}_{*}\frac{\bar{v}}{\check{C}},\check{C}^{*}\frac{\bar{v}}{\check{C}},\check{C}_{max}\frac{\bar{v}}{\check{C}}\right)$.

An increase in ambiguity of expected discount factor $D\in\mathbb{F}$ suggests a higher number of alternative recommendations to choose from. This may result in making a decision, which will be *ex post* associated with a profit lower than maximal, that is with a loss of chance. This kind of risk is called an ambiguity risk. The ambiguity risk of D is measured by energy measure $d(D)$.

An increase in the indistinctness of D, on the other hand, suggests that the differences between recommended and not recommended decision alternatives are harder to differentiate. This leads to an increase in the indistinctness risk, that is in the risk of choosing a not recommended option. The indistinctness risk of an expected discount factor D is measured by entropy measure $e(D)$. Imprecision risk consists of both ambiguity and indistinctness risk, combined.

From (15) we have that the return rate is a function of the future value of an asset, which is uncertain, since we don't know the future state of the world. Because of this, the investor is not sure whether they will gain or lose from the decision made. With the increase in uncertainty, the risk of making a wrong decision is higher. Here, uncertainty risk of a return rate will be measured by its variance σ^2.

As compared to energy and entropy measures of a return rate form a portfolio with component assets given imprecisely by a triangular or trapezoidal fuzzy number (Siwek, 2015), the simplicity of those measures calculated for discount factors encourages their use in analyzing portfolios burdened with imprecision. The criterion of minimizing the discounting factor may then substitute the criterion of return rate maximization, with same theoretical conclusions.

4. MULTI-ASSET PORTFOLIO

By a financial portfolio we will understand an arbitrary, finite set of financial assets. Each of this assets is characterized by its assessed PV and anticipated return rate.

Let us consider the case of a multi-asset portfolio π, consisting of financial assets Y_i $(i = 1,2,\dots,n)$. The PV of assets Y_i is estimated by fuzzy trapezoidal number $Tr\left(\check{C}_{min}^{(i)}, \check{C}_{*}^{(i)}, \check{C}^{*(i)}, \check{C}_{max}^{(i)}\right)$ where parameters are given as follows:

— $\check{C}^{(i)}$ is the market price,
— $\check{C}_{min}^{(i)} \in \left]0, \check{C}^{(i)}\right]$ is the maximal lower bound of PV,
— $\check{C}_{max}^{(i)} \in \left[\check{C}^{(i)}, +\infty\right[$ is the minimal upper bound of PV,
— $\check{C}_{*}^{(i)} \in \left[\check{C}_{min}^{(i)}, \check{C}^{(i)}\right]$ is the minimal upper assessment of prices visibly lower than the market price $\check{C}^{(i)}$,
— $\check{C}^{*(i)} \in \left[\check{C}^{(i)}, \check{C}_{max}^{(i)}\right]$ is the maximal lower assessment of prices visibly higher than the market price $\check{C}^{(i)}$.

We assume that for each security Y_i we know the simple return rate $\tilde{r}_t^i: \Omega \to \mathbb{R}$ appointed by (14) for the PV equal to the market price $\check{C}^{(i)}$. After Markowitz (1952) we assume that the n-dimensional variable $(\tilde{r}_t^1, \tilde{r}_t^2, \dots, \tilde{r}_t^n)^T$ has a cumulative normal distribution $N(\bar{\boldsymbol{r}}, \boldsymbol{\Sigma})$ where $\bar{\boldsymbol{r}} = (\bar{r}_1, \bar{r}_2, \dots, \bar{r}_n)^T$. We appoint an expected discount factor of security Y_i:

$$D^{(i)} = Tr\left(\check{C}_{min}^{(i)} \frac{\bar{v}_i}{\check{C}^{(i)}}, \check{C}_{*}^{(i)}, \check{C}^{*(i)} \frac{\bar{v}_i}{\check{C}^{(i)}}, \check{C}_{max}^{(i)} \frac{\bar{v}_i}{\check{C}^{(i)}}\right), \tag{22}$$

where $\bar{v}_i$ is an expected discount factor appointed using the expected return rate $\bar{r}_i$. According to (12), the energy measure of $D^{(i)}$ is given by

$$d\left(D^{(i)}\right) = \frac{\bar{v}_i}{2\check{C}^i}\left(\check{C}^{*(i)} + \check{C}_{max}^{(i)} - \check{C}_{min}^{(i)} - \check{C}_{*}^{(i)}\right), \tag{23}$$

and from (13), the entropy measure of a discounting factor can be calculated as

$$e\left(D^{(i)}\right) = \frac{\check{C}^{*(i)} - \check{C}_{min}^{(i)} + \check{C}_{max}^{(i)} - \check{C}_{*}^{(i)}}{-\check{C}_{*}^{(i)} - 3\check{C}_{min}^{(i)} + 3\check{C}_{max}^{(i)} + \check{C}^{*(i)}}. \tag{24}$$

We have that the market value $\check{C}^{(\pi)}$ of a portfolio π is equal to

$$\check{C}^{(\pi)} = \sum_{i=1}^{n} \check{C}^{(i)}. \tag{25}$$

Share p_i of an instrument Y_i in the portfolio π is given by

$$p_i = \frac{\check{C}^{(i)}}{\check{C}^{(\pi)}}. \tag{26}$$

We denote $\bar{\boldsymbol{p}} = (p_1, p_2, \dots, p_n)^T$. Then expected portfolio return rate $\bar{r}$ equals

$$\bar{r} = \bar{\boldsymbol{p}}^T \bar{\boldsymbol{r}}. \tag{27}$$

As for the present value of the portfolio, according to (11) it is also a trapezoidal fuzzy number

$$\begin{aligned} PV^{(\pi)} = Tr\left(\sum_{i=1}^{n} \check{C}_{min}^{(i)}, \sum_{i=1}^{n} \check{C}_{*}^{(i)}, \sum_{i=1}^{n} \check{C}^{*(1)}, \sum_{i=1}^{n} \check{C}_{max}^{(i)}\right) = \\ = Tr\left(\check{C}_{min}^{(\pi)}, \check{C}_{*}^{(\pi)}, \check{C}^{*(\pi)}, \check{C}_{max}^{(\pi)}\right). \end{aligned} \tag{28}$$

By (20), one can calculate the fuzzy expected discount factor of the portfolio π:

$$D^{(\pi)} = Tr\left(\check{C}_{min}^{(\pi)} \frac{\bar{v}}{\check{C}^{(\pi)}}, \check{C}_{*}^{(\pi)} \frac{\bar{v}}{\check{C}^{(\pi)}}, \check{C}^{*(\pi)} \frac{\bar{v}}{\check{C}^{(\pi)}}, \check{C}_{max}^{(\pi)} \frac{\bar{v}}{\check{C}^{(\pi)}}\right), \tag{29}$$

where $\bar{v}$ is a discounting factor calculated for expected return rate $\bar{r}$. We have

$$\frac{1}{\bar{v}} = \sum_{i=1}^{n} \frac{p_i}{\bar{v}_i}, \tag{30}$$

from which we obtain:

$$\bar{v} = \left(\sum_{i=1}^{n} \frac{p_i}{\bar{v}_i}\right)^{-1} = \left(\sum_{i=1}^{n} \frac{p_i}{\bar{v}_i}\right)^{-1} \sum_{i=1}^{n} p_i = \left(\sum_{i=1}^{n} \frac{p_i}{\bar{v}_i}\right)^{-1} \sum_{i=1}^{n} \frac{p_i}{\bar{v}_i} \bar{v}_i, \tag{31}$$

$$\begin{aligned} \frac{\bar{v}}{\check{C}^{(\pi)}} \check{C}_{min}^{(\pi)} = \frac{\bar{v}}{\check{C}^{(\pi)}} \sum_{i=1}^{n} \check{C}_{min}^{(i)} = \bar{v} \sum_{i=1}^{n} p_i \frac{\check{C}_{min}^{(i)}}{\check{C}^{(i)}} = \\ = \left(\sum_{i=1}^{n} \frac{p_i}{\bar{v}_i}\right)^{-1} \sum_{i=1}^{n} \frac{p_i}{\bar{v}_i} \left(\bar{v}_i \frac{\check{C}_{min}^{(i)}}{\check{C}^{(i)}}\right), \end{aligned} \tag{32}$$

$$\begin{aligned} \frac{\bar{v}}{\check{C}^{(\pi)}} \check{C}_{*}^{(\pi)} = \frac{\bar{v}}{\check{C}^{(\pi)}} \sum_{i=1}^{n} \check{C}_{*}^{(i)} = \bar{v} \sum_{i=1}^{n} p_i \frac{\check{C}_{*}^{(i)}}{\check{C}^{(i)}} = \\ = \left(\sum_{i=1}^{n} \frac{p_1}{\bar{v}_1}\right)^{-1} \sum_{i=1}^{n} \frac{p_i}{\bar{v}_i} \left(\bar{v}_i \frac{\check{C}_{*}^{(i)}}{\check{C}^{(i)}}\right), \end{aligned} \tag{33}$$

$$\begin{aligned} \frac{\bar{v}}{\check{C}^{(\pi)}} \check{C}^{*(\pi)} = \frac{\bar{v}}{\check{C}^{(\pi)}} \sum_{i=1}^{n} \check{C}^{*(i)} = \bar{v} \sum_{i=1}^{n} p_i \frac{\check{C}^{*(i)}}{\check{C}^{(i)}} = \\ = \left(\sum_{i=1}^{n} \frac{p_i}{\bar{v}_i}\right)^{-1} \sum_{i=1}^{n} \frac{p_i}{\bar{v}_i} \left(\bar{v}_1 \frac{\check{C}^{*(i)}}{\check{C}^{(i)}}\right), \end{aligned} \tag{34}$$

$$\frac{\bar{v}}{\check{C}^{(\pi)}}\check{C}_{max}^{(\pi)} = \frac{\bar{v}}{\check{C}^{(\pi)}}\sum_{i=1}^{n}\check{C}_{max}^{(i)} = \bar{v}\sum_{i=1}^{n}p_i\frac{\check{C}_{max}^{(i)}}{\check{C}^{(i)}} = \\ = \left(\sum_{i=1}^{n}\frac{p_i}{\bar{v}_i}\right)^{-1}\sum_{i=1}^{n}\frac{p_i}{\bar{v}_i}\left(\bar{v}_i\frac{\check{C}_{max}^{(i)}}{\check{C}^{(i)}}\right). \tag{35}$$

From the formulas given above, we can rewrite the fuzzy discount factor as

$$D^{(\pi)} = \left(\sum_{i=1}^{n}\frac{p_i}{\bar{v}_i}\right)^{-1} \odot \left(\oplus_{i=1}^{n}\frac{p_i}{\bar{v}_i}\odot D^{(i)}\right). \tag{36}$$

From (12) and (36), we obtain that the energy measure of an expected discounting factor $D \in \mathbb{F}$ is a linear combination of energy measures calculated for each of component assets

$$d\left(D^{(\pi)}\right) = \left(\sum_{i=1}^{n}\frac{p_i}{\bar{v}_i}\right)^{-1}\sum_{i=1}^{n}\frac{p_i}{\bar{v}_i}d\left(D^{(i)}\right). \tag{37}$$

The relation above suggests that the energy of a fuzzy expected discount factor of a portfolio π is, in fact, a linear combination of weighted energies of those factors calculated for its components. The weights calculated for the assets Y_i increase with their shares in the portfolio and, respectively, decrease with the value of their discount factor $\bar{v}_i$. This fact leads to a conclusion, that when trying to minimize the ambiguity risk of a portfolio, one should focus on minimizing the ambiguity of component assets, which are characterized by the highest expected return rates. On the other hand, the shares of an asset in the whole portfolio are, according to the theory, appointed *post factum,* by gathering available information on said assets. Condition (37) shows that, in the researched case, the portfolio diversification only ”averages” the risk of ambiguity.

According to (13), the entropy measure of expected discount factor is equal

$$e\left(D^{(\pi)}\right) = \frac{\check{C}_{*}^{(\pi)} - \check{C}_{min}^{(\pi)} + \check{C}_{max}^{(\pi)} - \check{C}^{*(\pi)}}{-\check{C}_{*}^{(\pi)} - 3\check{C}_{min}^{(\pi)} + 3\check{C}_{max}^{(\pi)} + \check{C}^{*(\pi)}}. \tag{38}$$

The variance of a portfolio return rate is calculated by

$$\sigma^2 = \bar{\boldsymbol{p}}^T\boldsymbol{\Sigma}\bar{\boldsymbol{p}}. \tag{39}$$

By constructing a portfolio which minimizes the variance Markowitz proved that portfolio diversification can ”minimize” the uncertainty risk.

5. CASE STUDY

The portfolio π consists of financial assets Y_1 and Y_2. Anticipated vector $(\tilde{r}_t^1, \tilde{r}_t^2)^T$ of their simple return rates has two-dimensional normal distribution

$$N\left((0.25,0.5)^T, \begin{bmatrix} 0.5 & -0.1 \\ -0.1 & 0.4 \end{bmatrix}\right).$$

For the asset Y_1 with market price $\check{C}^{(1)} = 24$, its PV is estimated by a trapezoidal fuzzy number $Tr(18, 23, 25, 37)$. Then according to (18), the fuzzy expected return rate from Y_1 is a fuzzy number $R_1 \in \mathbb{F}$ given by membership function $\rho_1 \in [0,1]^{\mathbb{R}}$

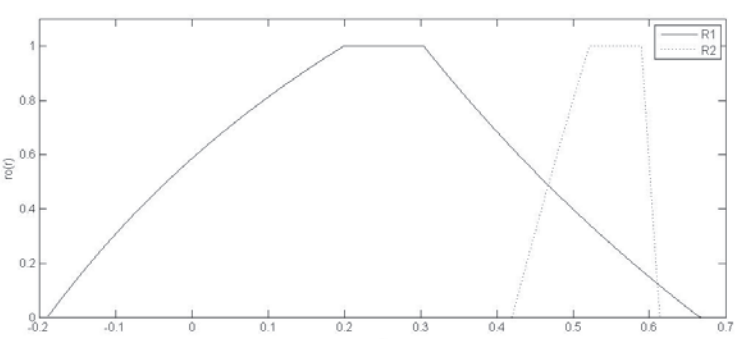
$$\rho_1(r) = \begin{cases} \dfrac{6}{1+r} - 3.6, & \text{for} \quad 0.67 \geq r > 0.30, \\ 1, & \text{for} \quad 0.30 \geq r > 0.20, \\ \dfrac{-2,5}{1+r} + 3.08 & \text{for} \quad 0.20 > r \geq -0.19, \\ 0 & \text{for} \quad r \notin [-0.19, \quad 0.67]. \end{cases}$$

For the asset Y_2 with market price $\check{C}^{(2)} = 69$, its PV is estimated by fuzzy trapezoidal number $Tr(66, 67, 70, 75)$. Then according to (18), the fuzzy expected return rate of Y_2 is a fuzzy number $R_2 \in \mathbb{F}$ with membership function $\rho_2 \in [0,1]^{\mathbb{R}}$

$$\rho_2(r) = \begin{cases} \dfrac{103.5}{1+r} - 66, & \text{for} \quad 0.57 \geq r > 0.54, \\ 1, & \text{for} \quad 0.54 \geq r > 0.48, \\ \dfrac{-20.7}{1+r} + 15 & \text{for} \quad 0.48 > r \geq 0.38, \\ 0 & \text{for} \quad r \notin [0.38, 0.57]. \end{cases}$$

Membership functions for return rates of both assets are presented in figure 1. It is easy to see that the expected return rate is not a trapezoidal fuzzy number.

Figure 1. Membership functions for expected return rates R_1 and R_2



Source: own study.

Using the expected return rate R_1 we may now appoint by means of (21) a fuzzy expected discount factor $D^{(1)} \in \mathbb{F}$. We have

$$D^{(1)} = Tr\left(18\frac{0.8}{24}, 23\frac{0.8}{24}, 25\frac{0.8}{24}, 37\frac{0.8}{24}\right) = Tr(0.6, 0.77, 0.83, 1.23).$$

According to (29), its energy measure equals

$$d\left(D^{(1)}\right) = \frac{0{,}8}{2 \cdot 24}(37 - 18 + 25 - 23) = 0.35.$$

and from (30), the entropy measure has the value of

$$e\left(D^{(1)}\right) = \frac{23 - 18 + 37 - 25}{-23 - 3 \cdot 18 + 3 \cdot 37 + 25} = 0.29.$$

The expected discount factor $D^{(2)} \in \mathbb{F}$ of the second asset calculated using R_2 equals

$$D^{(2)} = Tr\left(66\frac{0{,}67}{69}, 67\frac{0{,}67}{69}, 70\frac{0{,}67}{69}, 75\frac{0{,}67}{69}\right) = Tr(0.64,0.6,0.68,0.73).$$

Also, its energy measure equals

$$d\left(D^{(2)}\right) = \frac{0{,}67}{2 \cdot 69}(75 - 66 + 70 - 67) = 0.06.$$

and entropy measure has the value of

$$e\left(D^{(2)}\right) = \frac{67 - 66 + 75 - 70}{-67 - 3 \cdot 66 + 3 \cdot 75 + 70} = 0.2.$$

The market price of portfolio π is equal to

$$\check{C}^{(\pi)} = 24 + 69 = 93.$$

Corresponding to (26), shares p_1 and p_2 of Y_1 and Y_2 in the portfolio π are equal

$$p_1 = \frac{24}{93}, \quad p_2 = \frac{69}{93}.$$

We can appoint the fuzzy expected discount factor $D^{(\pi)} \in \mathbb{F}$ of the portfolio π. By (36), it is a fuzzy number of the form

$$D^{(\pi)} = \left(\left(\left(\frac{\frac{24}{93}}{0.8} + \frac{\frac{69}{93}}{0.67}\right)^{-1} \cdot \frac{\frac{24}{93}}{0.8}\right) \odot D^{(1)}\right) \oplus$$

$$\oplus \left(\left(\left(\frac{\frac{24}{93}}{0.8} + \frac{\frac{69}{93}}{0.67}\right)^{-1} \cdot \frac{\frac{69}{93}}{0.67}\right) \odot D^{(2)}\right) =$$

$$= \left(0.2256 \odot D^{(1)}\right) \oplus \left(0.7744 \odot D^{(2)}\right) = Tr(0.63, 0.68, 0.71, 0.84).$$

Its energy measure calculated by (37) equals

$$d\left(D^{(\pi)}\right) = 0.2256 \cdot 0.35 + 0.7744 \cdot 0.06 = 0.13.$$

Entropy measure can be calculated by (38)

$$e\left(D^{(\pi)}\right) = e\left(Tr(0.63, 0.68, 0.71, 0.84)\right) = \frac{0.68 - 0.63 + 0.84 - 0.71}{-0.68 - 3 \cdot 0.63 + 3 \cdot 0.84 + 0.71} = 0{,}27.$$

Let us note that we have

$$\left(\frac{p_1}{\bar{v}_1} + \frac{p_2}{\bar{v}_2}\right)^{-1} \left(\frac{p_1}{\bar{v}_1} e\left(D^{(1)}\right) + \frac{p_2}{\bar{v}_2} e\left(D^{(2)}\right)\right) =$$

$$= 0.2256 \cdot 0.29 + 0.7744 \cdot 0.2 = 0.2203 \neq 0{,}27 = e\left(D^{(\pi)}\right).$$

It implies that the portfolio entropy measure $e\left(D^{(\pi)}\right)$ cannot be calculated similarly to the portfolio energy measure $e\left(D^{(\pi)}\right)$ by the linear combination (37).

We obtain following relations between the energy measure and entropy measure appointed for fuzzy expected discount factors of portfolio and its components:

$$d(D_1) > d\left(D^{(\pi)}\right) > d(D_2),$$

$$e(D_1) > e\left(D^{(\pi)}\right) > e(D_2).$$

These inequalities show that the portfolio diversification can average the imprecision risk. Moreover, using (39) we calculate the variance of a return rate from portfolio:

$$\sigma^2 = 0.2175.$$

By increasing the number of assets in the portfolio, we can lower the variance (which approaches its limit with number of assets going to infinity). This means that creating a multi asset portfolio π results in minimizing the uncertainty risk.

Let us consider now any portfolio π consisting of financial assets Y_1 and Y_2. The contribution of the instrument Y_i in the portfolio π is equal to p_i. Then, according to (36), the expected discount factor $D^{(\pi)} \in \mathbb{F}$ of the portfolio π can be calculated in the following way

$$D^{(\pi)} = \left(\frac{p_1}{0.8} + \frac{p_2}{0.67}\right)^{-1} \odot$$

$$\odot \left(\left(\frac{p_1}{0.8} \odot Tr(0.6, 0.77, 0.83, 1.23)\right) \oplus \left(\frac{p_2}{0.67} \odot Tr(0.64; 0.65; 0.68; 0.73)\right)\right) =$$

$$= \frac{0.67p_1 \odot Tr(0.6, 0.77, 0.83, 1.23) \oplus 0.8p_2 \odot Tr(0.64; 0.65; 0.68; 0.73)}{0.667p_1 + 0.8p_2} =$$

$$= \frac{p_1 \odot Tr(0.402, 0.5159, 0.5561, 0.8241) \oplus p_2 \odot Tr(0.512, 0.52, 0.544, 0.584)}{0.67p_1 + 0.8p_2}.$$

We see that the expected fuzzy discount factor of portfolio can be expressed as a combination of securities contributions and their expected fuzzy discount factors. In an analogous way the ambiguity risk may be evaluated because of the energy measure for this factor by (37) is given as follows:

$$d\left(D^{(\pi)}\right) = \left(\frac{p_1}{0.8} + \frac{p_2}{0.67}\right)^{-1} \left(\frac{p_1}{0.8} 0.35 + \frac{p_2}{0.67} 0.06\right) = \frac{0.67p_1 0.35 + 0.8p_2 0.06}{0.67p_1 + 0.8p_2} =$$

$$= \frac{0.2345p_1 + 0.048p_2}{0.67p_1 + 0.8p_2}.$$

Above we have shown that entropy measure $e\left(D^{(\pi)}\right)$ cannot be expressed in analogous way. The last two equations can be applied to the mathematical programing task dedicated to portfolio optimization.

6. SUMMARY

The main purpose of this article was to analyse the possibility of managing the risk burdening a two-asset portfolio, built with use of an imprecise information stemming from present value of component assets. The imprecise present values were modelled with by trapezoidal fuzzy numbers. For this assumptions we have reached the following conclusions:

— The portfolio diversification can lower uncertainty risk,
— The portfolio diversification averages ambiguity risk,
— The portfolio diversification can to average indistinctness risk.

The results obtained suggest, on one hand, that the portfolio diversification does not help in lowering the imprecision risk, but on the other hand, it also does not increase it. Thus, research suggests that there exist portfolios, which imprecision risk will not be minimized with portfolio diversification, and thus it is vital to create a new risk minimization problem, including all of the risk types.

The results obtained above encourage for their broader analysis. Further research can focus on generalizing the representation of the present value to an arbitrary fuzzy number. Helpful here can be fundamental results obtained in Goetschel, Voxman (1986) and Stefaninia et al. (2006).

REFERENCES

Buckley I. J., (1987), The Fuzzy Mathematics of Finance, *Fuzzy Sets and Systems*, 21, 257–273.

Caplan B., (2001), Probability, Common Sense, and Realism: a Reply to Hulsmann and Block, *The Quarterly Journal of Austrian Economics,* 4 (2), 69–86.

Chiu C. Y., Park C. S., (1994), Fuzzy Cash Flow Analysis Using Present Worth Criterion, *The Engineering Economist*, 39 (2), 113–138.

Czerwiński Z., (1960), Enumerative Induction and the Theory of Games, *Studia Logica*, 10, 24–36.

Czerwiński Z., (1969), *Matematyka na usługach ekonomii*, PWN, Warszawa.

de Luca A., Termini S., (1972), A Definition of a Non-Probabilistic Entropy in the Settings of Fuzzy Set Theory*, Information and Control*, 20, 301–313.

de Luca A., Termini S., (1979), Entropy and Energy Measures of Fuzzy Sets, in: Gupta M. M., Ragade R. K., Yager R. R., (eds.), *Advances in Fuzzy Set Theory and Applications*, 321––338.

Dubois D., Prade H., (1978), Operations on Fuzzy Numbers, *International Journal System Sciences*, 9, 613–626.

Dubois D., Prade H., (1979), Fuzzy Real Algebra: Some Results, *Fuzzy Sets and Systems*, 2, 327––348.

Goetschel R., Voxman W., (1986), Elementary Fuzzy Calculus, *Fuzzy Sets and Systems*, 18, 31––43.

Greenhut J. G., Norman G., Temponi C. T., (1995), Towards a Fuzzy Theory of Oligopolistic Competition, IEEE Proceedings of ISUMA–NAFIPS, 286–291.

Gutierrez I., (1989), Fuzzy Numbers and Net Present Value, *Scandinavian Journal of Management*, 5 (2), 149–159.

Hiroto K., (1981), Concepts of Probabilistic Sets, *Fuzzy Sets and Systems*, 5, 31–46.

Huang X., (2007), Two New Models for Portfolio Selection with Stochastic Returns Taking Fuzzy Information, *European Journal of Operational Research*, 180 (1), 396–405.

Khalili S., (1979), Fuzzy Measures and Mappings, *Journal of Mathematical Analysis and Applications*, 68 (1), 92–99.

Klir G. J., (1993), Developments in Uncertainty-Based Information, *Advances in Computers*, 36, 255–332.

Kolmogorov A. N., (1933), *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin, Julius Springer.

Kosko B., (1986), Fuzzy Entropy and Conditioning, *Information Sciences*, 40, 165–174.

Kuchta D., (2000), Fuzzy Capital Budgeting, *Fuzzy Sets and Systems*, 111, 367–385.

Lambalgen M. von, (1996), Randomness and Foundations of Probability: Von Mises' Axiomatization of Random Sequences, Institute of Mathematical Statistics Lecture Notes – Monograph Series, 30, 347–367.

Lesage C., (2001), Discounted Cash-flows Analysis. An Interactive Fuzzy Arithmetic Approach, *European Journal of Economic and Social Systems*, 15 (2), 49–68.

Markowitz H. S. M., (1952), Portfolio Selection, *Journal of Finance*, 7 (1), 77–91.

Piasecki K., (2011), Behavioural Present Value, SSRN Electronic Journal, 01/2011.

Piasecki K., (2014), Behawioralna wartość bieżąca – nowe podejście, *Optimum Studia Ekonomiczne*, 67, 36–45

Piasecki K., Siwek J., (2015), Behavioural Present Value Defined as Fuzzy Number – a New Approach, *Folia Oeconomica Stetinensia*, 15 (2), 27–41.

Piasecki K., Siwek J., (2017), Portfel dwuskładnikowy z trójkątnymi rozmytymi wartościami bieżącymi – podejście alternatywne, *Przegląd Statystyczny*, 64, 59–77.

Sadowski W., (1976), *Decyzje i prognozy*, PWE, Warszawa.

Sadowski W., (1980), Forecasting and Decision Making, Quantitative Wirtschafts – und Unternehmensforschung, Springer–Verlag, Berlin Heidelberg, 92–102

Sheen J. N., (2005), Fuzzy Financial Profitability Analyses of Demand Side Management Alternatives from Participant Perspective, *Information Sciences*,169, 329–364.

Siwek J., (2015), Portfel dwuskładnikowy – studium przypadku dla wartości bieżącej danej jako trójkątna liczba rozmyta, Studia Ekonomiczne, *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, 241, 140–150.

Stefaninia L., Sorini L., Guerra M. L., (2006), Parametric Representation of Fuzzy Numbers and Application to Fuzzy Calculus, *Fuzzy Sets and Systems*, 157 (18), 2423–2455.

Tsao C.–T., (2005), Assessing the Probabilistic Fuzzy Net Present Value for a Capital, Investment Choice Using Fuzzy Arithmetic, *Journal* of the *Chinese Institute of Industrial Engineers*, 22 (2), 106–118.

von Mises L., (1962), *The Ultimate Foundation of Economic Science An Essay on Method*, D. Van Nostrand Company, Inc., Princeton.

von Mises R., (1957), *Probability, Statistics and Truth*, The Macmillan Company, New York.

Ward T. L., (1985), Discounted Fuzzy Cash Flow Analysis, Proceedings of 1985 Fall Industrial Engineering Conference, Institute of Industrial Engineers, 476–481.

Zadeh L., (1965), Fuzzy Sets, *Information and Control*, 8, 338–353.

PORTFEL WIELOSKŁADNIKOWY Z TRAPEZOIDALNYMI ROZMYTYMI WARTOŚCIAMI BIEŻĄCYMI

Streszczenie

Głównym celem niniejszego artykułu jest przedstawienie charakterystycznych cech portfela wieloskładnikowego w przypadku, kiedy bieżące wartości składników portfela są trapezoidalnymi liczbami rozmytymi. W ramach analizy portfelowej jest wyznaczany rozmyty oczekiwany czynnik dyskonta i oceny ryzyka nieprecyzyjności. Dzięki temu pojawia się możliwość opisania wpływu dywersyfikacji portfela na ryzyko nieprecyzyjności. Przedstawione teoretyczne rozważania i uzyskane wnioski są poparte przykładem liczbowym.

Słowa kluczowe: portfel wieloskładnikowy, wartość bieżąca, trapezoidalna liczba rozmyta, czynnik dyskontowy

MULTI-ASSET PORTFOLIO WITH TRAPEZOIDAL FUZZY PRESENT VALUES

Abstract

The main purpose of the following paper is to present characteristics of a multi-asset portfolio in case of present values of composing financial instruments being modelled by a trapezoidal fuzzy number. Throughout the analysis a fuzzy expected discount factor and imprecision risk assessments are calculated. Thanks to that, there arises a possibility to describe the influence of portfolio diversification on imprecision risk. Presented theoretical inference and obtained conclusions are supported by numerical example.

Keywords: multi-asset portfolio, present value, trapezoidal fuzzy number, discount factor

DOI: 10.5604/01.3001.0014.0536

PRZEGLĄD STATYSTYCZNY
TOM LXV – ZESZYT 2 – 2018

Robert SZÓSTAKOWSKI[1]

The use of the Hurst exponent to investigate the quality of forecasting methods of ultra-high-frequency data of exchange rates[2]

1. INTRODUCTION

In this article, a review of forecasting method was conducted using historical tick data from the foreign exchange market. Both ask and bid prices from one trading day – from 5 pm 05-03-2012 to 5 pm 06-03-2012 for every exchange rate AUD/CAD, AUD/USD, GBP/JPY, GBP/PLN, GBP/USD, USD/CHF, USD/JPY were analyzed separately as independent data. Forecasting methods used in the article range from simple statistical methods like moving average, linear regression to more advanced like Kalman filter, ARMA, ARIMA models. Finally, machine learning methods like linear discriminant analysis and logistic regression were tested.

In the first part of the research, point forecasts were calculated using statistical models. For them, AMAPE errors were calculated to compare results between methods. In the second part of the research, logistic regression and linear discriminant analysis models were trained on historical data to predict a direction of bid/ask price change (up, down, the same). Additionally, the point forecasts from the previous part of the research were transformed into predictions of a direction of bid/ask price change. For a comparison of results a metric called forecasting accuracy was used (a percent of accurate forecasts).

The models were optimized by selecting the best hyperparameters based on their performance on historical data. The time series were divided into subseries of 10000 values. Every model was trained and optimized on the previous subseries and tested on out-of-sample data (on next 10000 values). These two steps were repeated for all subseries.

---

[1] Poznań University of Economics, Faculty of Applied Mathematics, 53 Towarowa St., 61–896 Poznań, Poland, email: robertszostakowski@gmail.com.

[2] The author is very grateful to two anonymous referees for their helpful comments and suggestions. The author would also like to show his gratitude to the thesis supervisor for sharing pearls of wisdom during this research. Lastly, I would like to thank my second advisor, Dr Michał Galas for providing tick data from the repositories of University College London and for his guidance.

Moreover, for every time series, a rolling window of past 150 values was used to calculate values of the Hurst exponent. Every exponent value was assigned to the last 150^{th} bid/ask quote from the rolling window. These results were compared with forecasting accuracy metrics from both parts of the research. At the end based on the forecasting results and time series characteristics, a few discovered dependencies were presented.

The main hypothesis of this research is the fact that the forecasting error decreases and the percentage forecasting accuracy increases when the value of the Hurst exponent increases. It was proven that analyzing the time series characteristics based on the chaos theory like a value of the Hurst exponent can be helpful in achieving better forecasting results. Finally, the average forecasting accuracy was higher for machine learning methods than for statistical methods, regardless of values of the Hurst exponent.

The idea behind this research was to develop a methodology which can be applied to the art of forecasting to increase the performance of a variety of models. The article main contribution to the science is a set of rules how to use the Hurst exponent for developing more accurate forecasting models.

The paper is divided into following parts. The first section is an introduction to the article. It describes the idea behind the research. Moreover, this section explains the main hypothesis and article contribution to the science. The next section describes the current state of the science in the field of forecasting and the Hurst exponent analysis. The third section characterizes the data used in the research. The next section is a detailed introduction to the Hurst exponent analysis with references to its application in other papers. The fifth section describes forecasting methods which were used in the research. The sixth section shows how the research was conducted. It describes an optimization algorithm and explains how the forecasting models were adjusted for obtaining the best results. The next section puts attention to the problems which were encountered by the researcher while forecasting high-frequency data. The eighth section includes a description of benchmarks which were used to compare forecasting models. The ninth section shows the empirical results of forecasting using statistical and machine learning models applied to tick currency data. The article ends with a section which summarizes the research and proposes further extensions.

2. RELATED LITERATURE

Forecasts based on large data sets gained a significant importance in every branch of economics. The necessity of forecasting led to a discovery of a range of methods starting from simple, autoregressive models to complicated nonlinear specifications. In most cases, a level of model complexity does not correlate with the expected results which is shown in the work of Green, Armstrong (2015). Moreover, it is impossible to discover a true model which generates a given time

series data according to the effective market hypothesis developed by Fama (1970) or the fractal market hypothesis discovered by Peters (1994) which says that price changes in financial markets are random or come from the deterministic chaos process.

An attempt to measure long memory effects was done by Hurst (1951) in his work in the area of hydrology. Analyzing the Hurst exponent on financial data might lead to more accurate, competitive out-of-sample forecasts and it is described in the work of Mitra (2012), Castillo, Melin (1996). Their research was conducted on daily stock market returns. According to their conclusions, values of the Hurst exponent which differ from 0.5 (random series) can be used as a predictor for investment strategies. Based on this fact, it was assumed that we can increase our investment results by analyzing values of the Hurst exponent over time. Furthermore, Qian, Rasshed (2004) proved that investigating values of the Hurst exponent calculated from Dow-Jones daily returns can increase the forecasting performance of neural networks. Unfortunately, he proved that analyzing only values of the Hurst exponent bigger than 0.5 leads to more accurate forecasts. In this research, a similar approach was applied to the full order book data of exchange rates. This article shows that the Hurst exponent results calculated from tick data have totally different density plots than in the research done by Qian, Rasshed (2004).

In the art of forecasting, it is almost impossible to repeat the forecasting accuracy in out-of-sample data based on historical results. Usually, models are selected by their performance on historical data what leads to the assumption that the best historical models will perform with at least the same accuracy in the future. A research presented by Aiolfi, Timmermann (2006) shows that a significant persistence in the forecasting performance can be found. The authors used data sets of stock market returns, interest rates and spread from main G7 economies during the 1959 and 1999 year.

In this article, their research is extended by analyzing time series characteristics based on the fractal theory developed by Peters (1994) to measure the accuracy of forecasting methods on high-frequency data of exchange rates. The research of Andersen, Bollerslev (1997) showed the importance of long-memory dependence in financial market volatility. This dependence in market volatility is characterized by slowly mean-reverting fractionally integrated process. Their article proved that forecasts of low-frequency volatility are more precise when based on high-frequency data. It established a link between financial markets microstructure and lower-frequency data relevance. Moreover, Cheung (1993) showed an evidence of long memory in exchange rates data. His research indicates that the empirical evidence of unit roots in exchange rates data is not robust to long memory alternatives. The author mentioned that the dependence is also hard to detect using impulse-response function analysis. The model used in the research – ARFIMA (integrated autoregressive moving average) did not

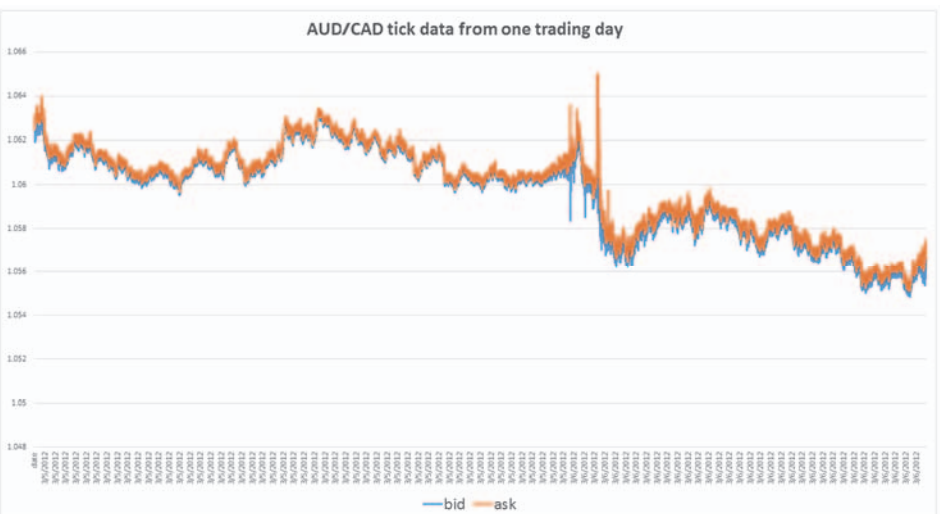
outperform a random walk in out-of-sample forecasts. Huang, Yang (1999) came to a similar conclusion using high-frequency data – one-minute time intervals in a given trading day for NYSE and NASDAQ. The researchers used the Modified Rescaled Range Analysis to discover a long-term memory in analyzed series. They showed sub periods during trading sessions when the random walk hypothesis was not supported. Their research conclusion is similar to Peters (1994) – local randomness and global determinism can actually coexist.

In the review of the forecasting methods, a wide range of models is used, starting from a simple moving average, ending with machine learning methods like linear discriminant analysis. Selecting two machine learning methods for the article was driven by analyzing the research results obtained in this area, particularly in the research of Ahmed et al. (2010), which shows how accurate predictions can be obtained using machine learning methods. Two models which were not used in the above-mentioned article: linear discriminant analysis and logistic regression were selected for the following research.

3. DATA AND PROGRAMMING FRAMEWORKS

The research was conducted on the ultra-high-frequency data set. The full order book data from one trading day – from 5 pm 05-03-2012 to 5 pm 06-03-2012 for seven major exchange rates was analyzed (AUD/CAD, AUD/USD, GBP/JPY, GBP/PLN, GBP/USD, USD/CHF, USD/JPY). Each exchange rate order book has a bid and ask series and they were both used in the forecasting research. As an example, one trading day from AUD/CAD exchange rate (bid and ask prices) is shown in figure 1.

Figure 1. Bid and ask time series from one trading day, AUD/CAD



The forecasts of the next value (in this case it was tick – a change in the price of a security from trade to trade) were conducted on both ask and bid prices separately. In summary, fourteen different time series were taken into the analysis, when we consider ask and bid series separately. Moreover, given time series were used to calculate factors necessary for forecasting models like the Hurst exponent, logarithmic returns, which are described in a more detailed oriented manner in the sixth section. The volume of analyzed data can be seen in table 1.

Table 1. NUMBER OF TICKS FOR BID AND ASK PRICES IN ANALYZED EXCHANGE RATES

Number of ticks	AUD/CAD	AUD/USD	GBP/JPY	GBP/PLN	GBP/USD	USD/CHF	USD/JPY
ask	143 395	533 049	359 564	93 417	307 348	441 409	483 167
bid	143 395	533 049	359 564	93 417	307 348	441 409	483 167

In analyzed time series several zeros were removed and replaced by the previous value. This algorithm was chosen as the safest cleansing approach since repeating data in exchange rates time series are very common.

The research was conducted in the framework developed in python 3.5.2. using libraries: scikit-learn 0.16.1, numpy 1.8.2, matplotlib 1.4.2, pandas 0.16.2.

4. HURST EXPONENT

In the article, the Hurst exponent analysis is used to discover subparts of the time series, which have different characteristics like persistency, randomness or anti-persistency. A value of the Hurst exponent is calculated by rescaled range analysis (R/S analysis) which is a statistical measure of the variability. It is calculated by dividing the range of the values exhibited in subseries by the standard deviation of the values over the same subseries. Namely, a series of length N is divided into a number of shorter time series d with a length n, where $d \cdot n = T$. Then for every sub-period with a length n, a rescaled range is calculated:

1. Calculation of the mean value

$$m = \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{1}$$

2. Calculation of the adjusted series Y

$$Y_t = X_t - m\,, \quad t = 1, 2, \dots, n, \tag{2}$$

3. Calculation of the cumulative deviation

$$Z_t = \sum_{i=1}^{t} Y_i, \quad \mathrm{t} = 1, 2, \ldots, \mathrm{n}, \tag{3}$$

4. Calculation of the range series – R

$$R(n) = max(Z_1, Z_2, \ldots, Z_n) - min(Z_1, Z_2, \ldots, Z_n), \tag{4}$$

5. Calculation of the standard deviation series – S

$$S(n) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2}, \tag{5}$$

where m is the mean value calculated in point 1.

6. Calculation of the rescaled range series $\frac{R}{S}$

$$\left(\frac{R}{S}\right)_n = \frac{R(n)}{S(n)}, \tag{6}$$

7. Hurst discovered that $\left(\frac{R}{S}\right)_n$ scales by power-law when the time increases which leads to the equation

$$E\left(\frac{R}{S}\right)_n = c\, n^H, \tag{7}$$

where c is a constant, independent of n and H is called the Hurst exponent.

The procedure described above is repeated for different values of $n = N, \frac{N}{2}, \frac{N}{4}, \frac{N}{8} \ldots$, where the minimum length of eight is usually chosen for the length of the smallest subseries. Finally, to estimate a value of the Hurst exponent, a simple, ordinary least squares regression is calculated on natural logarithms obtained from the equation from the 7th point.

$$lnE\left(\frac{\mathrm{R}}{\mathrm{S}}\right)_n = \ln c + H \ln n, \tag{8}$$

where H is the Hurst exponent.

The Hurst exponent divides time series into three groups:

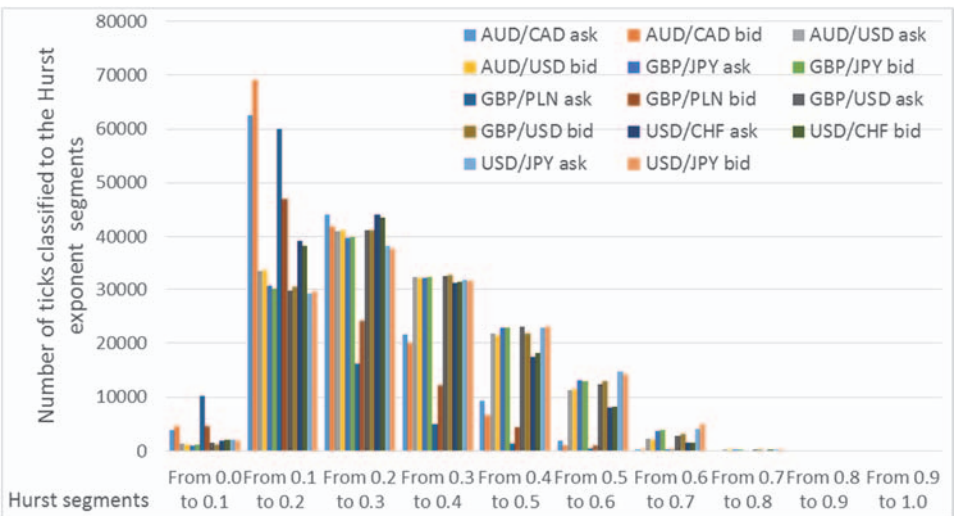
— anti-persistent when $H < 0.5$ – price in time series tends to come back to the long-term mean (weak trends),

— persistent when $0.5 < H < 1$ – this is the opposite of anti-persistent, which means there is a strong trend and long memory effects,

— random when $H = 0.5$.

In the research, the values of the Hurst exponent were calculated for all fourteen bid/ask time series with maximal n which is equal to 150 observations, which is very close to 2^7. That means the Hurst analysis was done for subseries with lengths $2^7, 2^6, 2^5, \ldots, 2^3$. After the Hurst exponent estimation, every value was classified into ten segments from 0 to 1 with a step 0.1. The candle density plot with the results of the classification can be seen in figure 2.

Figure 2. Number of ticks classified into the Hurst exponent segments for every exchange rate



Most of the values seem to be categorized to the anti-persistent group (from 0 to 0.5) which indicates that the time series show a mean-reverting tendency.

5. FORECASTING METHODS USED IN THE RESEARCH

Eight forecasting methods were analyzed in this article. Starting from simple statistical methods like moving average or exponential smoothing which were chosen mostly as benchmarks, going to more advanced, statistical methods like linear regression or Kalman filter. The last two methods were chosen with a hope that more advanced methods should provide better results than simple models.

Moreover, two time series models were chosen – ARMA and ARIMA as a representative of the econometric modeling approach. At the end, the approach to forecasting was changed and two methods based on machine learning were selected: logistic regression and linear quadratic regression.

A forecasted point result obtained for models from paragraphs 5.1 to 5.6 for selected time spans was always taken as a forecast for the next value. For models 5.7 and 5.8 a possible direction of the price change was forecasted. Possible three states were used:

— price increased,
— price decreased,
— price stayed the same.

In the second part of the research, the point forecasts for the models described in 5.1 to 5.6 were transformed to the forecasts of possible direction of bid/ask price change to make a comparison between statistical and machine learning models possible.

5.1. Simple moving average

”Moving average” referring to a type of stochastic process is an abbreviation of Wold’s (1939) process of moving average. It is an un-weighted mean of the previous n data which was used as a forecast for the next value. For calculating a simple moving average (SMA or running average) of n observations ($\boldsymbol{x_t}$, $\boldsymbol{x_{t-1}}$,…, $\boldsymbol{x_{t-(n-1)}}$) the following formula is used:

$$\mathrm{SMA} = \frac{x_t + x_{t-1} + \ldots + x_{t-(n-1)}}{n} = \frac{1}{n}\sum_{i=0}^{n-1} x_{t-i}. \tag{9}$$

5.2. Exponential smoothing

Historically, the method was independently developed by Robert Goodell Brown and Charles Holt and it was described in ”Smoothing, Forecasting and Prediction of Discrete Time Series” written by Brown (2004). The output of the exponential smoothing algorithm is defined as s_t, which can be regarded as the best estimate of what the next value of x will be. Given a sequence of observations which starts at the time $t = 0$ and a smoothing factor α, such as $0 < \alpha < 1$, the exponential smoothing model is given by the formulas:

$$s_0 = x_0, \tag{10}$$

$$s_t = \alpha\, x_t + (1 - \alpha)\, s_{t-1}. \tag{11}$$

5.3. Linear regression

It is an approach for modeling the relationship between a scalar dependent variable y and explanatory variables (or independent variables) denoted x. In linear regression, the relationships are modeled using linear predictor functions whose unknown parameters are estimated from the data.

In practice, there is an approximate linear relationship between the variables rather than exactly linear. This approximation can be represented by adding a non-observable variable ε, fancied as a collection of small errors. The variable is often called ”noise” and represents all other factors which influence the dependent variable y_t. Given n observations ($x_t, x_{t-1},\ldots, x_{t-(n-1)}$) called ”regressors”, ”exogenous variables” or ”independent variables” the following hypothesis describes the estimation of the future value y_t which is called a ”regressand” or ”endogenous variable”.

$$y_t = \beta_{t-1} x_{t-1} + \ldots + \beta_{t-(n-1)} x_{t-(n-1)} + \varepsilon, \tag{12}$$

where β are regression coefficients and ε is the error term or noise. This value captures all factors which influence the dependent value y_t other than the regressors x_t.

In order to estimate a linear regression model, few assumptions about the error term and data must be met (Greene, 2000):

— there is a random sampling of observations,
— the conditional mean should be zero – $E(\varepsilon \mid x) = 0$,
— there is no multicollinearity,
— the error terms should all have the same, finite variance (homoscedasticity),
— there is no autocorrelation (the error terms of different observation should not be correlated),
— the error term should be normally distributed.

In the research, the linear regression model was estimated few millions of times. Probably, for some estimations, the above-mentioned assumptions were met, but for some were not. The goal was to compare this linear method with other, more advanced methods (as a benchmark). For the same reasons, the linear regression was used in the articles of Altay (2005) and Ahangar et al. (2010).

Moreover, the linear regression was chosen for forecasting because of rapid price changes which can be observed in the analyzed time series. Moreover, this model is widely used in the scientific research, for example by Lin, Tsai (2015). For calculating linear regression forecasts the model from the scikit-learn library was used. It uses Ordinary Least Squares method to estimate the linear regression parameters.

5.4. Kalman Filter

It is the advanced statistical algorithm used widely in physics also known as linear-quadratic estimation (LQE). The model uses a series of measurements observed over time, containing statistical noise and other inaccuracies. It produces estimates of unknown variables that tend to be more precise than those based on a single measurement alone by using Bayesian Inference and by estimating a joint probability distribution over the variables for each time frame.

The Kalman filter can be written as a single equation but usually, it is described as two distinct equations which symbolize two phases: ”predict” and ”update”. The predict phase uses the state estimate from the previous time step to produce an estimate of the state at the current time step, which is also called ”a priori” state. The update phase combines the current ”a priori” prediction with the current observation information in order to refine the state estimate. Described phases are formulated by following equations:

Predict:

Predicted (a priori) state estimate

$$\hat{x}_{t|t-1} = F_{t-1}\,\hat{x}_{t-1|t-1} + B_t u_t, \tag{13}$$

Predicted (a priori) estimate covariance

$$P_{t|t-1} = F_t\,P_{t-1|t-1}F_t^T + Q_t, \tag{14}$$

Update:

Updated (a posteriori) state estimate

$$\hat{x}_{t|t} = \hat{x}_{t|t-1}\,K_t\big(y_t - H_t\hat{x}_{t|t-1}\big), \tag{15}$$

Optimal Kalman gain

$$K_t = P_{t|t-1}\,H_t^T\big(H_t P_{t|t-1}H_t^T + R_t\big), \tag{16}$$

Updated (a posteriori) estimate covariance

$$P_{t|t} = P_{t|t-1}\,(I - K_t H_t), \tag{17}$$

where:

$\hat{x}$ – estimated state,
F – state transition matrix (i.e., transition between states),
u – control variables,
B – control matrix (i.e., mapping control to state variables),
P – state variance matrix,
Q – process variance matrix (i.e., error due to process),
y – measurement variables,
H – measurement matrix (i.e., mapping measurements),
K – Kalman gain,
R – measurement variance matrix.

The complexity of the filter was explained in the paper ”Understanding the Kalman Filter” by Meinhold, Singpurwalla (1983).

5.5. ARMA – autoregressive–moving-average model

The model consists of autoregressive part $\mathrm{AR}(p)$ where p means the order, and of moving-average model $\mathrm{MA}(q)$ where q also means the order. The autoregressive part $\mathrm{AR}(p)$ uses a linear combination of its own lagged values. It is described using the following equation:

$$x_t = c + \sum_{i=1}^{p} \varphi_i x_{t-i} + \varepsilon_t, \tag{18}$$

where: φ_i are parameters of the linear model on lagged values, ε_t are errors of the model which are assumed to be independent identically distributed random variables (i.i.d.) sampled from a normal distribution with zero mean: $e_t \sim N(0, \sigma^2)$ where σ^2 is the variance.

The moving average part $\mathrm{MA}(q)$ models the error as a linear combination of error terms occurring contemporaneously in the past. It is described using the following equation:

$$x_t = \mu + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t, \tag{19}$$

where: θ_i are parameters of the linear model on past error terms, μ is the expectation of x_t.

Finally, the ARMA model consists two models described above $\mathrm{AR}(p)$ and $\mathrm{MA}(q)$ and has the following formula:

$$x_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i x_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}, \tag{20}$$

where: φ_i are parameters of the autoregressive part of the model, θ_i are parameters of the moving average.

A more detailed description of the model can be found in the work of Tsay (2002).

5.6. ARIMA model – autoregressive integrated moving average

This model is a generalization of the autoregressive moving average (ARMA) model. Non-seasonal ARIMA models are generally denoted $\mathrm{ARMA}(p, d, q)$ where parameters p, d, and q are non-negative integers and p is the order of the auto regression. This model is usually applied to the data which show evidence of non-stationarity. In such case, a differencing step can be applied to reduce the non-stationarity in the time series. The ”I” part of the model stands for ”Integrated” which means that the data were reduced by replacing non-stationary series by the difference between their values and the previous values. They can be estimated using the Box-Jenkins approach. A more detailed description of the model can be found in the research of Tsay (2002).

5.7. Logistic regression

This model measures the relationship between the categorical dependent variable and independent variables by estimating probabilities using a logistic function. It is used to estimate the probability of a binary response based on one or more predictors (or independent) variables (features). It is not a classification method, but in terms of economics, it is called qualitative response/discrete choice model. It can be seen as a special case of the generalized linear model which is analogous to linear regression, but it is based on different assumptions. The conditional distribution is a Bernoulli distribution rather than a Gaussian distribution because the dependent variable is binary. This method was developed by statistician David Cox in 1958. A detailed explanation of the logistic regression was described in the article by Yu et al. (2010).

5.8. Linear discriminant analysis

The model is a generalization of Fisher’s linear discriminant, which is used to find a linear combination of features that characterizes two or more classes of

objects. The resulting combination may be used as a linear classifier or for dimensionality reduction before conducting a classification. The method is related to ANOVA and to regression analysis, but it has continuous independent variables and a categorical dependent variable when ANOVA has categorical independent variables and a continuous dependent variable. It is widely used for face recognition problems, bankruptcy prediction and in marketing. In the research, the model LDA from the python machine learning library scikit-learn was used. It was implemented based on the description in the work of Hastie et al. (2008).

6. RECALCULATION OF MODELS AND OPTIMIZATION

6.1. Models calibration

In the article, eight forecasting methods were optimized by selecting the best hyperparameters based on their performance on historical data. Every time series out of fourteen was divided into subseries of 10000 values. After selecting the best model for particular subseries, the hyperparameters were used to conduct forecasts on the next subseries (10000 values, out-of-sample data). These two steps were repeated for the whole time series. The same procedure was applied to every time series. A model optimization for each subseries was conducted based on criteria's described below:

— the lowest AMAPE error – for models like moving average, exponential smoothing, linear regression, Kalman filter models, described in paragraphs 5.1 to 5.6,
— the highest accuracy of forecasting a direction of bid/ask price change, which is a percentage of accurate forecasts of possible up, down movements or without changes divided by the number of all forecasts. These criteria were used for logistic regression and linear discriminant analysis models described in paragraphs 5.7 and 5.8.

This approach for model calibration was chosen based on the research done by Katz, McCormick (2000).

6.2. Optimization

For the model optimization, a simple brute force algorithm was used – this approach checks all possible variants of the parameter from the set of possible combinations. For instance, for moving average parameter – window length, we start the optimization from taking 3 last prices, including the current one and then we check the following lengths: 6, 9, 12 with the step 3 until the length is smaller than 20. For every window length forecasting metrics are calculated (described in 4.1) and the best model (the optimal window length) is taken to forecast next 10000 values.

An explanation what parameters were optimized for every model can be found below:

1) moving average was trained on bid or ask prices. Only one parameter was optimized ”window length” – indicating how many quotations were used to calculate the forecasts of the model. The optimal value was selected from the set of values (3, 6, 9, 12, 15, 18).
2) linear regression model was also trained on bid or ask prices. It was assumed that endogenous variable is a predicted, next bid or ask price and exogenous prices are last n bid or ask prices, and the number of exogenous prices – n is chosen using the brute force optimization from the set of values (9, 12, 15, 18).
3) exponential smoothing model was trained on bid or ask prices. Only one parameter was optimized – a smoothing parameter called ”alpha”. The optimal value was selected from the set of values (0.1, 0.2, 0.3, 0.4, 0.5, 0.6),
4) Kalman filter was also trained on bid or ask prices. Two internal model parameters ”Q” and ”R” were optimized. Furthermore, for Kalman filter, two initial values for every calculation were taken arbitrary $P_o = 0,\ x_0 = 0$. These two parameters were selected from the set of values (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9),
5) ARMA and ARIMA models were calculated from logarithmic changes of bid/ask prices. Both models were chosen to the research due to the fact that the exchange rate time series tend to have stationary and non-stationary periods. The optimization of ARMA and ARIMA models was performed by checking the following combinations of autoregressive parts and moving average parts:
   — ARMA(p,q) – with the following autoregressive part p and moving average part q – (1, 0), (1,1), (1,2), (0,1) (2,1), (2,2), (2,0), (0,2),
   — ARIMA(p,1, q) – with the following autoregressive part p and moving average part q – (1,1,0), (1,1,1), (1,1,2), (0,1,1) (2,1,1), (2,1,2), (2,1,0), (0,1,2).
6) logistic regression and linear discriminant analysis were trained on logarithmic changes of bid/ask prices. These two machine learning models were trained by providing an input vector of logarithmic price changes with a labeled outcome based on the next bid/ask price:
   — 1 when the next bid/ask price was higher than the previous one,
   — 0 when the next bid/ask price was equal to the previous one,
   — 1 when the next bid/ask price was lower than the previous one.

The length of input vectors was optimized. The best model was selected after checking the historical performance of models trained on following input vectors lengths (3, 6, 9, 12, 15, 18). Given such training, these models were taught how to predict a direction of bid/ask price change.

Table 2 summarizes the optimization phase.

Table 2. OPTIMIZATION PARAMETERS WITH POSSIBLE VALUES

Method	Parameter	From	To	Step
Moving average	window length	3	20	3
Exponential smoothing	alpha	0.1	0.6	0.1
Linear regression	window length	9	20	3
Kalman filter	Q	0.1	0.9	0.1
	R	0.1	0.9	0.1
Logistic regression	values used for training	3	20	3
Linear discriminant analysis	values used for training	3	20	3

Surprisingly, after the optimization phase, the best forecasting results for many subseries were achieved, using the same internal model parameters like:
— for Kalman filter alpha=0.1 beta=0.1,
— for exponential smoothing alpha=0.5,
— for linear regression the minimal window length 9 observations,
— for moving average the minimal value equal 3 observations,
— for machine learning methods usually 3 last logarithmic returns were taken.

That leads to the conclusion that the last prices which are taken to calculate a forecast, tend to have the highest significance of all of the observations. Particularly, we can see that analyzing the exponential smoothing model. If it could have been possible for the alpha to take a value 1, it would have become a naive forecasting method, due to the process of the optimization. That is why the maximum value was restricted to 0.6.

7. PROBLEMS WITH FORECASTING HIGH-FREQUENCY DATA AND SOLUTION

During the research process, few problems appeared. First of them was connected with the repeated bid or ask values in the analyzed time series. It was caused by adjusting only one value of the tick while keeping the second value at the same level (i.e. ask value was changing when bid stayed the same for several ticks). This behavior can be seen in almost every analyzed time series. In such case, the data were not interrupted manually (for instance by cleansing), because of the bid and ask prices are unbreakable part of the tick, but even though it might have caused many problems during forecasting and less accurate predictions.

A solution for modeling price changes when the price is the same for several values, is developing models which can predict the fact that the price will not change in the next tick. Most often the subparts of the time series with the same prices are encountered when there is a lack of volatility, especially during the night. A predicting model which can detect such subparts of the time series can probably outperform described models.

That is the reason why it was worth to check the performance of the models forecasting a direction of bid/ask price change. In this case, the accuracy means how many times the method was correct forecasting that the price will go up, down or it will stay the same, divided by the overall number of forecasts. Given these three states (up, down, the same) it was possible to introduce machine learning algorithms from the family of supervised learning models which can be taught how to predict price changes.

In the work of Moody, Wu (1995), the authors found a relationship between the forecasting accuracy and bid/ask spreads on tick-by-tick data. Accordingly, this research was focused on finding a possible relationship between values of the Hurst exponent and the forecasting accuracy of statistical models and machine learning models. It is worth to check in further extensions of the research if there is a seasonality in intraday values of the Hurst exponent and how much it affects the forecasting results. The research conducted by Bayraktar et al. (2003) shows that a seasonality in intraday financial data (S&P 500) can affect the Hurst ratio estimation. In order to achieve the robustness of the Hurst exponent estimation, the authors suggest using wavelets with at least two vanishing moments.

8. BENCHMARKS AND FORECASTING ACCURACY MEASURES

Every forecasting technique must be evaluated before a possible application in solving a real economic problem. It is good to have some benchmarks which can be used for comparisons. Gately (1995) has shown that machine learning methods tend to have a better performance when predicting price changes rather than making point forecasts. Due to that fact, two machine learning models chosen in this research (described in paragraphs 5.7 and 5.8) were configured to predict price changes (increased, decreased, without change). This solution is recommended by the literature (i.e. Gately, 1995).

In the first research, the methods from the paragraphs 5.1 to 5.6 were analyzed and following accuracy measures MAE, MSE, RMSD, MAPE and AMAPE were calculated for every one of them[3]. The AMAPE (adjusted mean absolute percentage error) measure was used as a tool for comparing results. For comparing different currency data sets (varying descriptive statistics) AMAPE error, from all measures used to estimate how close forecasts or predictions are to the eventual outcomes, is the most accurate. It makes possible to compare currency data sets with each other in a fair manner which is described by (Doman, Doman, 2009). The metric is defined by the formula:

[3] The MAE, MSE, RMSD, MAPE results are available on public repository for every exchange rate in unstructured output from the scripts on: https://github.com/rszostakowski/phd/tree/master/1Article/resultsForEveryExchangeRate.

$$AMAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_{t+i} - \hat{y}_{t+i}}{y_{t+i} + \hat{y}_{t+i}}\right|, \tag{21}$$

where:

N – number of forecasts,
i – the length of the forecast period,
$\hat{y}_{t+i}$ – a point forecast in the future,
t – number of observations in the testing data set.

Moreover, in this part of the research, a naive forecasting was used as a benchmark. In this model, the current price is a forecast, without adjusting them or attempting to establish causal factors:

$$\hat{y}_{t+h|t} = y_t. \tag{22}$$

In the second research, the highest accuracy of forecasting a direction of bid/ask price change was used to compare methods between each other. Machine learning methods were trained to forecast a possible up, down price movements or without changes. For making it comparable to other methods, point forecasts obtained using methods from 5.1 to 5.6 were also classified into these three states, i.e. if exponential smoothing forecast indicates that the price will go up we have a classification to the “up” state. Owing to that it was possible to calculate the accuracy of every described method in the article. For benchmarking two artificial time series were created:

— ”Random up and down” – a realization of the Markov process with two possible states (price went up or down) with the equal probability 0.5 of appearance,
— ”Random up and down or the same” – a realization of the Markov process with three possible states (up or down, stayed the same) with the equal probability$\left(\frac{1}{3}\right)$ of appearance.

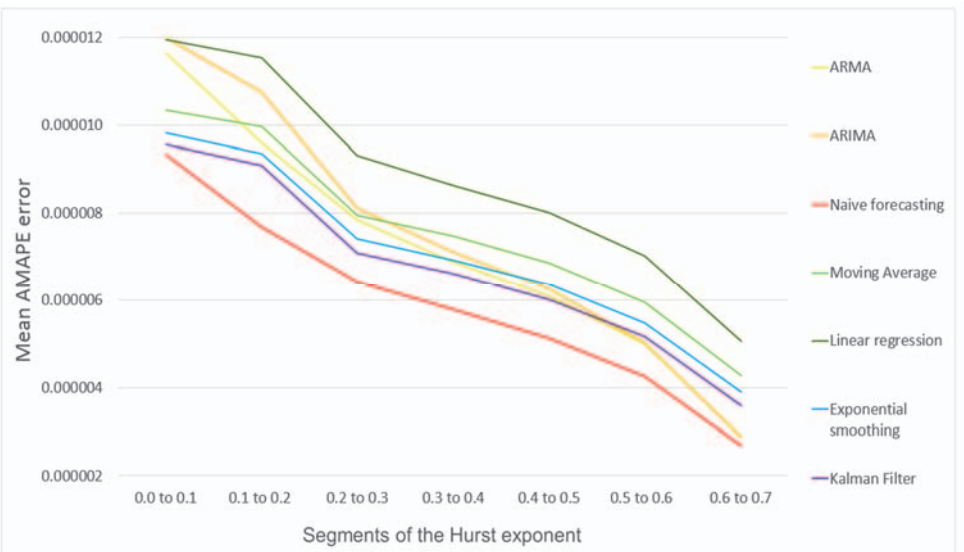
9. EMPIRICAL STUDY

9.1. Statistical models forecasts

The Hurst exponent analysis was performed for every rolling window of 150 values and the result was assigned to the last tick from these 150 values. It was conducted to check, if subseries with a higher or lower value of the Hurst exponent than 0.5 (random time series) can be used to make more ac-

curate forecasts, than for the series with a value of the Hurst exponent oscillating around 0.5. Thereafter, for every tick, the assigned value of the Hurst exponent was classified into the segments from 0 to 1 with a step 0.1. The results of the Hurst exponent classification can be seen in figure 2. Surprisingly, the values of the Hurst exponent bigger than 0.8 were not obtained. They are typical for time series with a strong trend. The majority of Hurst exponent ratios was classified into the first two segments (from 0 to 0.1 and from 0.1 to 0.2).

Figure 3. The Hurst exponent segment-based classification of forecasting accuracy measured by AMAPE, historical, tick-by-tick foreign exchange rates data from one trading day – 5pm 05-03-2012 to 5pm 06-03-2012



For every statistical model, an average adjusted mean absolute percentage error (AMAPE) was calculated to show how close forecasts of these models are to the eventual outcomes[4]. It was done to show a general performance of the forecasting methods, according to the Hurst exponent classification. The results can be seen in figure 3.

Unfortunately, average AMAPE for statistical models seems to be higher than AMAPE for naive forecasting. It proves, that the forecasting using sophisticated statistical methods on the high-frequency data cannot outperform the results of

[4] The partial results of AMAPE measures for every exchange rate are available on github repository: https://github.com/rszostakowski/phd/blob/master/1Article/ResultsWithHurstExponent.xlsx.

naive forecasting. The same conclusion can be found in the work of Cheung (1993), Green, Armstrong (2015) and Kilian, Tylor (2001). The closest method to achieving this goal was Kalman Filter. Surprisingly, the average performance of ARMA and ARIMA methods was similar for most of the exchange rates. Moreover, their forecasts were more inaccurate than the predictions of trivial methods like exponential smoothing for the Hurst ratio between 0 and 0.3 but for the Hurst exponent segments above 0.4 they increased their performance and were more accurate than any other method except for the naive forecasting. Unsatisfactory results were obtained due to a large number of data, which were repeated (bid and ask prices). The worst method for all Hurst exponent segments was the linear regression. Even strong persistence observed for the highest values of the Hurst exponent did not improve significantly the results of the linear regression model.

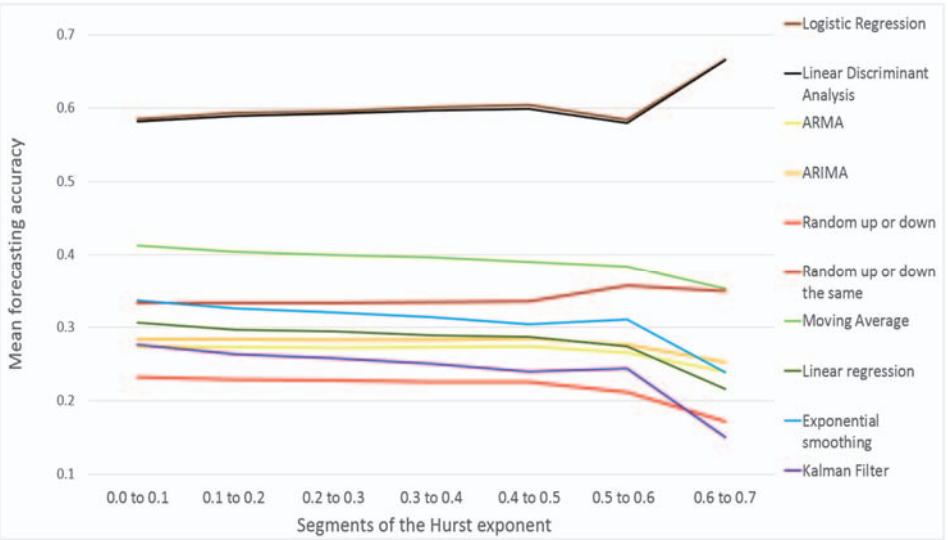
The most important discovery found in this research is the fact that the forecasting errors tend to decrease for every forecasting method when the value of the Hurst exponent increases. This observation is persistent for all forecasting models. Even for the Hurst exponent ratio close to 0.5, which indicates a random time series, better forecasts were obtained than for values of the Hurst exponent close to 0 (a mean-reverting time series). This fact proves the hypothesis of the research that forecasting methods are more accurate when the value of the Hurst exponent increases.

9.2. Forecasting a direction of bid/ask price change

In the second part of the research, an alternative approach to forecasting a point value was introduced. Forecasting a direction of bid/ask price change was chosen to overcome some of the drawbacks of the previous approach. For every method, an average accuracy of forecasting a direction of bid/ask price change was calculated and classified to the Hurst exponent ranges[5]. It is worth to mention that for time series without many repeated values a benchmark process “random up and down” should have an average accuracy of forecasting close to 0.5. As we can see in figure 4 a higher mean forecasting accuracy was obtained for the “random up and down or the same” process which indicates that analyzed times series have many repeated bid/ask prices. It is approximately the one-third of ticks. The second process slightly improves its performance for values of the Hurst exponent above 0.6 due to the fact that the statistical sample for these segments is too small (figure 2).

[5] The accuracy of forecasting a direction of bid/ask price change for every exchange rate are available on github repository: https://github.com/rszostakowski/phd/blob/master/1Article/ResultsWithHurstExponent.xlsx.

Figure 4. The Hurst exponent segment-based classification of mean accuracy of forecasting a direction of price change, historical, tick-by-tick foreign exchange rates data from one trading day – 5pm 05-03-2012 to 5pm 06-03-2012



In this part of the research, the most accurate forecasting results were obtained by models based on machine learning techniques – the logistic regression and the linear discriminant analysis. For both of them, the average accuracy was close to 0.6, which means that in average the machine learning models were accurate in predicting 60% price directions. Such good results indicate that this approach can be successfully applied in liquidity forecasting or in active investment. Moreover, in most cases, the moving average model outperformed the random processes and other statistical models like ARMA and ARIMA. The most sophisticated statistical method like Kalman Filter appeared to be inaccurate. It might have been caused by too narrow model calibration.

Interestingly, for GBP/PLN ask and bid series the average forecasting accuracies were higher than for any other currency. Moreover, for these two series even models like Kalman filter, moving average and exponential smoothing were more accurate than the naïve forecasting. It might have been caused by a lower liquidity observed on GBP/PLN exchange rate. It means that almost every new tick in the series was changing both of the prices (bid and ask) simultaneously.

This part of the research has revealed that the statistical models tend to have a contrary tendency to the one, discovered in the first part of the research. For them, an average accuracy of forecasting a direction of bid/ask price change decreases when the value of the Hurst exponent increases. Fortunately, this tendency is not observed in the forecasting results of machine learning methods. These methods seem to benefit from the increase of the Hurst exponent – the

bigger the value of the Hurst exponent the higher their average performance. That indicates that a further research into statistical learning methods should be pursued.

10. CONCLUSIONS AND FURTHER RESEARCH

In this paper, seven different exchange rates series were tested using the Hurst exponent analysis (full order book, bid and ask prices). For most of the data, the values of the Hurst exponent oscillated around 0.2. These results differ from the study of Mitra (2012), who has shown that the Hurst exponent calculated from daily returns of stock market indices oscillates around 0.5.

Unfortunately, the first part of the research revealed that analyzed statistical methods tend to be more inaccurate than the naive forecasting on high--frequency-data. Moreover, their average AMAPE was higher than the average adjusted mean absolute percentage error of naïve forecasting for all Hurst exponent segments. From the other side, the results from the second part of the research were far more satisfying. The machine learning methods used for predicting a direction of bid/ask price change were more precise than the random process and the statistical methods. Finally, none of the methods were able to show a significant increase in their performance for mean reverting parts of the time series.

The most important research discovery is the fact that the average forecasting errors tend to decrease with an increase of the Hurst exponent for statistical and machine learning methods. Moreover, for the values of the Hurst exponent close to 0.5 forecasts from all analyzed models, seem to be more accurate than for the values which indicate a mean reverting time series (from 0.1 to 0.2). A positive effect of analyzing the Hurst exponent in forecasting was observed in the research paper of Mitra (2012) for time series with the Hurst exponent bigger than 0.5. Unfortunately, the author did not analyze the effect of the Hurst exponent below 0.5 on forecasting, which was done in this article.

Furthermore, for machine learning methods the accuracy of forecasting a direction of price change tends to increase when the value of the Hurst exponent increases. According to the author's best knowledge, this approach has not been analyzed in any other research paper. Finally, only machine learning methods discovered nonlinear structures hidden in the data and proved to be successful in predicting price changes. Their accuracy usually oscillated around 60%.

Forecasting high-frequency data of exchange rates seems to be very complex, due to the currency market characteristics (volume, number of investors). The research has shown several areas for an improvement in the art of forecasting which tend to follow in three directions.

First of them, is the path of analyzing market characteristics. It is worth to try creating time series regimes when it is highly probable to outperform the naive forecasting based on other metrics than the Hurst exponent. A second path leads to developing better forecasting models. Vengertsev (2014) and Ghahramani (2001) have shown that more advanced machine learning methods like support vector machine or deep learning models which can be successfully applied in the forecasting. Moreover, based on the papers written by Menkhoff, Taylor (2007) and Osler (2003) it is worth to apply technical analysis indicators as input vectors to machine learning models.

Finally, it is worth to try to develop a statistical model which would detect the fact, that a bid or ask price can be unchanged in the future tick. It can increase forecasting performance and lead to outperforming the naive forecasting model.

REFERENCES

Ahangar R. G., Yahyazadehfar M., Pournaghshband H., (2010), The Comparison of Methods Artificial Neural Network with Linear Regression Using Specific Variables for Prediction Stock Price in Tehran Stock Exchange, *arXiv preprint arXiv*:1003.1457.

Ahmed N. K., Atiya A. F., El Gayar N., El-Shishiny H., (2010), An Empirical Comparison of Machine Learning Models for Time Series Forecasting, *Econometric Reviews*, 29 (5), 594–621.

Aiolfi M., Timmermann A., (2006), Persistence in Forecasting Performance and Conditional Combination Strategies, *Journal of Econometrics*, 1–2, 31–53.

Altay E., Satman M. H., (2005), Stock Market Forecasting: Artificial Neural Network and Linear Regression Comparison in an Emerging Market, *Journal of Financial Management and Analysis* 18 (2), 18–33.

Andersen T. G., Bollerslev T., (1997), Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns, *The Journal of Finance,* 52 (3), 975–1005.

Bayraktar E., Poor H. V., Sircar K. R., (2003), Efficient Estimation of the Hurst Parameter in High Frequency Financial Data with Seasonalities using Wavelets, *Computational Intelligence for Financial Engineering, Proceedings of the IEEE 2003 International Conference*, 309–316.

Brown R. G., (2004), *Smoothing, Forecasting and Prediction of Discrete Time Series*, Dover Publications, Dover Phoenix Ed edition.

Castillo O., Melin P., (1996), Automated Mathematical Modelling for Financial Time Series Prediction Using Fuzzy Logic, Dynamical Systems and Fractal Theory, *Computational Intelligence for Financial Engineering, Proceedings of the IEEE/IAFE* 1996 *International Conference*.

Cheung Y.-W., (1993), Long Memory in Foreign-Exchange Rates, *Journal of Business & Economic Statistics* 11 (1), 93–101.

Doman M., Doman R., (2009), *Modelowanie zmienności i ryzyka*, Wolters Kluwer Polska Sp. z.o.o, Warszawa.

Fama E., (1970), Efficient Capital Markets: a Review of Theory and Empirical Work. *Journal of Finance*, 25, 2, 383–417

Gately E., (1995), *Neural Networks for Financial Forecasting*, John Wiley & Sons Inc.

Ghahramani Z., (2001), An Introduction to Hidden Markov Models and Bayesian Networks, *International Journal of Pattern Recognition and Artificial intelligence*, 15 (1), 9–42.

Green K. C., Armstrong J. S., (2015), Simple Versus Complex Forecasting: The Evidence, *Journal of Business Research*, 68, 8, 1678–1685.

Greene W. H., (2000), *Econometric Analysis,* International edition.
Hastie T., Tibshirani R., Friedman J., (2008), *The Elements of Statistical Learning*, Section 4.3, 106–119.
Huang B.-N., and Yang C. W., (1999), An Examination of Long-Term Memory Using the Intraday Stock Returns, *Technical Report* 99–03, Clarion University of Pennsylvania, Clarion.
Hurst H. E., (1951), Long-Term Storage Capacity of Reservoirs, *Transactions of American Society of Civil Engineers,* 116, 770–799.
Katz J. O., McCormick D., (2000), *The Encyclopedia of Trading Strategies,* McGraw-Hill.
Kilian L., Taylor M., (2001), Why Is It so Difficult to Beat the Random Walk Forecast of Exchange Rates, Europan Central Bank, *Working paper Series, 88*.
Lin T., Tsai C. C., (2015), A Simple Linear Regression Approach to Modeling and Forecasting Mortality Rates, *Journal of Forecasting,* 34, 543–559.
Meinhold R. J., Singpurwalla N. D., (1983), Understanding the Kalman Filter, *The American Statistician*, 37, 2, 123–127.
Menkhoff L., Taylor M., (2007), The Obstinate Passion of Foreign Exchange Professionals: Technical Analysis, *Journal of Economic Literature,* 45 (4), 936–972.
Mitra S. K., (2012), Is Hurst Exponent Value Useful in Forecasting Financial Time Series, *Asian Social Science,* 8 (8), 111–120.
Moody J., Wu L., (1995), Price Behavior and Hurst Exponents of Tick-By-Tick Interbank Foreign Exchange Rates, *Computational Intelligence for Financial Engineering*, Proceedings of the IEEE/IAFE 1995.
Osler C., (2003), Currency Orders and Exchange Rate Dynamics: An Explanation for the Predictive Success of Technical Analysis, *The Journal of Finance*, 7 (5), 37–58.
Peters E., (1991), *Chaos and Order in the Capital Markets*, John Wiley & Sons Inc.
Peters E., (1994), *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*, John Wiley & Sons Inc.
Qian B., Rasheed K., (2004), Hurst Exponent and Financial Market Predictability, *Proceedings of the IASTED conference on Financial Engineering and Applications* (FEA '04), 203–209.
Tsay R. S., (2002), *Analysis of Financial Time Series Financial Econometrics*, John Wiley & Sons Inc.
Vengertsev D., (2014), Deep Learning Architecture for Univariate Time Series Forecasting, CS229 Technical Report, 3–7.
Wold H., (1939), A Study in the Analysis of Stationary Time Series, *Journal of the Royal Statistical Society*, 102 (2), 295–298.
Yu H. F., Huang F. L., Lin C. J., (2010), Dual Coordinate Descent Methods for Logistic Regression and Maximum Entropy Models, *Machine Learning Journal,* 85, 1, 41–75.

JAKOŚĆ PROGNOZOWANIA CEN W ZALEŻNOŚCI OD WYKŁADNIKA HURSTA PRZY WYKORZYSTANIU DANYCH WYSOKIEJ CZĘSTOTLIWOŚCI Z RYNKU WALUTOWEGO

Streszczenie

Na przestrzeni ostatniego wieku przeprowadzono wiele badań na temat użyteczności metod statystycznych w prognozowaniu cen na rynkach finansowych. Niniejszy artykuł wyjaśnia, dlaczego większość z nich zawiodła, bazując na teorii rynku fraktalnego oraz na podstawie badań przeprowadzonych przy użyciu da-

nych wysokich częstotliwości z głównych par walutowych. Dla wykorzystanych modeli statystycznych i metod uczenia maszynowego zostały policzone miary takie jak AMAPE oraz trafność prognozowania kierunku zmian cen w zależności od wykładnika Hursta. Artykuł pokazuje, że średni błąd prognozowania zmniejsza się wraz ze wzrostem wartości wykładnika Hursta dla zastosowanych modeli prognostycznych. Zaprezentowana w artykule metodyka prognozowania może być skutecznie wykorzystana do podejmowania trafniejszych decyzji inwestycyjnych oraz do budowy automatycznych systemów decyzyjnych.

Słowa kluczowe: dane wysokich częstotliwości, prognozowanie, uczenie maszynowe, metody statystyczne, mikrostruktura rynku, wykładnik Hursta

THE USE OF THE HURST EXPONENT TO INVESTIGATE THE QUALITY OF FORECASTING METHODS OF ULTRA-HIGH-FREQUENCY DATA OF EXCHANGE RATES

Abstract

Over the last century a variety of methods have been used for forecasting financial time data series with different results. This article explains why most of them failed to provide reasonable results based on fractal theory using one day tick data series from the foreign exchange market. Forecasting AMAPE errors and forecasting accuracy ratios were calculated for statistical and machine learning methods for currency time series which were divided into sub-segments according to Hurst ratio. This research proves that the forecasting error decreases and the forecasting accuracy increases for all of the forecasting methods when the Hurt ratio increases. The approach which was used in the article can be successfully applied to time series forecasting by indicating periods with the optimal values of the Hurst exponent.

Keywords: high-frequency data, forecasting, machine learning, statistical models, microstructure, Hurst exponent

DOI: 10.5604/01.3001.0014.0537

PRZEGLĄD STATYSTYCZNY
TOM LXV – ZESZYT 2 – 2018

Iwona BĄK[1]
Katarzyna CHEBA[2]

Study of spatial uniformity of sustainable development of the European Union before, during and after the economic crisis[3]

1. INTRODUCTION

Dynamic changes noted on the world markets, which are predominantly connected with economic slowdown coerce the consideration of the uniformity of development of each regions. A particularly interesting area of research in this field is the impact analysis of social and economic development for example on the basis of sustainable development indicators: before, during and after the economic slowdown in 2007–2008. The analyses of that type allow to track changes in individual EU countries, forming a single organism, but they are characterized by differing levels of development, with different resistance to the crisis of 2007––2008 and often completely different social and economic realities.

The basic question we ask, whether it is possible to talk about balancing the socio-economic development in the European Union? Posing such questions is particularly important in the case of such political and economic structures such as European Union. The basic, strategic developmental objectives of the European Union include the aspiration to harmonious development of all of its members, however, it is extremely difficult task as both the statistical data and the operational experience prove. A separate, and extremely important issue is the measuring of homogeneity (heterogeneity) of particular regions of the European Union.

The purpose of the paper is study of spatial uniformity in the field of sustainable development of European Union and geographical regions of Europe ana-

[1] West Pomeranian University of Technology Szczecin, Faculty of Economics, Department of Applied of Mathematics in Economy, 31 Janickiego St., 71–270 Szczecin, Poland, corresponding author – e-mail: iwona.bak@zut.edu.pl.

[2] West Pomeranian University of Technology Szczecin, Faculty of Economics, Department of Applied of Mathematics in Economy, 31 Janickiego St., 71–270 Szczecin, Poland.

[3] Research results presented in this paper are an element of research project implemented by the National Science Center Poland under the grant OPUS13 no UMO-2017/25/B/HS4/02172.

lyzed by the prism of EU countries located in these regions before, during and after the economic crisis in 2007–2008. In the work to study the spatial differentiation of social and economic development of European Union, on the basis of sustainable development indicators presented by Eurostat, the taxonomic measure of development based on median Weber vector as well vector calculus were used. The previous study by the authors (Bąk, Cheba, 2017) confirmed the existence of significant heterogeneity of spatial development of individual geographical regions of the European Union. Therefore, further research will concentrate on studying the applications of discussed methods will be based on data sustainable development indicators, analyzed separately before, during and after the period of economic slowdown. The results presented in the work will contribute to increasing knowledge about methods testing homogeneity (heterogeneity) of the development in the regional aspect and methods showing the direction of the analyzed changes in the situation of economic crisis.

The paper is organized as follows: the second part describes the methodological issues of the empirical analysis presented in the paper, including indicators and statistical methods description. The following part presents study results which were divided into two topics: results of EU Member States' ranking in the field of sustainable development and uniformity of the balanced development of the European Union. The final part of the article put forward conclusions.

2. METHODOLOGICAL ISSUES OF THE EMPIRICAL ANALYSIS

2.1. Background

The question how to measure the social and economic development is particularly important in the face of growing crises that have economic, political or social origins (Peacock et al., 1988; Rigobon, 2003; Lopez, 2005; Autor et al., 2008; Klenert et al., 2015; Kobayashi, Shirai, 2016; Moomaw et al., 2017). As it has been known for a long time that classical measures of economic development don't reflect well enough the actual development of countries, the measures that describe also the qualitative aspects of their prosperity (including the social and environmental ones) have been sought (Eagle et al., 2010). One of the studies' direction on new economic development measures was the idea of sustainable development which was born in response to the criticism of overexploitation of natural environment that led to increased global threat of natural disasters (Duran et al., 2015). Information about risks related with excessive use of natural environment had been published in U Thant's report of 1967 (Meadows, 1973). These threats, particularly the ones related with the depletion of natural resources and the degradation of ecosystems are also mentioned in the 1972 Club of Rome report *The Limits to growth* (Berger, Zwirner, 2008). While, the concept of sustainable development was first formulated explicitly during the

Third UNEP Program in 1975, as ”(...) such a course of inevitable and desirable economic development that would not materially and irreversibly affect the human environment and would not lead to the degradation of the biosphere and would not undermine the laws of nature, economics and culture” (UN, 1975). Next this concept was presented in 1987 in ”Our Common Future”, a publication also known as Brundtland Report. The Report was created by the UNO commission established with the intention of developing a global programme of changes in the concept and practice of development. It states that the rapid growth of civilisation, equated to the increased general well-being, leads to overexploitation of natural resources and, in effect, can endanger the global ecosystem. In this report sustainable development was defined as ”sustainable development to meet current needs without the risk that future generations will not be able to meet their needs” (WCED, 1987).

The idea of sustainable development is not contradictory to the growth in prosperity. However, the emphasis is on the optimalisation of economy with simultaneously minimised consumption of raw materials, energy and water as well as the reduced human environmental impact. Consequently, the principal rule of the sustainable development is the need to address the three pillars: the society, economy and environment (van den Bergh, Hofkes, 1998; Hopwood et al., 2005). Also, this concept points to the need to cross both the institutional and geographical borders in order to coordinate strategies and make proper decisions in the framework of the cooperation of governmental agencies from different countries. It means that it is necessary to look at the current problems faced by the European Union not only from the Union’s or individual countries’ perspective, but also from the perspective of individual regions that are functionally or geographically related. More than a decade after the first EU enlargement following the accession of the East European countries in 2004, the divisions within the EU, such as distinguishing between the old and new EU Member States, still seem to exist. These divisions are also noticeable when we compare the indices of the EU sustainable development changed before, during and after the economic slowdown in 2007–2008. The European Commission announces the results of monitoring the sustainable development (SD) indices on the biannual basis. Its latest report was published in 2015 (European Union, 2015). The implementation of the EU Sustainable Development Strategy (EU SDS) is monitored by means of the sustainable development indicators (SDI) published by Eurostat. Until recently (the change of the way of SDI presentation according the Agenda 2030 took place on 15.11.2017) the SDIs had a hierarchic structure whose components were divided into three levels. At the top there were 11 Headline Indicators that were intended to give an overall picture of the progress in terms of the key challenges of the EU SDS. The second level was represented by 31 Operational Indicators that related to the operational objectives of the strategy, while on the third, lowest level there were 84 Explanatory Indicators

that illustrated the progress of the actions described in the SDS. In this paper these indicators were used to study the spatial differentiation of sustainable development of European Union countries.

The first step of these studies in the area of sustainable development is usually the analysis of the EU achievements in subsequent years and the assessment of their compliance with the strategic targets. In spite of dynamic changes in individual areas of the EU sustainable development, it is necessary to analyse as well the internal homogeneity of the EU in this aspect. The majority of published studies (Mulder, van den Bergh, 2001; Stefanescu, On, 2012, Boda et al., 2015; Gnimassoun, Mignon, 2015) are based on the assessment if the EU is moving towards the adopted targets. Therefore, the authors concentrate more on assessing the existing level of sustainable development than on the very process of balancing the sustainable development. However, the analysis of internal imbalances among member states proves that developmental differences are significant. The inequalities exist both on the international and regional level. Information obtained from the analyses of individual SD indicators, concerning both individual countries and geographical regions, was used in a study on the SD level in the EU countries in 2004, 2008 and 2014. Apart from the above analyses, the purpose of which was to assess the impact of the 2007/2008 crisis on the sustainable development in individual EU Member States, the collected data helped conduct a spatial analysis of the SD distribution across the EU geographical regions and their countries.

2.2. Objectives, scope and methodology of the study

The objectives of the study were to find an answer to the following questions:

1. Is it possible to talk about balancing the sustainable development in the European Union?
2. How big is the unevenness of sustainable development of particular UE regions, namely:
   a) How spatially homogeneous (heterogeneous) are those regions?
   b) Are the identified changes in time homogenous (heterogeneous)?
3. How has the position of the European Union countries in the field of sustainable development changed before, during and after the economic slowdown in 2007–2008?

The analysis of similarities and differences between the European Union countries was based on sustainable development indicators at the EU level (Eurostat, 2017). At the beginning of the study database was set up. In the paper SD indicators presented by Eurostat were used. The original data base included 47 indicators describing 12 themes of the European sustainable development from 2004, 2007 and 2014.

In the next step diagnostic features were selected for the study. The selection criteria is usually divided into two groups: the content related and formal/statistical ones. In the first approach the set of diagnostic features contains such values that, according to the obtained knowledge about the phenomena under study, are the most typical of the compared objects. In the second approach, the selection of features follows a specific formal procedure. The most appropriate is a two–stage selection procedure where both approaches are simultaneously used. After defining and gathering data concerning the initial set of features, proper verification actions are usually performed against two most important criteria:

1. Variability – the features should be diverse, i.e. effectively discriminating the objects.

   To assess the variability, a diversity coefficient, calculated from the formula (1), is used:

$$V_j = \frac{S_j}{\bar{x}_j}, \tag{1}$$

   where: $\bar{x}_j$ – arithmetic mean of X_j, S_j value – standard deviation of *j*-th feature X_j, $j = 1, 2, \ldots m, m$ – number of features.

   Taking into account the former of this criterion, 6 diagnostic features were eliminated from the study, because the coefficients of variation calculated for them were low throughout the whole period of study (at 10% or lower).
2. Correlation – two strongly correlated features carry similar information; therefore one of them is redundant. For this reason, the correlation indicators of all the features should be taken into account, and then, the most suitable verification method should be applied to eliminate features most similar to others. The starting point here is to create a matrix of feature correlations:

$$R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{21} & 1 & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1} & r_{m2} & \cdots & 1 \end{bmatrix}, \tag{2}$$

   where: r_{jk} – the Pearson linear correlation coefficient of the j^{th} and k^{th} feature.

In the next step, the matrix of correlations among the features was constructed for every analysed year separately. When examining the similarity of the features by means of the coefficients of variation, it was found that some indicators

were very strongly correlated. Therefore, the formal approach, a parametric method proposed by Hellwig (Nowak, 1990)[4] was used to select a final set of diagnostic features. The starting point in this method is the matrix of the coefficients of correlation (formula 2) among the potential diagnostic features. The classification criterion is the parameter r^* also called a critical value of the correlation coefficient so that $0<r^*<1$. The value of r^* can be chosen at the researcher's discretion or determined in a formal way[5].

The features from the preliminary list can be similar to one another due to their strong correlation, hence they can form clusters. The clusters are such subsets of features where the least similarity between them is not smaller than r^*. The clusters contain one central features and several satellite ones. A satellite feature of an individual central feature is the one whose similarity to the central feature is not smaller than r^*. The features form a cluster if they consist of one central feature and at least one satellite feature. Then, they are called systemic features. The features that are not attributed to any cluster are called isolated features. The central and isolated features create the so called base configuration of features and they are considered to be diagnostic features. According the proposal of Zeliaś (2000) the final set of data was created by the features (both central and satellite) whose frequency of occurrence was the highest in the whole analyzed period.

To the final set of diagnostic features, which has become a basis for further empirical studies, the following indicators have been selected[6]:

a) in the area of socio-economic development (3 indicators): young people neither in employment nor in education or training (NEET) (15–24 years), % of the total population in the same age group – (x_{1aDO}); total R&D expenditure, % of GDP – (x_{2aSE}); total unemployment rate, % – (x_{3aDE});
b) in the area of sustainable consumption and production (2 indicators): generation of waste excluding major mineral wastes, kg per capita – (x_{4bDO}); final energy consumption[7], 1000 tons of oil equivalent – (x_{5bDE});

[4] It is the most commonly used method of diagnostic characteristics selection. However, the method is not perfect: it is sensitive to outliers (or asymmetric distribution of variables) and it takes into account only direct relationships of a given characteristic with other ones, ignoring indirect relationships. Improved resistance of the method to outliers can be achieved by replacing in the first step the sum of elements in a column (or a row) of the correlation coefficient matrix by their median. The second fault can be eliminated by means of the matrix inverse method (Nowak, 1990).

[5] In the paper it was assumed: $r = 0.5$.

[6] Symbols:, a, b, c, d…j – denote the SD theme, S denotes the stimulant, D – the destimulant, while the symbols H, O, E, C – the indicator level: H – headline indicator, O – operational, E – explanatory and C – contextual.

[7] According to the Eurostat: "this indicator expresses the sum of the energy supplied to the final consumer's door for all energy uses. It is the sum of final energy consumption in industry, transport, households, services, agriculture, etc. Final energy consumption in industry covers the consumption in all industrial sectors with the exception of the "Energy sector".

c) in the area of social inclusion (4 indicators): early leavers from education and training[8], % – (x_{6cDO}); tertiary educational attainment, by sex, age group 30–34, % – (x_{7cSO}); long–term unemployment rate – (x_{8cDE}); adult participation in learning (lifelong learning)[9], % – (x_{9cSE});
d) in the area of demographic changes (3 indicators): employment rate of older workers, % – (x_{10dSH}); total fertility rate, number of children per woman- (x_{11dSE}); old-age dependency ratio[10], per 1000 persons of working age (15––64) – (x_{12dDC});
e) in the area of public health (1 indicator): life expectancy at birth of males, years – (x_{13eSH});
f) in the area of climate change and energy (4 indicators): primary energy consumption[11], million TOE (tons of oil equivalent) – (x_{14fDH}); share of renewable energy in gross final energy consumption, % – (x_{15fSO}); electricity generated from renewable sources, % – (x_{16fSE}); share of renewable energy in fuel consumption of transport, % – (x_{17fSE});
g) in the area of sustainable transport (2 indicators): energy consumption of transport relative to GDP, index (2010–100%) – (x_{18gDH}); energy consumption by transport mode – road transport, 1000 tons of oil equivalent – (x_{19gDE});
h) in the area of natural resources: no indicators;
i) in the area of global partnership (1 indicator): CO_2 emissions per inhabitant in the EU and in developing countries, tons – (x_{20iDE});
j) in the area of good governance (2 indicators): shares of environmental taxes in total tax revenues from taxes and social contributions, % – (x_{21jDO}); level of citizens´confidence in EU institutions (for sub-theme policy coherence and effectiveness), % – (x_{22jSO}).

[8] According to the Eurostat: „the indicator is defined as the percentage of the population aged 18–24 with at most lower secondary education and who were not in further education or training during the last four weeks preceding the survey. Lower secondary education refers to ISCED (International Standard Classification of Education) 2011 level 0–2 for data from 2014 onwards and to ISCED 1997 level 0–3C short for data up to 2013. The indicator is based on the EU Labour Force Survey".

[9] According to the Eurostat: „the indicator is defined as the percentage of the population aged 18––24 with at most lower secondary education and who were not in further education or training during the last four weeks preceding the survey. Lower secondary education refers to ISCED (International Standard Classification of Education) 2011 level 0–2 for data from 2014 onwards and to ISCED 1997 level 0–3C short for data up to 2013. The indicator is based on the EU Labour Force Survey".

[10] According to the Eurostat: "this indicator is the ratio between the number of persons aged 65 and over (age when they are generally economically inactive) and the number of persons aged between 15 and 64. The value is expressed per 100 persons of working age (15–64)".

[11] According to the Eurostat: by "Primary Energy Consumption" is meant the Gross Inland Consumption excluding all non-energy use of energy carriers (e.g. natural gas used not for combustion but for producing chemicals). This quantity is relevant for measuring the true energy consumption and for comparing it to the Europe 2020 targets. The "Percentage of savings" is calculated using these values of 2005 and its forecast for 2020 targets in Directive 2012/27/EU; the Europe 2020 target is reached when this value reaches the level of 20%".

The set of diagnostic indicators chosen for the description of the compared objects can contain the variables whose influence on the phenomenon under study has different direction, i.e. stimulants and destimulants. The stimulants are variables whose bigger values indicate a higher level of progress of a given phenomenon, while the destimulants are diagnostic characteristics whose smaller values indicate a higher level of development[12] (Nowak, 1990). The classification of diagnostic characteristics selected for the study into stimulants and destimulants is shown in table 1.

Table 1. DIVISION OF DIAGNOSTIC FEATURES INTO STIMULANTS AND DESTIMULANTS

Stimulants	Destimulants
x_{2aSE}, x_{7cSO}, x_{9cSE}, x_{10dSH}, x_{11dSE}, x_{13eSH}, x_{15fSO}, x_{16fSE}, x_{17fSE}, x_{22jSO}	x_{1aDO}, x_{3aDE}, x_{4bDO}, x_{5bDE}, x_{6cDO}, x_{8cDE}, x_{12dDC}, x_{14fDH}, x_{18gDH}, x_{19gDE}, x_{20iDE}, x_{21jDO}

S o u r c e: own elaboration.

2.3. Description of used mathematical methods

In the work to study the spatial differentiation of development of individual countries in the European Union, on the basis of sustainable development indicators selected to the study[13], the following methods were used: a) taxonomic measure of development based on median Weber vector and b) vector calculus.

The linear assignment of European countries and defining typological groups of objects was conducted using the method based on the median Weber (1971) vector[14]. The positional option of the linear object assignment takes a different normalization formula, in comparison with the classical approach, based on a quotient of the feature value deviation from the proper coordinate of the Weber median and a weighed absolute median deviation, using the Weber median (Lira et al., 2002; Młodak et al., 2016)[10]:

$$z_{ij} = \frac{x_{ij} - \theta_{0j}}{1{,}4826 \cdot m\tilde{a}d(X_j)}, \tag{3}$$

where: $(\theta_{01}, \theta_{02}, \ldots, \theta_{0m})$ is the Weber median, $m\tilde{a}d(X_j)$ is the absolute median deviation, in which the distance from the features to the Weber vector is meas-

[12] Sometimes the category of *nominants* is used. In their case the most favourable situation is when they reach a fixed value or number interval.

[13] The median Weber vector was calculated on the basis of features by transforming destimulants into stimulants on the basis of the following formula: $x'_{ij} = c - x_{ij}, i = 1, 2, \ldots, n, c = 0$.

[14] The median Weber is a multi-dimensional generalization of the classical notion of the median. It is a vector that minimizes the sum of Euclidean distance (Euclidean distance) of the data points representing the considered objects, and therefore is somehow ”in the middle” of them, but it is also robust to the presence of outliers (Młodak, 2006, 2014).

ured[15], i.e.: $m\tilde{a}d(X_j) = \underset{i=1,2,\ldots,n}{\text{med}}|x_{ij} - \theta_{0j}|$ $(j = 1, 2, \ldots, m)$. The synthetic measure μ_i is calculated on the basis of maximum values of normalized features, similarly to the Hellwig (1968) method:

$$\varphi_j = \max_{i=1,2,\ldots,n} z_{ij}, \tag{4}$$

according to the following formula:

$$\mu_i = 1 - \frac{d_i}{d_-}, \tag{5}$$

where: $d_- = \text{med}(\mathbf{d}) + 2{,}5\text{mad}(\mathbf{d})$, where $\mathbf{d} = (d_1, d_2, \ldots, d_n)$ is a distance vector calculated using the formula: $d_i = \underset{j=1,2,\ldots,m}{\text{med}}|z_{ij} - \varphi_j| \cdot i = 1, 2, \ldots, n, \varphi_j -$ the i-th coordinate of the development pattern vector, which is constituted of the maximum values of the normalized features.

The assignment of objects with a positioning measure is the basis for a division of objects into four classes. The most commonly used grouping method in the positioning scope is called the *three medians method*. It involves indicating a median of vector coordinates $\mu = \mu_1, \mu_2, \ldots, \mu_n$, which is denoted $med(\mu)$, then dividing the population of objects into two groups Ω_k: those, for which the measure values exceed the median (are higher than it – Ω_1) and those, for which the measure values do not exceed the median (are equal or lower than it – Ω_2). Next the indirect medians are defined as: $\text{med}_k(\mu) = \underset{i:\Gamma_i \in \Omega_k}{\text{med}}(\mu_i)$, where k = 1, 2. This way the following groups of objects are created[16]:

— Group I: $\mu_1 > \text{med}_1(\mu)$,
— $\mu_i > \text{med}_1(\mu)$,
— Group II: $\text{med}(\mu) < \mu_i \leq \text{med}_1(\mu)$,
— Group III: $\text{med}_2(\mu) < \mu_i \leq \text{med}(\mu)$,
— Group IV: $\mu_1 \leq \text{med}_2(\mu)$.

The vector calculus was used for the examination of homogeneity of the European Union. The theoretical foundations of vector calculus and the proposal of its implementation with regard to the examination of the level of development of socio-economic objects were presented for instance in the publications by Nermend (2009) and Nermend, Tarczyńska-Łuniewska (2013). This method is characterized by the high level of flexibility, especially in the case of vector

[15] The Weber median was calculated in *R program: l1median* of package: *pcaPP*.

[16] Groups equinumerous are getting when the number of objects in the community is divisible by four.

measure constructed on the basis of scalar product and the arithmetic of the increase proposed by Borawski (2012). It allows to achieve additional information about the uniformity of diagnostic objects included in the analysed object (in the paper considered by the prism of the countries located in European geographical regions).

The vector calculus, depending on the adopted manner of computation of incremental standard deviation and/or the increment of variance might be implemented to research: a) spatial homogeneity of a set of elements located on a bigger spatial unit, for example the homogeneity of EU Member States located over a bigger region and b) time homogeneity of identified changes, for instance over the years. Calculations using synthetic vector measure starts with the designation of so-called ordered twos, which are used for further calculations instead of actual values. These twos form: the mean and the standard deviation and the mean and the variance. In the case of testing the spatial homogeneity of the objects, the values of the analyzed indicators for smaller objects (subobjects, in the work: EU countries) located in the bigger area (in the work: in geographical regions of Europe) are taken into account and mean value $(\underset{j}{\eta_i})$, standard deviation $(\underset{j}{\sigma_i})$ and the variance $(\underset{j}{\sigma_i}^2)$ are computed on the basis on following formulas:

a) mean value $(\underset{j}{\eta_i})$:

$$\underset{j}{\eta_i} = \frac{\sum_{k=1}^{N} \underset{j}{x_{i,k}}}{N}, \tag{6}$$

where: $\underset{j}{\eta_i}$ – the mean value of i-th feature for j-th object, N – the number of objects considered in study of spatial homogeneity for j-th object, $\underset{j}{x_{i,k}}$ – the value of i-th feature for k-th subobject in j-th object,

b) standard deviation $(\underset{j}{\sigma_i})$:

$$\underset{j}{\sigma_i} = \sqrt{\frac{\sum_{k=1}^{N} \left(\underset{j}{x_{i,k}} - \underset{j}{\eta_i} \right)^2}{N}}, \tag{7}$$

c) variance $(\underset{j}{\sigma_i}^2)$:

$$\sigma_{\substack{i\\j}}^2 = \frac{\sum_{k=1}^{N}\left(x_{\substack{i,k\\j}} - \eta_{\substack{i\\j}}\right)^2}{N}. \tag{8}$$

Mean and standard deviation as well as mean and variance are determined on the basis of the values of the analyzed indicators for smaller objects located on the area of larger objects form ordered twos, and the calculations for them are performed in parallel.

The next step is to determine increases based on which further calculations are conducted. Similar calculations are performed also for a pair consisting of mean value and variance (Nermend, Tarczyńska-Łuniewska, 2013).

$$\left(\Delta\,\eta_{\substack{i\\j}}, \Delta\,\sigma_{\substack{i\\j}}\right) = \left(\eta_{\substack{i\\j}} - \eta_0, \sigma_{\substack{i\\j}} - \sigma_0\right), \tag{9}$$

$$\left(\Delta\,\eta_{\substack{i\\j}}, \Delta\,\sigma_{\substack{i\\j}}^2\right) = \left(\eta_{\substack{i\\j}} - \eta_0, \sigma_{\substack{i\\j}}^2 - \sigma_0^2\right), \tag{10}$$

where: $\eta_{\substack{i\\j}}$ is the mean i-th variable j-th object, $\sigma_{\substack{i,\\j}}$ is standard deviation of i-th variable j-th object, η_o, σ_o are reference points, respectively for the growth of the mean and the standard deviation. Reference points can be arbitrarily chosen and should be identical for all increments of mean values, standard deviations and variances. In practice, in order to simplify a calculation most frequently it is taken as it equals zero. This means that by adding zero to the increment of the mean value, standard deviation or variance we obtain the mean value, standard deviation and variance[17].

In the next stage, the normalization of the designated values pairs (ordered twos) is carried out with the following formula (Nermend, Tarczyńska-Łuniewska 2013):

$$\left(\eta'_{\substack{i\\j}}, \Delta\,\sigma_{\substack{i\\j}}'\right) = \left(\frac{\Delta\eta_{\substack{i\\j}} - \Delta\overline{\eta_\iota}}{\sigma_{\eta_i}}, \frac{\Delta\sigma_{\substack{i\\j}}}{\sigma_{\eta_i}}\right), \tag{11}$$

[17] This is possible until the reference point doesn't change (Nermend, Tarczyńska-Łuniewska, 2013).

and:

$$\left(\eta_{j}^{\prime}, \Delta\, \sigma_{j}^{\prime}\right) = \left(\frac{\Delta\eta_{j} - \Delta\bar{\eta}_{\iota}\ \ \Delta\sigma_{j}^{\ 2}}{\sigma_{\eta_i}}, \frac{}{\sigma_{\eta_i}^{\ 2}}\right), \tag{12}$$

where: $\Delta\bar{\eta}_i$ – is an mean value of mean values, σ_{η_i}i $\sigma_{\eta_i}^2$ are their standard deviation and variance, respectively.

Prior to the delimitation of synthetic measure a pattern ($\Delta\underset{w}{\eta_i^{\prime}}$), which shows the most favorable values of the analyzed feature and anti-pattern ($\Delta\underset{aw}{\eta_i^{\prime}}$), which illustrates the least favorable values are determined. For this purpose, the value of the first and third quartile is used, which for the stimulant pattern ($\Delta\underset{w}{\eta_i^{\prime}}$) assumes the values of the third quartile[18] for stimulant and the first quartile for the destimulant as follows (Nermend, Tarczyńska-Łuniewska, 2013):

$$\underset{w}{\Delta}\,\eta_i^{\prime} = \begin{cases} \underset{k_{III}}{\Delta}\,\eta_i^{\prime} & \text{for stimulants,} \\ \underset{k_I}{\Delta}\,\eta_i^{\prime} & \text{for destimulants,} \end{cases}$$

where: $\Delta\underset{w}{\eta_i^{\prime}}$ is the value of the i-th normalized variable for the pattern, $\Delta\underset{k_I}{\eta_i^{\prime}}$ is the value of the i-th normalized variable for the first quartile, $\Delta\underset{k_{III}}{\eta_i^{\prime}}$ is the value of the i-th normalized variable for the third quartile.

While, in the case of the anti-pattern ($\Delta\underset{aw}{\eta_i^{\prime}}$), the procedure is reversed – as its coordinates, the values of the first quartile for the stimulant and the third quartile for the destimulant are assumed. If the pattern is determined and based on quartiles it represents an unreal, idealized object. There is therefore no need to determine the deviation increases for its coordinates. Determination of synthetic vector measure based on the scalar ratio of vectors representing the objects and vectors pattern and anti-pattern is determined on the basis of the formula (Nermend, Tarczyńska-Łuniewska 2013):

$$\underset{j}{\Delta \mathrm{m}_{s\eta}} = \frac{\sum_{i=1}^{M}\left(\underset{j}{\Delta\eta_i^{\prime}} - \underset{aw}{\Delta\eta_i^{\prime}}\right)\left(\underset{w}{\Delta\eta_i^{\prime}} - \underset{aw}{\Delta\eta_i^{\prime}}\right)}{\sum_{i=1}^{M}\left(\underset{w}{\Delta\eta_i^{\prime}} - \underset{aw}{\Delta\eta_i^{\prime}}\right)^2}. \tag{13}$$

[18] They can also be determined based on the real object.

The next step is to assign the tested objects (in this case: the geographic regions of Europe) to the appropriate classes with following way (Nermend, Tarczyńska-Łuniewska, 2013):

$$\underset{j}{\mathrm{cl}} = \begin{cases} 1 \quad \text{for} \quad \underset{j}{\Delta m_{S\eta}} \geq \overline{\Delta m_S} + \sigma_{\Delta m_S}, \\ 2 \quad \text{for} \quad \underset{j}{\Delta m_{S\eta}} \geq \overline{\Delta m_S} \quad \wedge \quad \underset{j}{\Delta m_{S\eta}} < \overline{\Delta m_S} + \sigma_{\Delta m_S}, \\ 3 \quad \text{for} \quad \underset{j}{\Delta m_{S\eta}} \geq \overline{\Delta m_S} - \sigma_{\Delta m_S} \quad \wedge \quad \underset{j}{\Delta m_{S\eta}} < \overline{\Delta m_S}, \\ 4 \quad \text{for} \quad \underset{j}{\Delta m_{S\eta}} < \overline{\Delta m_S} - \sigma_{\Delta m_S}, \end{cases}$$

where: $\overline{\Delta m_S}$ is the mean value of the mean value increment, $\sigma_{\Delta m_S}$ – is the standard deviation of the mean value increment and $\underset{j}{\mathrm{cl}}$ – is class number for the j-th object.

The first class includes the best objects with the highest values of the synthetic vector measures and the fourth class the worst ones with the lowest values.

On the basis of the increments of standard deviations the maximum value of the standard deviation increment is determined, as follows (Nermend, Tarczyńska-Łuniewska 2013):

$$\underset{j}{\Delta m_{s\sigma max}} = \frac{\max\limits_{i}(\underset{j}{\Delta\sigma_i})}{\sqrt{\sum_{i=1}^{M}\left(\underset{w}{\Delta\eta_i} - \underset{aw}{\Delta\eta_i}\right)^2}}. \tag{14}$$

This maximum value of the increments of standard deviation can be interpreted as a measure of the spatial homogeneity ($\underset{j}{\mathrm{hl}}_{\sigma}$) of development. The lower is the value of this measure the greater is homogeneity and the smaller are the differences between the objects and reverse. Next, the ratio of this maximum value of the increments of standard deviation to the width of the class can be estimated. The division into classes according the level of homogeneity of sustainable development and the width of these classes can be carried out in the following way:

$$\underset{j}{\mathrm{cl}}_{\sigma} = \begin{cases} 1 & \text{for} \quad \Delta\underset{j}{m}_{S\sigma} < p_1\sigma_{mS}, \\ 2 & \text{for} \quad \Delta\underset{j}{m}_{S\sigma} \geq p_1\sigma_{mS} \wedge \Delta\underset{j}{m}_{S\sigma} < p_2\sigma_{mS}, \\ 3 & \text{for} \quad \Delta\underset{j}{m}_{S\sigma} \geq p_2\sigma_{mS} \wedge \Delta\underset{j}{m}_{S\sigma} < p_3\sigma_{mS}, \\ 4 & \text{for} \quad \Delta\underset{j}{m}_{S\sigma} \geq p_3\sigma_{mS}, \end{cases}$$

where: $\underset{j}{\mathrm{cl}}_{\sigma}$ is class number for maximum value of standard deviation of j-th object, p_1, p_2, p_3 – scaling factors chosen by the researcher.

3. STUDY RESULTS

Table 2 shows the results of the classification and the typological groups of the EU countries obtained by means of the taxonomic measure of development (formulas 3–5) calculated on the basis of the characteristics of their situation in the area of sustainable development. The positions of individual countries in the obtained rankings were usually different, with the exception of Sweden and Denmark whose positions (the first and the second, respectively) did not change in the years of study. Finland and Italy did not move further than by one or two positions. The greatest leaps were observed in the case of Slovakia which was one second last in the 2004 ranking, then in 2008 jumped 7 positions higher to get to the 7th position in 2014.

Sixteen EU countries did not see any fall in the ranking due to the crisis, while four countries went up in the ranking by at least six positions. The situation in the area of sustainable development in 2008 compared to 2004 deteriorated in 12 countries – the most affected were Hungary (the fall from the 8th to the 22nd position) and France (the fall from the 9th to the 17th position). Cyprus, Ireland, Luxembourg and Spain went down in 2008 by five positions in relation to 2004. Over the decade of 2004–2014 four member states (Cyprus, Greece, Ireland and Portugal) were ranked lower in each of the years 2008 and 2014 in relation to previously studied year. It should be noted that in 2014 only ten EU countries improved their situation in comparison to 2004. A half of them fell in the ranking for 2014 in relation to 2008 with Greece and the Czech Republic going down by 11 and 10 positions, respectively.

Because the position in the ranking of individual EU countries in the years of study is not the same (in some cases the movements in the ranking are considerable), Kendall's tau coefficients were determined in order to assess the conformity of ordering the objects under study (table 3)[19]. High values of the coeffi-

[19] Kendall's tau coefficients adopt values from the interval [–1, 1]. The closer their value is to 1, the greater is the conformity of ordering (Stanisz, 2007, p. 313–314).

cients confirm the conformity of linear ordering of countries, despite the differences in positions taken by some objects. The highest correlation coefficient was obtained for the 2004 and 2008 rankings. Sometimes, even one diagnostic feature was decisive for belonging to a particular group, the level of which clearly distinguished countries themselves. Due to this, it was decided to determine the measures ω_j that can be interpreted as the scales defining the relative importance of individual diagnostic features[20]. These measures were calculated according to the formula (Nowak, 1990, p. 34–35):

$$\omega = \frac{V_j}{\sum_{j=1}^{m} V_j} \cdot 100\%, \tag{15}$$

where: V_j – classic coefficient of variation calculated for the *j*-th diagnostic feature.

Table 2. THE EU COUNTRIES SORTED BY THE SUSTAINABLE DEVELOPMENT IN: 2004, 2008 AND 2014

Country	Value of synthetic measure (μ_i)	Position in the ranking	Group
2004			
Sweden	0.689	1	I
Denmark	0.604	2	
Ireland	0.551	3	
Finland	0.526	4	
Luxembourg	0.428	5	
Slovenia	0.397	6	
Austria	0.350	7	
Hungary	0.348	8	II
France	0.341	9	
Czech Republic	0.322	10	
Latvia	0.300	11	
Lithuania	0.294	12	
Estonia	0.294	13	
Cyprus.....	0.289	14	
Portugal	0.285	15	III
Belgium	0.250	16	
Greece	0.221	17	
United Kingdom ...	0.216	18	
Netherlands	0.215	19	
Germany	0.202	20	
Malta	0.181	21	
Romania	0.177	22	IV
Spain	0.125	23	
Croatia	0.108	24	
Italy	0.077	25	
Bulgaria	0.046	26	
Slovakia	0.013	27	
Poland	–0.076	28	

[20] The higher the value of the measure, the greater the importance of the *j*-th diagnostic feature.

Table 2. THE EU COUNTRIES SORTED BY THE SUSTAINABLE DEVELOPMENT IN: 2004, 2008 AND 2014 (cont.)

Country	Value of synthetic measure (μ_i)	Position in the ranking	Group
2008			
Sweden	0.716	1	I
Denmark	0.658	2	
Finland	0.526	3	
Austria	0.482	4	
Latvia	0.467	5	
Slovenia	0.434	6	
Czech Republic	0.380	7	
Ireland	0.375	8	II
Lithuania	0.354	9	
Luxembourg	0.342	10	
Estonia	0.320	11	
Romania	0.316	12	
Belgium	0.283	13	
Germany	0.277	14	
Netherlands	0.270	15	III
Portugal	0.262	16	
France	0.253	17	
Greece	0.251	18	
Cyprus	0.247	19	
Slovakia	0.240	20	
United Kingdom	0.222	21	
Hungary	0.210	22	IV
Bulgaria	0.160	23	
Poland	0.157	24	
Malta	0.141	25	
Croatia	0.131	26	
Italy	0.073	27	
Spain	0.003	28	
2014			
Sweden	0.820	1	I
Denmark	0.688	2	
Lithuania	0.633	3	
Luxembourg	0.623	4	
Finland	0.608	5	
Latvia	0.557	6	
Slovakia	0.512	7	
Austria	0.478	8	II
Slovenia	0.469	9	
France	0.465	10	
United Kingdom	0.445	11	
Ireland	0.410	12	
Poland	0.397	13	
Estonia	0.371	14	
Germany	0.339	15	III
Belgium	0.334	16	
Hungary	0.325	17	
Netherlands	0.312	18	
Portugal	0.309	19	
Czech Republic	0.306	20	
Cyprus	0.281	21	
Romania	0.266	22	IV
Croatia	0.219	23	
Bulgaria	0.183	24	
Malta	0.105	25	
Italy	0.074	26	
Spain	0.067	27	
Greece	0.026	28	

S o u r c e: own calculations.

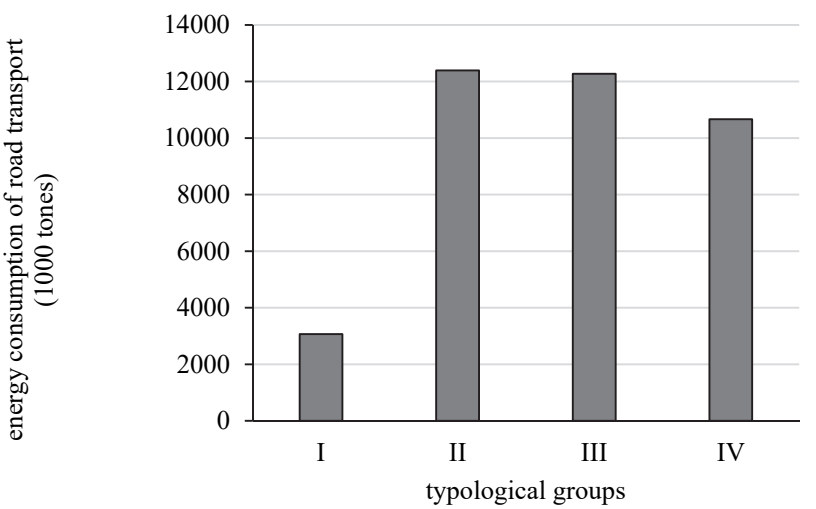
Table 3. KENDALL'S T COEFFICIENTS CALCULATED FOR THE RANKS OF COUNTRIES ACCORDING TO TAXONOMIC MEASURES OF DEVELOPMENT

Year	2004	2008	2014
2004	1.0000	**0.6138**	0.4815
2008	**0.6138**	1.0000	**0.5714**
2014	0.4815	**0.5714**	1.0000

S o u r c e: own calculations.

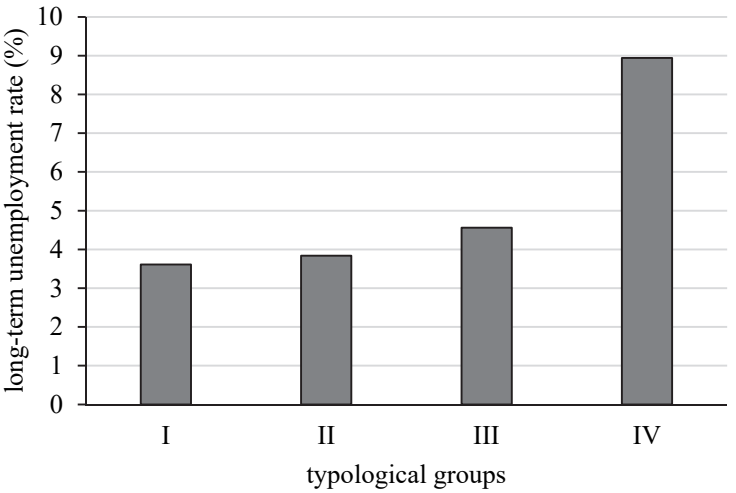
It turned out that in the study of the sustainable development of EU countries based on data from the last analyzed period (2014) the most important are: primary energy consumption (x_{14fDH} – 11.52%), energy consumption of road transport relative to GDP (x_{19gDE} – 11.40%), final energy consumption (x_{5bDE} – 11.30%), long-term unemployment rate (x_{8cDE} – 6.35%), share of renewable energy in gross final energy consumption (x_{15fSO} – 6.20%). These five diagnostic features were characterized by the highest variability in the set of attributes accepted for testing, their share exceeded 46% of the total value of the sum of variability coefficients and therefore they significantly influenced the classification of objects (EU Member States). In order to show the differences in the level of listed characteristics in individual groups, average values in groups were calculated and presented in figures 1–2.

Figure 1. Average energy consumption of road transport, 1000 tons of oil equivalent in typological groups



Source: own elaboration.

Figure 2. Average longiterm unemployment rate in typological groups



Source: own elaboration.

In the first group there were seven countries for which the mean values of diagnostic features were definitely higher than the EU mean in the case of stimulants and lower in the case of destimulants. The objects from this group were mainly characterized by low final energy consumption, high share of energy from renewable sources in total energy, low long-term unemployment rate, high level of education (low participation of early school leavers and high share of people continuing education) and higher compared to the average EU citizens' confidence. The priority for the classification of countries in the second group was mainly high expenditure on R&D, good level of education (low share of early school leavers and high share of people with the third level of education) and high level of primary energy consumption. Objects that were classified in the third group were characterized by similar average values of the analyzed diagnostic features in comparison with the second group. However, the lower rating of the third group was the result of the higher level of the total unemployment rate, long-term unemployment rate and the lowest among all groups, the average value for the share of energy from renewable sources in total energy. In the worst situation in terms of sustainable development were EU countries classified into the fourth group characterized by unfavorable values of the majority of diagnostic features accepted for study[21].

[21] A similar analysis of typological groups can be made in the years 2008–2013. While examining the importance of diagnostic features according to the formula 15, it was also noticed that in the study of the sustainable development of the EU countries the same features that differentiated objects in 2014 were of the greatest importance.

The next stage of the study was the analysis of spatial homogeneity of the EU countries located in geographical regions of Europe. In this case the results for EU countries located in 4 following geographical regions of Europe were analysed:

a) Western Europe (Austria, Belgium, France, Germany, Luxembourg, the Netherlands),
b) Northern Europe (Denmark, Estonia, Finland, Ireland, Latvia, Lithuania, Sweden, the United Kingdom),
c) Southern Europe (Cyprus, Croatia, Greece, Italy, Malta, Portugal, Slovenia, Spain),
d) Eastern Europe (Bulgaria, the Czech Republic, Hungary, Poland, Romania, Slovakia).

The analysis results are presented in table 4 which shows both the ordering of the regions according to the level of development of the average country located in these regions (according the formula 13) and the results of the spatial homogeneity of sustainable development of geographical regions of Europe (formula 14).

Table 4. THE DIVISION OF GEOGRAPHICAL REGIONS OF EUROPE INTO CLASSES ACCORDING TO THE VALUES OF SYNTHETIC VECTOR MEASURE AND THEIR SPATIAL HOMOGENEITY

Year	The division of Europe due to:	Europe			
		Northern	Western	Southern	Eastern
2004	the values of synthetic vector measure $(\Delta m_{sn})_j$[a]	1st rank/ Class I	2nd rank/ Class II	3rd rank/ Class III	4th rank/ Class IV
	spatial homogeneity $(hl_\sigma)_j$[b]	28.36%	29.04%	30.75%	21.04%
2008	the values of synthetic vector measure $(\Delta m_{sn})_j$[a]	1st rank/ Class I	2nd rank/ Class II	3rd rank/ Class III	4th rank/ Class IV
	spatial homogeneity $(hl_\sigma)_j$[b]	34.25%	29.80%	37.63%	22.33%
2014	the values of synthetic vector measure $(\Delta m_{sn})_j$[a]	1st rank/ Class I	2nd rank/ Class II	4th rank/ Class IV	3rd rank/ Class III
	spatial homogeneity $(hl_\sigma)_j$[b]	24.59%	19.41%	21.24%	26.62%

a The level of development of an average country in geographic European region. *b* The ratio of the maximum increase of the standard deviation to the width of the class.

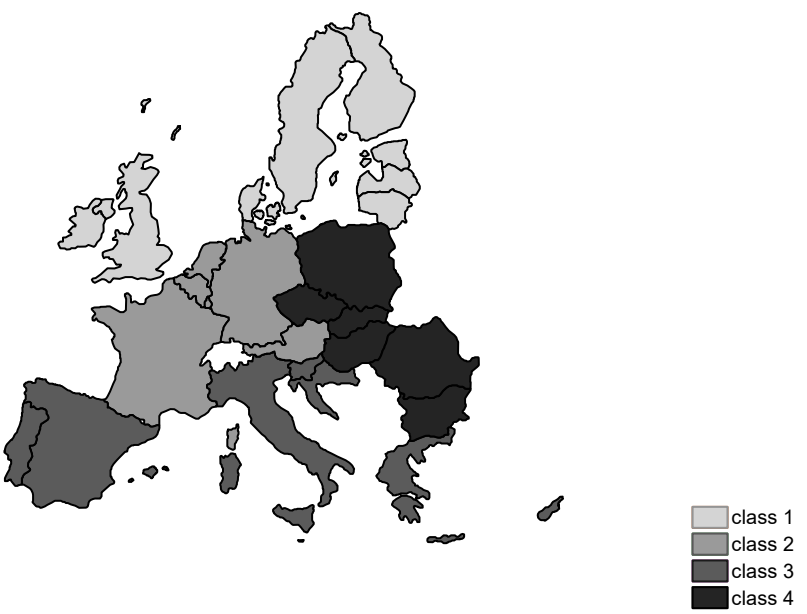
Source: own calculations.

The impact of the economic crisis on the ordering of the EU countries in four geographical regions of Europe is particularly obvious in the Southern and Eastern European Union countries. The position of the Southern countries, that had to cope with the world economic and financial crisis deteriorated and they fell in 2014 into the fourth typological group at the lowest level of development. Before the crisis those countries belonged to the third group. In contrast, the position of

the East European countries improved – they appeared the most immune to the crisis and in 2014 were promoted to the third typological group from the group IV where they were classified in 2004 and 2008. According to the study results, geographical regions of the EU represent an average level of homogeneity of the sustainable development distribution. The results of spatial homogeneity of sustainable development in 2008 and 2014 are shown on figures 3–4.

The influence of the world crisis on the development slowdown can be clearly seen. In 2008 in the case of three out of four analysed geographical regions (Northern, Southern and Western Europe) more disturbances in their spatial homogeneity ($\underset{j}{\mathrm{hl}}_{\sigma} > 29\%$) were observed than in the case of the Eastern Europe ($\underset{j}{\mathrm{hl}}_{\sigma} = 22.33\%$). This situation was probably due to the aforementioned resistance to the crisis of such economies as Poland, the Czech Republic, Slovakia or Hungary. However, in 2014 the countries of this part of the EU saw more imbalance in the sustainable development levels than other regions (the highest value of $\underset{j}{\mathrm{cl}}_{\sigma}$ in the case of this regions of Europe). However, it should be noted that in that year in none of the regions the imbalance exceeded 27%.

Figure 3. The division into classes due to the value of the synthetic measure of the sustainable development in 2004.



Source: own elaboration.

Figure 4. The division into classes due to the value of the synthetic measure of the sustainable development in 2014.

Source: own elaboration.

4. CONCLUSIONS

In the paper the results of the analysis of the spatial uniformity on the basis of sustainable development indicators published by Eurostat were presented. To the study of the spatial uniformity the taxonomic measure of development based on median Weber vector as well vector calculus were used. On the basis of the results of the analysis the spatial differences between the EU countries and the European geographical regions was confirmed. It should be noted, that according of the results of these analysis the improvement of the position taken by Eastern EU countries in the ranking and the deterioration of this position taken by the Southern EU countries were observed.

The same results was noted by others authors (e.g. Klenert et al., 2015; Kobayashi, Shirai, 2016). These authors indicate that the division of the European Union into 'better' West European countries and 'worse' Eastern Europe, or 'old' developed and 'new' developing Union or the founding countries and the remaining member states, are still synonymous to the differences in the EU development. The map of divided Europe has changed a little after the economic and financial crisis when it turned out that it was the Eurozone countries in the South that suffered most of all. According to the report *Central Europe Fit for the Future* (Nic, Świeboda, 2014) published by the think-tanks of the Central European

Policy Institute in Bratislava and the Warsaw demosEuropa: ”North-South axis has largely replaced the old one between ’West’ and ’East’.” The authors of the report also point out that the term ’new Europe’ no longer denotes ’the newly introduced to the club of rich old democracies’ but refers to the countries which, despite their difficult history, have proven their capacity to transform politically and socially and managed to cope with the crisis better than the countries in the South, and even in the North of the Europe. A good example are the Baltic countries that suffered from the 2009 recession at the level of almost 20% of their GDP, but several years later, having implemented painful reforms, met the Eurozone membership criteria and today are developing at the faster rate than any member of the EU. According to Eurostat data base, in 2015 Polish GDP per capita reached 38.7% of the EU average. In the same year in the Czech Republic, which entered the Union with better economy than Poland, GDP per capita was at 54.1% of the European average, in Slovakia – 50.1%, and in Slovenia – 65.0% – i.e. more than in Portugal (60.3%) or Greece (56.4%). What is more, most of these countries have much worse transport infrastructure and their expenditure on R&D is much lower than in the rest of Europe, except Slovenia which spends 2.39% of its GDP (in comparison to Germany with 2.87%).

Catching up with the rest of Europe in this and other areas will take another decade. The negative effect of the crisis on the sustainable development of the EU countries is particularly present in the South European countries the majority of which found it difficult to survive the economic slowdown. However, the situation has improved in Eastern Europe. Moreover, the Western and Northern European Union countries have strengthened their position in the rankings measuring the rate of their sustainable development.

The results obtained in this study can be used in subsequent years to examine the direction of changes in sustainable development levels observed both from the point of view of the EU Member States and geographical regions. The analyses of the Union’s internal homogeneity in this aspect will be particularly useful. The methods applied in this study, such as Weber point and vector analysis, as well as the adopted procedure of selecting diagnostic features allowed for tracking the changes in sustainable development levels not only through the prism of individual SD indicators, but also in reference to many features explaining the EU sustainable development.

REFERENCES

Autor D. H., Katz L. F., Kearney M. S., (2008), Trends in U. S. Wage Inequality: Revising the Revisionists, *The Review of Economics and Statistics,* 90 (2), 300–323.

Bąk I., Cheba K., (2017), Multidimensional Comparative Analysis of Sustainable Development in European Union, Proceedings of the 2017 International Conference Economic Science for Rural Development, 45, Jelgava, LLU ESAF, 14–20.

Berger G., Zwirner, W., (2008), The Interfaces Between the EU SDS and the Lisbon Strategy: Objectives, Governance Provisions, Coordination and Future Developments, ESDN Quarterly Report December 2008.

Boda F., Munteanu L., Raducanu D., (2015), Digital Elevation Modeling Based on Multimodal Aerospace Data in the Context of Sustainable Development of Romania, *Procedia Economics and Finance*, 32, 992–996.

Borawski M., (2012), Vector Space of Increments, *Control and Cybernetics*, 41 (1), 145–170.

Duran D. C., Gogan L. M., Artene A., Duran V., (2015), The Components of Sustainable Development – a Possible Approach, *Procedia Economics and Finance*, 26, 812–817.

Eagle N., Macy M., Claxton R., (2010), Network Diversity and Economic Development, *Science*, 21, 1029–1031.

European Union (EU), (2015), Sustainable Development in the European Union 2015. Monitoring Report of the EU Sustainable Development Strategy, Eurostat, Brussel.

Eurostat, (2017), http://ec.europa.eu/eurostat, access time July 2017.

Gnimassoun B., Mignon V., (2015), How Do Macroeconomic Imbalances Interact? Evidence from a Panel Var Analysis, *Macroeconomic Dynamics*, 20 (7), 1717–1741.

Hellwig Z., (1968), Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*, Przegląd Statystyczny*, 15 (4), 307–327.

Hopwood B., Mellor M., O'Brien G., (2005), Sustainable Development: Mapping Different Approaches, *Sustainable Development*, 13 (1), 38–52.

Klenert D., Mattauch L., Edenhofer O., Lessmann K., (2015), Infrastructure and Inequality: Insights from Incorporating Key Economic Facts About Household Heterogeneity, *Macroeconomic Dynamics*, available on CJO2016.

Kobayashi K., Shirai D., (2016), Heterogeneity and Redistribution in Financial Crises, *Macroeconomic Dynamics*, 20 (6), 1527–1549.

Lira J., Wagner W., Wysocki F., (2002), Mediana w zagadnieniach porządkowania obiektów wielocechowych, in: Paradysz J., (ed.), *Statystyka regionalna w służbie samorządu lokalnego i biznesu*, Internetowa Oficyna Wydawnicza Centrum Statystyki Regionalnej, Akademia Ekonomiczna w Poznaniu, Poznań, 87–99.

Lopez F., (2005), Trade and Growth: Reconciling the Macroeconomic and Microeconomic Evidence, *Journal of Economic Survey*, 19 (4), 623–648.

Meadows D. H., Meadows D. L., Randers J., Behrens III W. W., (1973), *Granice Wzrostu*, Państwowe Wydawnictwo Ekonomiczne, Warszawa.

Młodak A., (2006), *Analiza taksonomiczna w statystyce regionalnej,* Centrum Doradztwa i Informacji DIFIN, Warszawa.

Młodak A., (2014), On the construction of an aggregated measure of the development of interval data, *Computational Statistics,* 29, 895–929.

Młodak A., Józefowski T., Wawrowski Ł., (2016), Zastosowanie metod taksonomicznych w estymacji wskaźników ubóstwa, *Wiadomości Statystyczne,* 61 (2), 1–24.

Moomaw W. R., Bhandry R. R., Kuhl L., Verkooijen P., (2017), Sustainable Development Diplomacy: Diagnostics for the Negotiation and Implementation of Sustainable Development, *Global Policy*, 8 (1), 73–81.

Mulder P., van den Bergh J. C. J. M., (2001), Evolutionary Economic Theories of Sustainable Development, *Growth and Change*, 32 (4), 110–134.

Nermend K., (2009), *Vector Calculus in Regional Development Analysis,* Springer, Berlin.

Nermend K., Tarczyńska-Łuniewska M., (2013), Badanie jednorodności przestrzennej i czasowej rozwoju obiektów społeczno-gospodarczych, *Przegląd Statystyczny,* 60 (1), 85–100.

Nic M., Świeboda, P., 2014, Central Europe Fit for the Future: 10 Years After EU Accession, GLOBSEC Policy Institute, available online (accessed on 15 June 2017):

http://www.cepolicy.org/publications/central-europe- fit-future-10-years-after-eu-accession

Nowak E., (1990), *Metody Taksonomiczne w Klasyfikacji Obiektów Społeczno-gospodarczych*, PWE, Warszawa.

Peacock W. G., Hoover G. A., Kilian C. D., (1988), Divergence and Convergence in International Development: A Decomposition Analysis of Inequality in the World System, *American Sociological Review*, 6, 838–852.

Rigobon R., (2003), Identification Through Heteroskedasticity, *The Review of Economics and Statistics***,** 85 (4), 772–792.

Stanisz A., (2007), Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny, Statystyki podstawowe, 1, StatSoft, Kraków.

Stefanescu D., On A., (2012), Entrepreneurship and Sustainable Development in European Countries Before and During the International Crisis, *Procedia – Social and Behavioral Sciences*, 58, 889–898.

UN (United Nations), (1975), United Nations Environment Programme. Report of the Governing Council on the work pf this third session*,* United Nations, New York, USA*.*

van den Bergh J. C. J. M., Hofkes M. W., (eds.), (1998), *Theory and Implementation of Economic Models for Sustainable Development*, Kluwer Academic Publishers, Dordrecht.

WCED (World Commission on Environment and Development), (1987), Our Common Future. Un Documents: Gathering a Body of Global Agreements has been compiled by the NGO Committee on Education of the Conference of NGOs from United Nations web sites with the invaluable help of information & communications technology, United Nations, New York, USA.

Weber A., (1909, reprint 1971), *Theory of Location of Industries*, Russel & Russel, New York,

Zeliaś, A., (ed.), 2000, *Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym*, AE, Kraków.

BADANIE PRZESTRZENNEJ JEDNORODNOŚCI ZRÓWNOWAŻONEGO ROZWOJU UNII EUROPEJSKIEJ PRZED, W TRAKCIE I PO KRYZYSIE EKONOMICZNYM

Streszczenie

Celem pracy jest analiza przestrzennej jednorodności w obszarze zrównoważonego rozwoju Unii Europejskiej oraz regionów geograficznych Europy rozpatrywanych z punktu widzenia krajów członkowskich UE położonych w tych regionach przed, w trakcie i po kryzysie ekonomicznym z lat 2007–2008. W analizach podobieństw i różnic rozwojowych występujących pomiędzy krajami członkowskimi Unii Europejskiej i w przypadku regionów geograficznych Europy wykorzystano wskaźniki zrównoważonego rozwoju publikowane przez Eurostat. Do ostatecznego zbioru cech diagnostycznych, które stały się podstawą dalszych badań empirycznych, wybrano 22 wskaźniki. Do badania przestrzennego zróżnicowania w obszarze zrównoważonego rozwoju wykorzystano taksonomiczny miernik rozwoju wyznaczony w oparciu o medianę Webera oraz rachunek wek-

torowy. Wpływ kryzysu na uszeregowanie krajów członkowskich UE jest szczególnie widoczny w przypadku krajów Europy Południowej i Wschodniej. Pozycja krajów Europy Południowej, które najgorzej poradziły sobie ze światowym kryzysem gospodarczym i finansowym pogorszyła się, a kraje te zostały zaklasyfikowane do grupy o najniższym poziomie rozwoju przeciętnego kraju członkowskiego. Otrzymane wyniki mogą być wykorzystane w kolejnych latach do badania kierunków zmian zachodzących w obszarze zrównoważonego rozwoju analizowanych zarówno z punktu widzenia pojedynczych państw członkowskich UE, jak i regionów geograficznych Europy.

Słowa kluczowe: zrównoważony rozwój, Unia Europejska, wielowymiarowa analiza porównawcza, mediana Webera, rachunek wektorowy

STUDY OF SPATIAL UNIFORMITY OF SUSTAINABLE DEVELOPMENT OF THE EUROPEAN UNION BEFORE, DURING AND AFTER THE ECONOMIC CRISIS

Abstract

The purpose of the paper is study of spatial uniformity in the field of sustainable development of European Union and geographical regions of Europe analyzed by the prism of EU countries located in this regions before, during and after the economic crisis from 2007–2008.

Material and methods The analysis of similarities and differences between the EU Member States countries or in the case of geographic regions of Europe has been based on sustainable development indicators published by Eurostat. To the final set of diagnostic features, the 22 indicators have been selected. To study the spatial differentiation of sustainable development the taxonomic measure of development based on median vector Weber as well vector calculus were used. The impact of the economic crisis is particularly obvious in the Southern and Eastern European Union countries. The position of the Southern countries, that failed to cope with the world economic and financial crisis, deteriorated and they fell into the group at the lowest level of development. The results obtained in this study can be used in subsequent years to examine the direction of changes in sustainable development levels observed both from the point of view of the EU Member States and geographical regions.

Keywords: sustainable development, multidimensional comparative analysis, the European Union, vector calculus, Weber median