GUS

Indeks 371262
e-ISSN 2657-9545
ISSN 0033-2372

Przegląd Statystyczny
Statistical Review

Vol. 69 | No. 1 | 2022

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

INFORMATION FOR AUTHORS

Przegląd Statystyczny. Statistical Review publishes original research papers on theoretical and empirical topics in statistics, econometrics, mathematical economics, operational research, decision science and data analysis. The manuscripts considered for publication should significantly contribute to the theoretical aspects of the aforementioned fields or shed new light on the practical applications of these aspects. Manuscripts reporting important results of research projects are particularly welcome. Review papers, shorter papers reporting on major conferences in the field, and reviews of seminal monographs are eligible for submission, but only on the Editor's request.

Since 1 May 2019, the journal has been publishing articles in English.

Any spelling style is acceptable as long as it is consistent within the manuscript.

All work should be submitted to the journal through the ICI Publishers Panel
(https://editors.publisherspanel.com/pl.ici.ppanel-app-war/ppanel/index).

For details of the submission process and editorial requirements please visit
https://ps.stat.gov.pl/ForAuthors.

Przegląd Statystyczny
Statistical Review

Vol. 69 No. 1 2022

EDITORIAL OFFICE'S ADDRESS

University of Gdańsk, ul. Armii Krajowej 101, 81-824 Sopot, Poland

Language editing: Scientific Journals Division, Statistics Poland
Technical editing and typesetting: Statistical Publishing Establishment – team supervised by Wojciech Szuchta

Zakład Wydawnictw Statystycznych

Printed and bound: Statistical Publishing Establishment
al. Niepodległości 208, 00-925 Warsaw, Poland, zws.stat.gov.pl

Website: ps.stat.gov.pl

ISSN 0033-2372
e-ISSN 2657-9545
Index 371262

Information on the sales of the journal: Statistical Publishing Establishment
phone no. +48 22 608 32 10, +48 22 608 38 10

Order no. 139/2022 – 215 printed copies

LETTER FROM THE EDITOR

Dear Readers,

Since 2020 *Przegląd Statystyczny. Statistical Review* has been publishing articles exclusively in English. The current issue is the 9th edition fully in English. Until now, 26 research articles, two anniversary articles and two conference reports have been released. With more than 12 thousand text views recorded on the journal's web page, the content of *Przegląd Statystyczny. Statistical Review* is becoming increasingly more visible for the members of the academia and practitioners. I would like to thank the authors for their submissions, the reviewers for their excellent service, and you, the readers, for showing a growing interest in the journal.

I also wish to announce that we are continually seeking new submissions. We welcome high-quality papers addressing significant issues from all branches of economics, finance and management. Those on theoretical and empirical topics in statistics, econometrics, mathematical economics, operational research, decision sciences and data analysis authored by PhD candidates are particularly appreciated. The full editorial process – from the paper's submission to its publishing is free of charge. The final decision on the acceptance or rejection of the article is issued within approximately two months.

On behalf of the Board of Editors,
Paweł Miłobędzki
Editor-in-Chief

CONTENTS

Przegląd Statystyczny. Statistical Review, 2022, vol. 69, 1, 1–20 DOI: 10.5604/01.3001.0015.8669
Received: 3.01.2022, revised: 23.02.2022, accepted: 7.03.2022

Extended Enders and Siklos test for threshold cointegration

Magdalena Osińska,[a] Maciej Gałecki[b]

Abstract. In our previous studies, we modified the Enders and Siklos test for threshold error correction to a version allowing the individual threshold variable to be responsible for the asymmetric mechanism of the system. The idea was to learn about the threshold mechanism both in the long and short run. In this paper, we tested for the asymmetry of the adjustment of the error correction mechanism towards the long-run path. The subsamples within regimes differ in size with respect to the threshold value. The novelty lies in the division of both short and long-run variables according to a threshold variable with a given threshold value (assumed or estimated). We named the test extended Enders and Siklos test (exE-S). The present study focuses on the power and size of the modified procedure. A simulation study was designed, computed and conducted. The results are favourable for the proposed approach, although they strongly depend on the difference in values between the adjustment parameters in the regimes.

Keywords: threshold error correction test, power, size, Monte Carlo, economic growth
JEL: C22, O47

1. Introduction

The paper aims to evaluate the size and power of a novel extended Enders and Siklos test for threshold cointegration and compare it with the size and power of the original one. The test was first described and applied in Boehlke et al. (2017, 2018), and Gałecki and Osińska (2019). However, its size and power were not examined in the previous publications. This paper fills the gap which arose in that area.

Enders and Siklos (2001) defined their test in the context of a threshold cointegration. A threshold cointegration, as the opposite of a linear one, assumes asymmetry in the short-run speed of adjustment to the steady-state, mainly when the bottom-up and top-down adjustment directions are considered. As the previous literature, this paper considers a threshold cointegration (1,1).

The concept of threshold cointegration refers to both cointegration and nonlinearity of the threshold type. The literature on this issue relates to approaches involving a single equation and a system of equations. Balke and Fomby (1997) applied the idea of nonlinear threshold modelling developed by Chan (1993) and Tong (1990) and joined it with the concept of cointegration. Enders and Siklos (2001) developed the testing scheme for nonlinear cointegration and asymmetry, assuming that the

[a] Nicolaus Copernicus University in Toruń, Faculty of Economic Sciences and Management, ul. Gagarina 13a, 87-100 Toruń, e-mail: emo@umk.pl, ORCID: https://orcid.org/0000-0002-9796-2892.
[b] University of Łódź, Faculty of Economics and Sociology, Chair of Econometric Models and Forecasts, ul. Rewolucji 1905 r. 37/39, 90-214 Łódź, e-mail: maciej.galecki@uni.lodz.pl,
ORCID: https://orcid.org/0000-0002-6402-8489.

system's reaction is asymmetric around lagged ECM or lagged momentum ECM. Since the ECM is stationary, the threshold value is supposed to be zero. Stigler (2010) provided a broad overview of the different methods related to Threshold Error Correction Modelling (TECM), including both univariate and multivariate models.

The original Enders and Siklos test was extended by Kapetanios et al. (2006), who considered two cases referred to as threshold variables. In the first one, a threshold variable enters the cointegrating vector and in the second case, it is not present in the long-run equation. In the former case, the threshold variable is responsible for both long-run and short-run dynamics. In the latter one – the cointegrating vector remains independent of the threshold, since it works in the short run only. The approach that fits the idea above was proposed by Kapetanios et al. (2006), who applied it in the smooth transition regression model framework. The threshold model can be considered as a special case of a smooth transition model. Bruzda (2007) fitted the Kapetanios et al. test to the threshold cointegration case.

Tsay (1998) further developed testing for threshold cointegration and examined whether the variable of interest is generated by a linear or nonlinear process. The null hypothesis assumes that Y_t is generated by a linear data generating process, while the alternative hypothesis assumes that it follows the multivariate threshold process. Hansen and Seo (2002) proposed a new test for threshold cointegration, where the test statistic depends on the covariance structure of the processes under examination. The starting point for the test is the linear vector error correction model (VECM). They assumed that each process is integrated of order one. There is only one cointegrating vector in the model. The null hypothesis carries a linear model with cointegration, whereas the alternative one a threshold model with cointegration. In the present paper, the Tsay, and Hansen and Seo tests were used to check the robustness of the proposed procedure in empirical studies.

Many authors found that economic and financial processes often exhibited non-equal reactions to positive or negative stimulus. Granger and Lee (1989), using a threshold model with a sign function revealed asymmetries in sales, production and inventories in the United States. The most frequent asymmetric relationships are those related to price transmission. Frey and Manera (2007) provided a broad overview of the existing literature on asymmetries in price transmission, finding that a threshold-type asymmetry is quite common in a wide range of markets, mainly financial – Martens et al. (1998), fuel – Ghassan and Banerjee (2015), Leszkiewicz-Kędzior and Welfe (2014), Gosińska et al. (2020), as well as the wheat market – Hassouneh et al. (2017). Piłatowska and Włodarczyk (2017) showed a threshold error correction relationship between CO_2 emissions and economic growth. Boehlke et al. (2019) found a vast array of applications related to economic growth modelling.

This study describes the extended Enders and Siklos test and provides a series of simulations showing its size and power. The paper's novelty is that it shows evidence that threshold cointegration can be led not only by the $ECM_{t-1}/\Delta ECM_{t-1}$ term but also individual variables responsible for the threshold mechanism.

The paper is organised as follows: Section 2 briefly describes the procedure of Enders and Siklos and its extensions and discusses a modified testing approach using a TECM basis, in Section 3 the simulation results are presented, while the empirical example is shown in Section 4. The conclusions are presented in the last part of the paper.

2. Extended Enders and Siklos test

Enders and Siklos (2001) assumed no cointegration in the null hypothesis, whereas nonlinearity is assumed under the alternative hypothesis applying a two-regime threshold model. The threshold variable is defined as a SETAR variable, which is either lagged error correction term ECM_{t-1} or the M-TAR variable, i.e. momentum error correction variable ΔECM_{t-1}. The value of a threshold can be estimated or assumed to be constant. The authors adopted zero as the natural threshold value for the mentioned variables in the original paper. The consequences of the Enders and Siklos test are related to the following cases: threshold cointegration and no threshold cointegration, which implies a linear cointegration, a stationary TAR model, or a partial cointegration.

Enders and Siklos' (2001) procedure consists of the stages listed below.

1. It is assumed that a linear cointegrating equation exists under the conditions defined in Engle and Granger (1987):

$$Y_t = \alpha_0 + \sum_{i=1}^{k} \alpha_i X_{it} + u_t, \tag{1}$$

2. The testing regression is estimated as:

$$\Delta\hat{u}_t = I_t \rho_1 \hat{u}_{t-1} + (1 - I_t)\rho_2 \hat{u}_{t-1} + \sum_{i=1}^{p} \beta_i \Delta\hat{u}_{t-i} + \varepsilon_t, \tag{2}$$

where

$$\hat{u}_t = Y_t - \hat{Y}_t = ECM_t.$$

$$I_t = \begin{cases} 1 & \text{for} \quad \hat{u}_{t-1} \geq \gamma \\ 0 & \text{for} \quad \hat{u}_{t-1} < \gamma \end{cases} \quad \text{or} \quad I_t = \begin{cases} 1 & \text{for} \quad \Delta\hat{u}_{t-1} \geq \gamma \\ 0 & \text{for} \quad \Delta\hat{u}_{t-1} < \gamma \end{cases}$$

and $\gamma = 0$.

It is assumed that the threshold in Equation (2) is defined in terms of the error correction mechanism: (ECM) $\hat{u}_{t-1}$ or $\Delta\hat{u}_{t-1}$.

3. The set of two null hypotheses to be tested takes the following form:

$$H_0^1: \rho_1 = \rho_2 = 0, \tag{3}$$

$$H_0^2: \rho_1 - \rho_2 = 0. \tag{4}$$

H_0^1 is for the case of no threshold cointegration; consequently, the Engle-Granger linear cointegration is confirmed, H_0^2 assumes a symmetric reaction, being the argument for linear cointegration. If both hypotheses are rejected, the Enders and Siklos procedure indicates threshold cointegration around the long-run equilibrium. The short-run adjustment is asymmetric with respect to positive and negative changes. A precise interpretation of the set of hypotheses to be tested (3-4) was provided by Balke and Fomby (1997). This interpretation is presented in Table 1.

Table 1. Possible models under the no-TECM hypotheses

System characteristics	Linearity	Nonlinearity
No cointegration	H_0^1: No linear cointegration	H_1^2: No cointegration. Nonlinear residual process
Cointegration	H_1^2: Linear cointegration	H_1^2: Nonlinear cointegration

Source: based on Balke and Fomby (1997).

Stigler (2010) emphasised that testing for threshold cointegration involves two issues that must be solved simultaneously: cointegration and nonlinearity. Hence, the following cases are possible: cointegration and threshold effects, cointegration and no threshold effects, no cointegration and no threshold effects and, finally, no cointegration and threshold effects.

The results of the Enders and Siklos approach allow the identification of asymmetric reactions around the entire cointegrating vector (which can be unknown). Still, they do not indicate individual threshold variables responsible for the asymmetric mechanism of the system. In many cases, single variables can diversify the mechanism of a short-run adjustment. Two possible cases are considered: the first, when a threshold variable enters the cointegrating vector and the second, where it is not present in the long run. The threshold variable is responsible for both long-run and short-run dynamics in the first case. In contrast, in the second case, the cointegrating vector remains independent of the threshold since it only works in the short run.

The approach that fits the idea above was partially proposed by Kapetanios et al. (2006) and modified by Bruzda (2007). Having (1) unchanged, the testing of Equation (2) is a matter of the re-formulation into the following form:

$$\Delta Y_t = I_t\,\rho_1\hat{u}_{t-1} + (1-I_t)\rho_2\hat{u}_{t-1} + \omega\Delta X_t + \sum_{j=1}^{p}\psi_j\Delta Z_{t-j} + \varepsilon_t, \tag{5}$$

where indicator functions I_t remain the same as defined above and $\gamma = 0$. This test can be extended by allowing for other than $\hat{u}_{t-1} = 0$ and $\Delta\hat{u}_{t-1} = 0$ threshold variables. The set of possible threshold variables is defined in vector Z_t:

$$Z_t = (Y_t, X_{1t}, X_{2t}, \ldots, X_{kt})'.$$

Then the threshold value (empirical level of γ) is a subject of estimation, where

$$I_t = \begin{cases} 1 & \text{for} \quad Z_{t-i} \geq \hat{\gamma} \\ 0 & \text{for} \quad Z_{t-i} < \hat{\gamma} \end{cases} \tag{6}$$

or

$$I_t = \begin{cases} 1 & \text{for} \quad \Delta Z_{t-i} \geq \hat{\gamma} \\ 0 & \text{for} \quad \Delta Z_{t-i} < \hat{\gamma} \end{cases} \tag{7}$$

and

$$-\infty < \hat{\gamma} < \infty; \quad \hat{\gamma} = arg\min_{\gamma} AIC(\gamma). \tag{8}$$

This approach allows the identification of asymmetric reactions in the long run, although it is possible for individual variables to be the threshold. In this approach, the number of observations in the short run remains equal in both regimes.

Boehlke et al. (2018) proposed a new testing procedure based on the entire set of variables available in long-run and short-run equations. This procedure extends the set of possible thresholds and determines the way they impact the identification of the periods of intense economic growth within the observed sample. Long-run equation (1) remains the same. The testing equation is modified to the form:

$$\begin{aligned} \Delta Y_t = I_t\,\rho_1\hat{u}_{t-1} + (1-I_t)\rho_2\hat{u}_{t-1} + \sum_{s=1}^{p_{II}} I_t\beta_{s1}\Delta Y_{t-s} + \\ + \sum_{s=1}^{p_I}(1-I_t)\,\beta_{s2}\,\Delta Y_{t-s} + \sum_{i=0}^{k} I_t\alpha_i\Delta X_{it} + \sum_{i=0}^{k}(1-I_t)\alpha_i\Delta X_{it} + \\ + \sum_{j=1}^{q_{II}}\sum_{i=1}^{k} I_t\gamma_{j1}\Delta X_{it-j} + \sum_{j=1}^{q_I}\sum_{i=1}^{k}(1-I_t)\gamma_{j2}\Delta X_{it-j} + \varepsilon_t, \end{aligned} \tag{9}$$

where

$$I_t = \begin{cases} 1 & \text{for} \quad Z_{t-i} \geq \hat{\gamma} \\ 0 & \text{for} \quad Z_{t-i} < \hat{\gamma} \end{cases} \quad \text{or} \quad I_t = \begin{cases} 1 & \text{for} \quad \Delta Z_{t-i} \geq \hat{\gamma} \\ 0 & \text{for} \quad \Delta Z_{t-i} < \hat{\gamma} \end{cases}$$

$$Z_t = (Y_t, X_{1t}, X_{2t}, \ldots, X_{kt})'$$

and

$$-\infty < \hat{\gamma} < \infty \,;\ \hat{\gamma} = arg \min_{\gamma} AIC(\gamma).$$

In the proposed model, the short-term equations differ between the regimes in terms of the following: a vector of explanatory variables, number of observations and parameters estimate. The approach seems fairly complex, because it shows asymmetries around the long-run and in the short-run dynamics. The advantage of such an approach is that in the final TECM different sets of variables can act in different regimes having the long-run relationship unchanged. However, its limitation is related to the number of observations; if the time series of interest is short, some results may remain unverified.

Three approaches to the TECM specification described above should be considered as nested – the last one nests the second approach, and the second nests the first one. The sequence of testing from the simplest to the broadest course validates the results. If they can be confirmed by Enders and Siklos, Kapetanios et al. and the extended Enders and Siklos approach, the nonlinear mechanism underlying the relationship in question becomes very likely.

3. Simulation results

3.1. The experiment

In the experiment, thresholds ECM_{t-1} and ΔECM_{t-1} were used to ensure the comparability of the original and extended Enders and Siklos test results. The experiment was based on the Monte Carlo method. The simulations included the following steps:

1. generating time series Y_t and X_t (both 1(1)), with the error terms defined as: $\varepsilon_t \sim N(0,1), \eta_t \sim N(0,1)$;
2. generating long-run relationship $Y_t = 0.5 - 0.2X_t + u_t$, where $\hat{u}_t = ECM_t$;
3. checking the stationarity of the residuals ($\hat{u}_t \sim I(0)$);
4. calculating threshold variables $\hat{u}_{t-1}$ and $\Delta\hat{u}_{t-1}$;

5. determining the threshold values to satisfy the following sample proportion between the regimes:
   a) 50%–50%;
   b) 60%–40%;
   c) 80%–20%.

   To estimate the threshold values from $\hat{u}_t$, the following rules were applied: (a) at the median level, (b) at decile_6 level, and (c) at decile_8 level.

   The assumed sample sizes including 50; 100; 500; 1,000 and 2,000 correspond to different situations observed in practice. Typical economic time series observed monthly, quarterly or at an annual frequency consist of 50 or 100 observations. The numbers 500; 1,000 and 2,000 enable the verification whether the longer time series increase the power of the test.
6. performing the Enders and Siklos test (E-S) based on the equation:

$$\Delta\hat{u}_t = I_t \rho_1 \hat{u}_{t-1} + (1 - I_t)\rho_2 \hat{u}_{t-1} + \varepsilon_t,$$

   where $\varepsilon_t \sim N(0,1)$ and I_t were defined in Equation (2);
7. performing the extended test (exE-S) based on the short-run model of the form:

$$\Delta Y_t = I_t \rho_1 \hat{u}_{t-1} + (1 - I_t)\rho_2 \hat{u}_{t-1} + I_t a_1 \Delta x_t + (1 - I_t) a_2 \Delta x_t + I_t b_1 \Delta Y_{t-1} + (1 - I_t) b_2 \Delta Y_{t-1} + \varepsilon_t.$$

In the experiment, parameters ρ_1 and ρ_2 were assumed to change in the range of [–0.99; –0.09] with 0.1 steps. The parameters in the short-run equation (a_1, a_2, b_1, b_2) are defined as follows:

- symmetric negative (–0.3; –0.3; –0.3; –0.3);
- asymmetric negative (–0.3; –0.3; –0.6; –0.3);
- symmetric positive (0.3; 0.3; 0.3; 0.3);
- asymmetric positive (0.3; 0.3; 0.6; 0.3);
- symmetric mixed (–0.3; 0.3; –0.3; 0.3);
- asymmetric mixed (0.3; –0.3; 0.6; –0.3).

A total of 10,000 replications were carried out and the simulation procedure was performed in the Gretl package. Threshold variables ECM_{t-1} and ΔECM_{t-1} were taken from the long-run regression. The threshold value was presumably known and equal to zero. After each sampling, the observations were assigned to one of the two regimes, and models were tested for parameters significance. Insignificant variables were excluded from the model.

In the case of the exE-S test, H_0^1 is tested using the Wald test, like in the case of the E-S. When H_0^2 is considered, the significance of the difference between parameters of error correction mechanism, i.e. ρ_1 and ρ_2 is subject to testing. Two null hypotheses are defined in (3) and (4). The distribution of the Wald test is typically analysed in the form of a chi-squared test or F test, whereas the latter is appropriate for small samples. It is proven that if $X \sim F(n_1, n_2)$, the limiting distribution of n_1X as $n_2 \to \infty$ is the chi-square distribution with n_1 degrees of freedom (Hogg et al., 2005). Taking into account a large sample, the exE-S test was compared to the chi-squared distribution with n_1 degrees of freedom, where n_1 is the number of restrictions (for H_0^1 $n_1 = 2$, and for H_0^2 $n_1 = 1$). Using the Kolmogorov-Smirnov goodness of fit test, the exE-S test did not allow the rejecting of the null hypothesis assuming that its distribution fits the chi-squared distribution at the significance level of 1%.

3.2. Power of the extended and original Enders and Siklos test

Power is an essential characteristic of the statistical test. Power refers to the probability of rejecting H_0 when it is false. On the other hand, size is defined as the probability of rejecting the null when it is true. The standard approach of Neyman and Pearson is to maximise the power while limiting the size by a pre-specified significance level of α (Lloyd, 2006).

In the study, the power and size of the original and extended tests were checked for sensitivity according to:

a) the sample size;
b) the number of observations in the regimes;
c) the values of parameters ρ_1 and ρ_2 and their difference;
d) the parameter values and asymmetry in the short run;
e) the significance levels.

These imply that the number of the results of the simulations is vast. Therefore, only the crucial ones are presented in the paper. First of all, the power of all tests for H_0^1 is equal to 1 for all cases; the refore it is not presented here.[1] The power results for H_0^2 are presented in Figures 1 and 2, and Table 2. The simulated values are shown at the median level calculated over 10,000 replications. Figure 1 refers to the 0.1 difference between the values of parameters ρ_1 and ρ_2, while Figure 2 shows the results when the difference is 0.2. For more considerable differences, the power approaches 1. The figures presenting power results include the *a–d* characteristics mentioned above. The significance level was assumed to take the value of 1%.

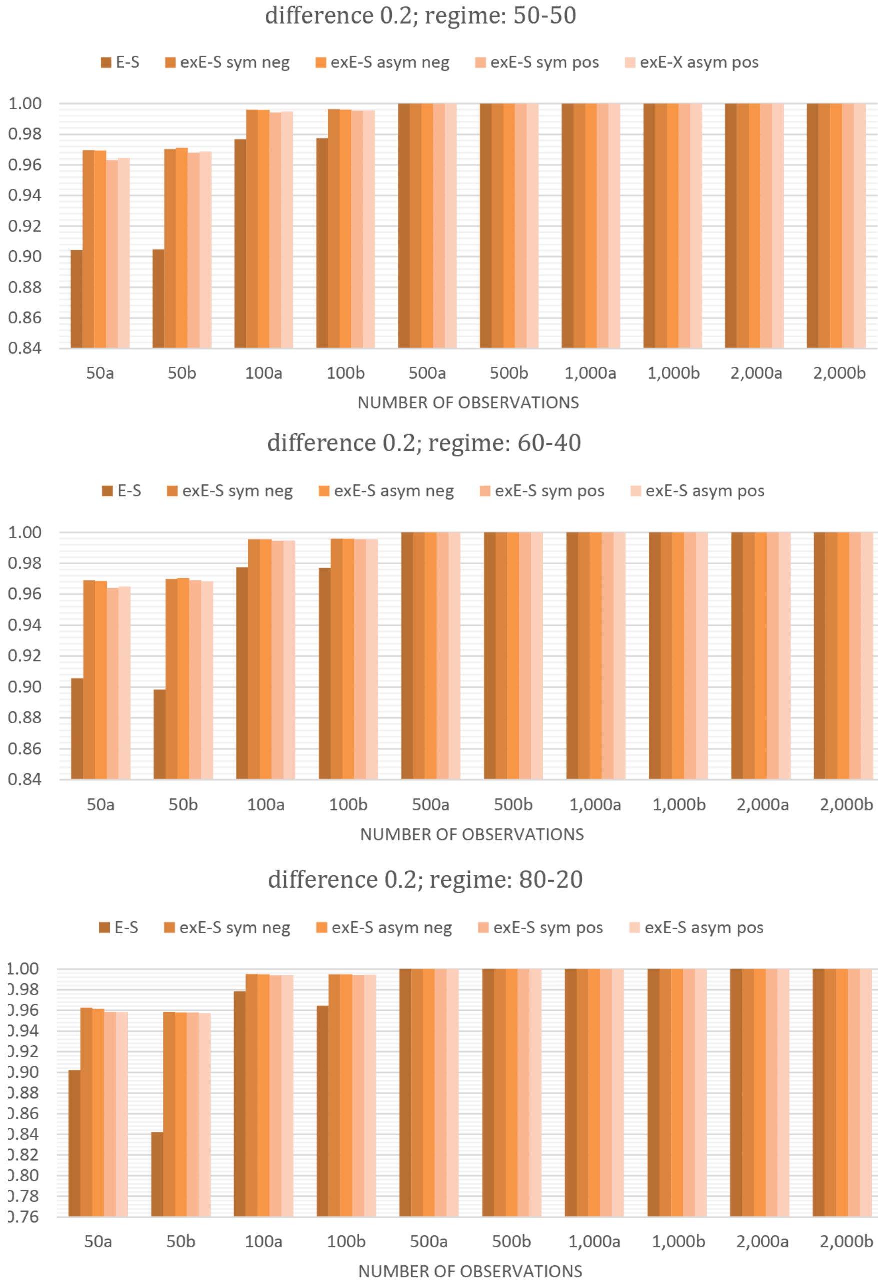
[1] All results are available upon request.

Figure 1. Power of the exE-S and E-S test, difference: 0.1

Note. Difference=$|\rho_1\text{-}\rho_2|$; a – ECM_{t-1}; b – ΔECM_{t-1}; sym neg – parameters in the short-run equation equal in regimes, negative; asym neg – parameters in the short-run equation non-equal in regimes, negative; sym pos – parameters in the short-run equation equal in regimes, positive; asym pos – parameters in the short-run equation non-equal in regimes, positive. Significance level: 1%.

Source: authors' work.

Figure 2. Power of the exE-S and E-S, difference: 0.2



Note. Difference=$|\rho_1 - \rho_2|$; a – ECM_{t-1}; b – ΔECM_{t-1}; sym neg – parameters in the short-run equation equal in regimes, negative; asym neg – parameters in the short-run equation non-equal in regimes, negative; sym pos – parameters in the short-run equation equal in regimes, positive; asym pos – parameters in the short-run equation non-equal in regimes, positive. Significance level: 1%.

Source: authors' work.

Table 2. Power of the exE-S and E-S tests

Obs. no.	Z_t	Regime	ρ1–ρ2	H_0^1: E-S	H_0^2: E-S	H_0^1:exE-S[1]	H_0^2:exE-S[1]	H_0^1:exE-S[2]	H_0^2:exE-S[2]
100	ECM_{t-1}	50–50	0.1	1.00	0.4157	1.00	0.6919	1.00	0.7077
100	ΔECM_{t-1}	50–50	0.1	1.00	0.4134	1.00	0.6943	1.00	0.7072
100	ECM_{t-1}	60–40	0.1	1.00	0.4174	1.00	0.6872	1.00	0.7041
100	ΔECM_{t-1}	60–40	0.1	1.00	0.4197	1.00	0.6946	1.00	0.7040
100	ECM_{t-1}	80–20	0.1	1.00	0.4228	1.00	0.6851	1.00	0.7026
100	ΔECM_{t-1}	80–20	0.1	1.00	0.3846	1.00	0.7241	1.00	0.7341
1,000	ECM_{t-1}	50–50	0.1	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ΔECM_{t-1}	50–50	0.1	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ECM_{t-1}	60–40	0.1	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ΔECM_{t-1}	60–40	0.1	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ECM_{t-1}	80–20	0.1	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ΔECM_{t-1}	80–20	0.1	1.00	1.00	1.00	1.00	1.00	1.00
100	ECM_{t-1}	50–50	0.2	1.00	0.9768	1.00	0.9945	1.00	0.9958
100	ΔECM_{t-1}	50–50	0.2	1.00	0.9772	1.00	0.9950	1.00	0.9957
100	ECM_{t-1}	60–40	0.2	1.00	0.9775	1.00	0.9944	1.00	0.9958
100	ΔECM_{t-1}	60–40	0.2	1.00	0.9770	1.00	0.9951	1.00	0.9959
100	ECM_{t-1}	80–20	0.2	1.00	0.9783	1.00	0.9921	1.00	0.9949
100	ΔECM_{t-1}	80–20	0.2	1.00	0.9645	1.00	0.9930	1.00	0.9949
1,000	ECM_{t-1}	50–50	0.2	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ΔECM_{t-1}	50–50	0.2	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ECM_{t-1}	60–40	0.2	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ΔECM_{t-1}	60–40	0.2	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ECM_{t-1}	80–20	0.2	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ΔECM_{t-1}	80–20	0.2	1.00	1.00	1.00	1.00	1.00	1.00
100	ECM_{t-1}	50–50	0.5	1.00	1.00	1.00	1.00	1.00	1.00
100	ΔECM_{t-1}	50–50	0.5	1.00	1.00	1.00	1.00	1.00	1.00
100	ECM_{t-1}	50–50	0.5	1.00	1.00	1.00	1.00	1.00	1.00
100	ΔECM_{t-1}	50–50	0.5	1.00	1.00	1.00	1.00	1.00	1.00
100	ECM_{t-1}	60–40	0.5	1.00	1.00	1.00	1.00	1.00	1.00
100	ΔECM_{t-1}	60–40	0.5	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ECM_{t-1}	60–40	0.5	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ΔECM_{t-1}	60–40	0.5	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ECM_{t-1}	80–20	0.5	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ΔECM_{t-1}	80–20	0.5	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ECM_{t-1}	80–20	0.5	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ΔECM_{t-1}	80–20	0.5	1.00	1.00	1.00	1.00	1.00	1.00
100	ECM_{t-1}	50–50	0.9	1.00	1.00	1.00	1.00	1.00	1.00
100	ΔECM_{t-1}	50–50	0.9	1.00	1.00	1.00	1.00	1.00	1.00
100	ECM_{t-1}	50–50	0.9	1.00	1.00	1.00	1.00	1.00	1.00
100	ΔECM_{t-1}	50–50	0.9	1.00	1.00	1.00	1.00	1.00	1.00
100	ECM_{t-1}	60–40	0.9	1.00	1.00	1.00	1.00	1.00	1.00
100	ΔECM_{t-1}	60–40	0.9	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ECM_{t-1}	60–40	0.9	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ΔECM_{t-1}	60–40	0.9	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ECM_{t-1}	80–20	0.9	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ΔECM_{t-1}	80–20	0.9	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ECM_{t-1}	80–20	0.9	1.00	1.00	1.00	1.00	1.00	1.00
1,000	ΔECM_{t-1}	80–20	0.9	1.00	1.00	1.00	1.00	1.00	1.00

Note. The symbols: H_0^1:E-S, H_0^2:E-S refer to the E-S test, and H_0^1:exE-S[1], H_0^2:exE-S[1], H_0^1:exE-S[2], H_0^2:exE-S[2] to the exE-S test with short-run parameters (1) *a*={–0.3; 0.3} and *b*={–0.3; –0.3}, and (2) *a*={0.3; –0.3}, and *b*={0.6; –0.3}, respectively. Parameters in the short term equation are asymmetric, both positive and negative (1) *a*={–0.3; 0.3} and *b*={–0.3; –0.3}, and (2) *a*={0.3; –0.3}, and *b*={0.6; –0.3}. Significance level: 1%.

Source: authors' calculations.

The results presented above show that the power of the exE-S test strongly depends on the difference between ρ_1 and ρ_2. If $\rho_1 - \rho_2 = 0.1$, the results presented in Figures 1–2 exhibit insufficient power, particularly for 50 and 100 observations. It is due to the weak asymmetry effect between the regimes. Also, the E-S test loses its power in such a case. The results align with the power of the threshold cointegration tests presented in Bruzda (2007, pp. 326–327). She considered the Kapetanios et al. test in the form presented in Equation (5), particularly the case when the difference varied between 0.0 and 0.4 and the number of observations was 100. The power of the test when the threshold value was known and equal to 0 for the 5% significance level was between 0.062 and 0.684 in the case of ECM_{t-1} and between 0.01 and 0.996 in the case of ΔECM_{t-1}. In the case of the test based on Equations (5–7), the power for 0.1 asymmetries is much higher, so it is for higher asymmetry.

The results presented in Table 2 show that the difference in power when ρ_1- ρ_2 change from 0.1 to 0.2 is substantial. Suppose the difference increases to 0.5, the power of both tests is entirely satisfactory. Table 2 contains results for 100 and 1,000 observations when the parameters in the short run change their signs and values. These have no impact on the examined test's power. Also, the division between regimes related to the number of observations in the regimes does not influence the results.

3.3. Size of the extended and original Enders and Siklos tests

Results of the power are reliable only when the size of the test is kept across various assumptions. In simulations, the size was assumed to take the following values: 0.01; 0.05; 0.1. Figures 3 and 4 show empirical sizes for H_0^1 and H_0^2, respectively. The number of observations was 100 and 1,000.

Figure 3. Size of exE-S and E-S for H_0^1

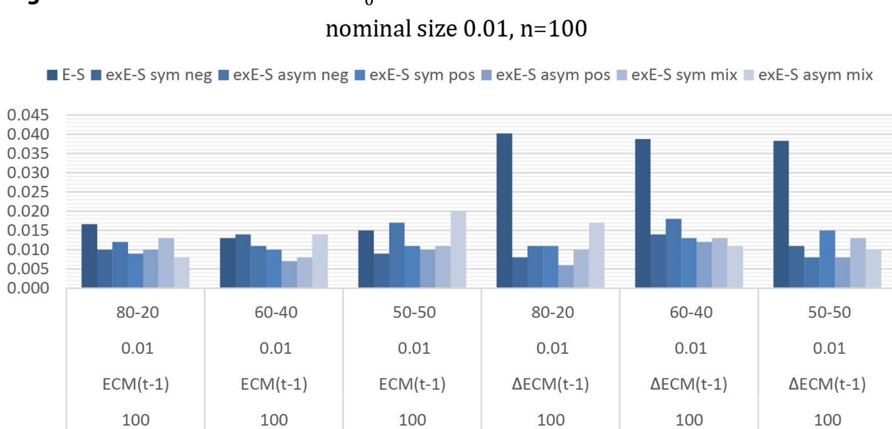
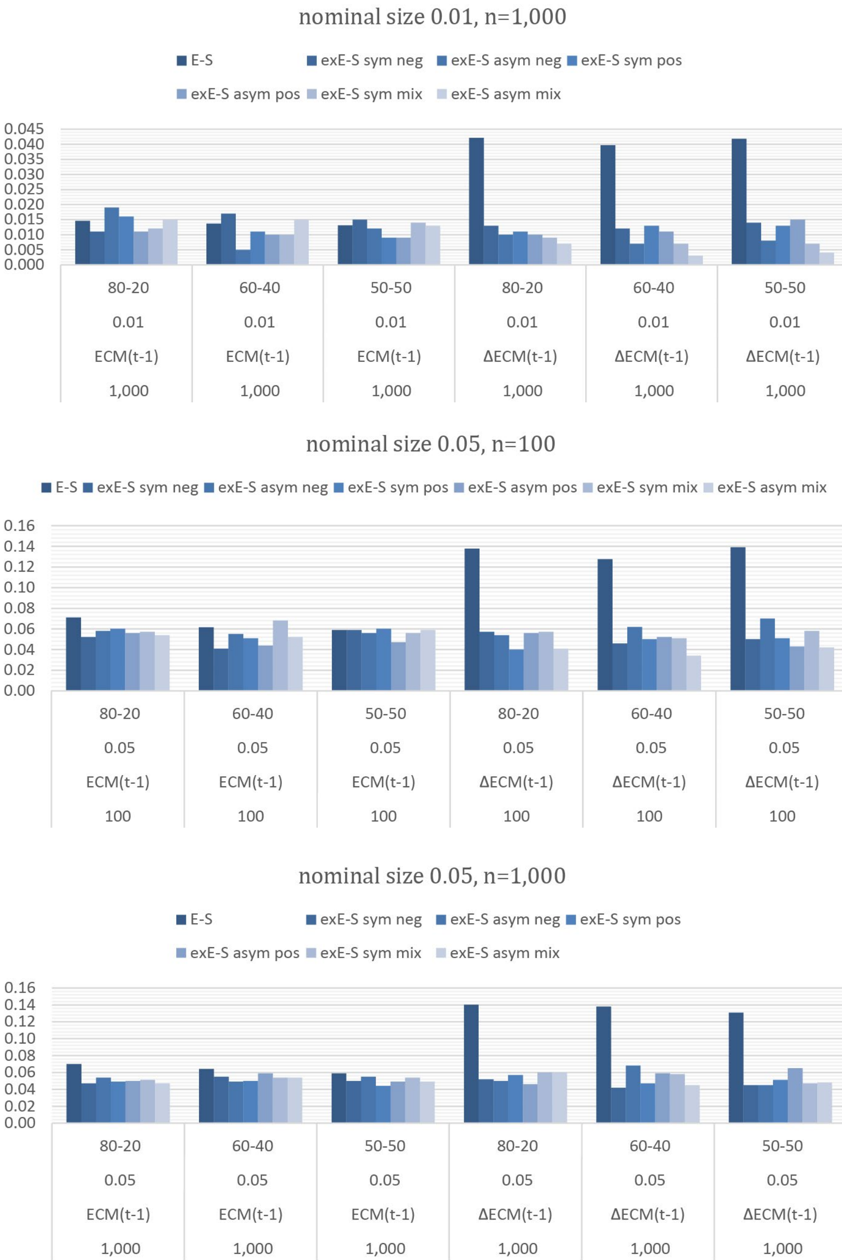


Figure 3. Size of exE-S and E-S for H_0^1 (cont.)



Note. sym neg – parameters in the short-run equation equal in regimes, negative; asym neg – parameters in the short-run equation non-equal in regimes, negative; sym pos – parameters in the short-run equation equal in regimes, positive; asym pos – parameters in the short-run equation non-equal in regimes, positive; sym mix – parameters in the short run equal in modulus, opposite signs; asym mix – parameters in the short run non-equal in modulus, opposite signs.

Source: authors' work.

Figure 4. Size of exE-S and E-S for H_0^2

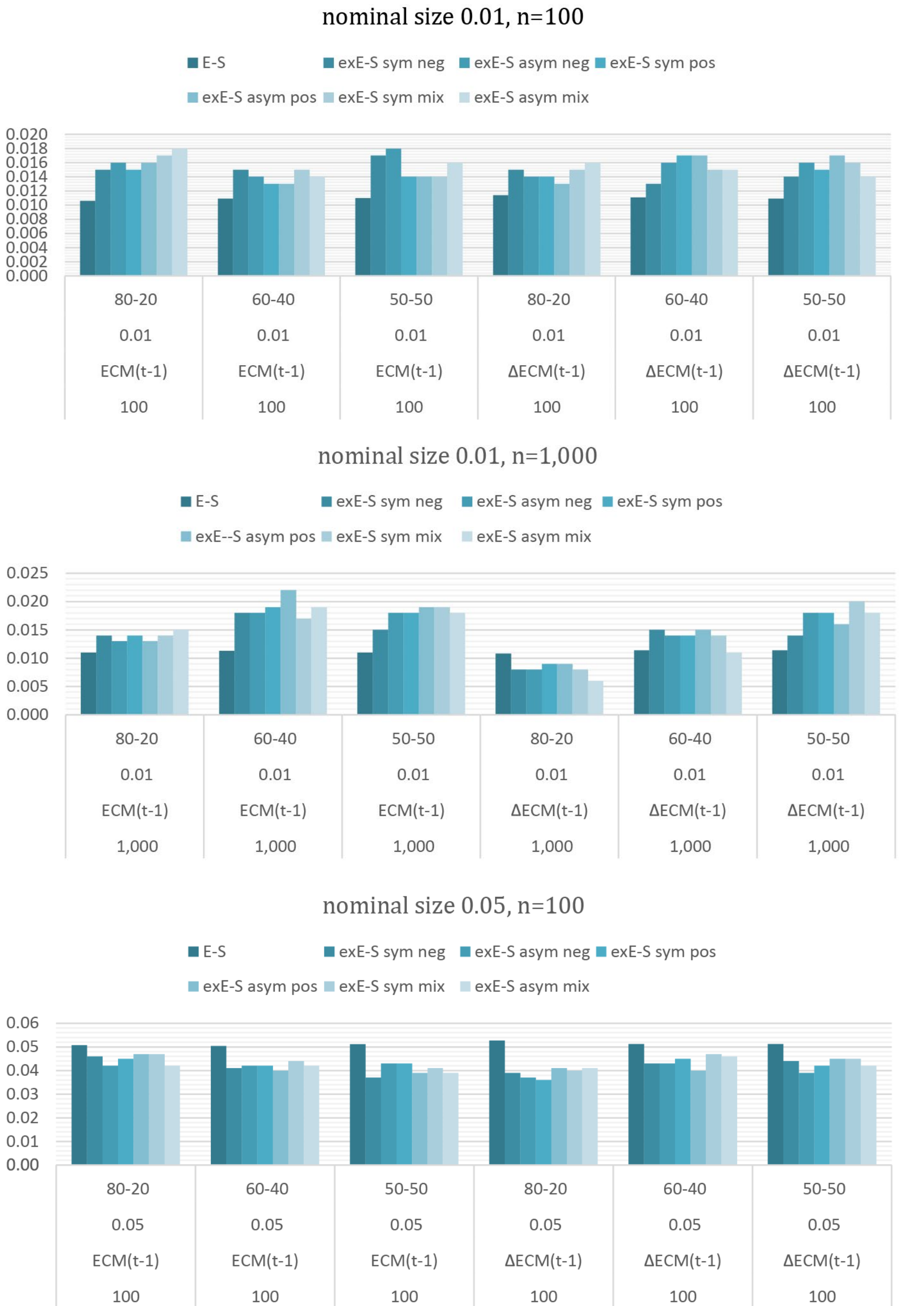
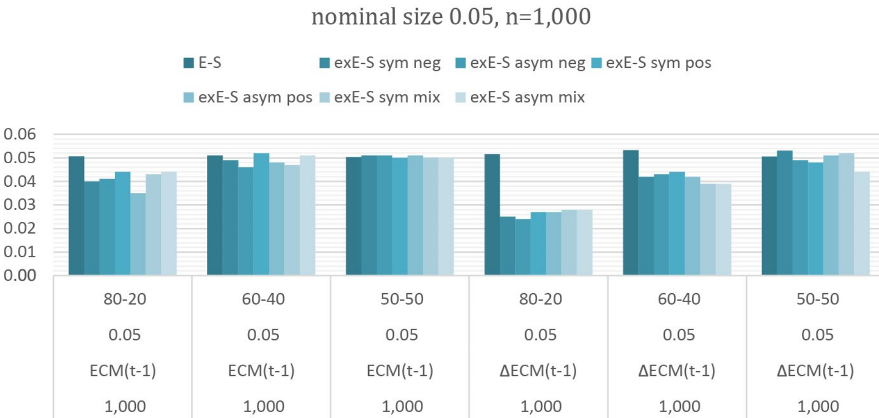


Figure 4. Size of exE-S and E-S for H_0^2 (cont.)



Note. sym neg – parameters in the short-run equation equal in regimes, negative; asym neg – parameters in the short-run equation non-equal in regimes, negative; sym pos – parameters in the short-run equation equal in regimes, positive; asym pos – parameters in the short-run equation non-equal in regimes, positive; sym mix – parameters in the short run equal in modulus, opposite signs; asym mix – parameters in the short run non-equal in modulus, opposite signs.

Source: authors' work.

Figure 3 presents the size for H_0^1. The E-S test size is higher than the nominal one {0.01; 0.05}, but close to it when the threshold variable is assumed to be ECT_{t-1}. In case the threshold variable is ΔECT_{t-1}, the size is much larger than the assumed one. In the case of 0.01, the estimated size is more than three times higher than the assumed one, while for 0.05, the estimated size is over twice as large. The exE-S test keeps its nominal size more stable. The observed differences concern the number of observations in regimes and short-run parameters. The test was insensitive to the threshold variable.

Figure 4 shows the empirical size for H_0^2, which distinguishes asymmetric effects between the regimes. In this case, the E-S test size is very close to the nominal one, disregarding the assumptions. The extended test gave the best results when the nominal size was 0.05. In the case of 0.01, the size of the extended test was larger than the nominal one. The number of observations in both regimes mattered when the proportion was 80%–20% in the respective regimes. In that instance the size was smaller than the nominal one. It is evident if the threshold variable was ΔECM_{t-1}.

4. Empirical example

The successful applications presented in Boehlke et al. (2019) implied further interest in using the exE-S test in the area of economic growth. In the empirical illustration, the example of the Israeli economy is presented. The economy of Israel was the subject of numerous analyses of the factors of its success. Trajtenberg (2001) characterised R&D expenditures, Chorev and Anderson (2006) analysed success in Israeli high-tech start-ups, and Aharoni (2014) provided an in-depth insight into the Israeli economic processes. The paper uses annual data for the years 1980–2017 to uncover the signs of threshold cointegration while GDP *per capita* is considered an endogenous variable. The data were downloaded from OECD (http://stats.oecd.org/), the Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org/), and the Central Bureau of Statistics in Israel (https://www.cbs.gov.il/EN/pages/default.aspx).

Figure 5 presents the Israeli GDP *per capita* expressed in US dollars in constant prices of 2010 and transformed into logarithms. One can notice a structural break around 2002. The Quandt (1958) test results confirmed it with a value of 29.5 (p-value 0.0013). The structural break was strongly related to the dot-com bubble, which significantly affected the Israeli start-ups (Zilberfarb, 2006).

Figure 5. GDP *per capita* in Israel in 1980–2017

Note. Israeli GDP *per capita* expressed in US dollars in constant prices as of 2010, then transformed into logs.
Source: authors' calculations.

As threshold variables, the following were tested: R&D expenditures (R&D), short interest rate (IRs), military expenditures (MilExp), the exchange rate of Israeli shekel to USD (EXR), and savings (Sav). All the potential thresholds are lagged. Above these, the standard threshold variables, i.e. ECM_{t-1} and ΔECM_{t-1} were tested using both E-S and exE-S tests. The Tsay test and Hansen and Seo tests validated the results. The results are presented in Table 3.

Table 3. Results of testing for threshold cointegration for GDP *per capita* using E-S, exE-S, Tsay and Hansen and Seo tests

Threshold variable	Test	H_0^1: (ρ_1=ρ_2=0)	H_0^2: (ρ_1–ρ_2=0)	Tsay $H_0: \psi = 0$	H-S $H_0: A_1 = A_2$
ECM(*t*–1)	E-S	0.0002	0.6230	NA	NA
	exE-S	0.3712	NA		
ΔECM(*t*–1)	E-S	0.0003	0.6938	0.0037	0.9312
	exE-S	0.0021	0.0000		
R&D(*t* –2)	exE-S	0.0071	0.0000	0.0000	0.2461
ΔIRs(*t* –2)	exE-S	0.0110	0.0000	0.4825	0.0001
MilExp(*t* –3)	exE-S	0.0084	0.0000	0.0083	0.0110
EXR(*t* –4)	exE-S	0.0009	0.0000	0.0000	0.9999
Sav(*t* –4)	exE-S	0.0552	0.0035	0.3629	0.0001

Note. Only *p*-values are presented in the table. E-S – original Enders and Siklos test, exE-S – extended Enders and Siklos test, Tsay – Tsay test, H-S – Hansen and Seo test. NA – not available. The results indicating threshold cointegration are shadowed. Significance level: 5%.

Source: authors' calculations.

The results indicate that three out of four tests did not confirm threshold cointegration taking ECM_{t-1} and ΔECM_{t-1} with a zero threshold value. Only the Tsay test showed threshold cointegration for ΔECM_{t-1}. The proposed exE-S test displayed three possible threshold variables: $R\&D_{t-2}$, ΔIRs_{t-2}, and Sav_{t-4}. The values of the threshold were estimated and set at the following levels: 1.0300, 0.0002 and 3.0887, respectively. The first threshold variable was additionally confirmed by the Tsay test and the two other by the Hansen and Seo test. Empirical results should be confronted with economic facts and foundations. As the Israeli economy is based on innovations, both variables, R&D expenditures and savings, are reasonable. The short-term interest rate also refers to savings and investments. It is worth noting that R&D investment is closely related to government contracts and therefore they are also economically worthwhile (Lichtenberg, 1995).

5. Conclusions

In the paper, the power and size of the exE-S were analysed. A simulation experiment was conducted in order to present the advantages and limitations of the test. Moreover, an empirical example was provided. The results of both the simulation and empirical analysis are promising and allow formulating several conclusions. The power of the exE-S test is satisfactory for all parameter values, nevertheless, it depends on the difference between the ρ_1 and ρ_2 parameters. If the number of observations is relatively small (i.e. 50 and 100), the power is lower when the difference is 0.1. It corresponds to a weak asymmetry effect in the regimes and is similar to the E-S test results. However, the power of the exE-S is bigger than that of the E-S test. For greater values of $\rho_1 - \rho_2$ differences, the power of both tests is high.

The signs of short-run parameters in TECMs do not influence the results. The number of observations in each regime is not meaningful for power, however, it is important for the TECM model construction. The simulation results for size are slightly different. In the case of H_0^1, both the E-S and exE-S tests have their size close to the nominal one. The E-S test performs worse if the threshold variable is ΔECM_{t-1}. In the case of H_0^2, the E-S test preserves its size despite the parameters change. The exE-S test has its size higher than the assumed 1% and identical for the 5% significance level. In the case of the size of the extended test, the number of observations in regimes in the 80%–20% proportion decreases the size. The empirical example concerning economic growth in Israel indicates that the testing results using the exE-S test give an in-depth insight into threshold variables for the TECM model in comparison to the E-S test. The results are either supported by the Tsay or the Hansen and Seo test.

Using statistical tests in an empirical study is uncertain due to a low number of observations, differences between the model and the original data generating process, and many other circumstances. Therefore, it is recommended to apply a hierarchical procedure, i.e. to start with the E-S test first to recognise whether a threshold error cointegration around ECM_{t-1} (or ΔECM_{t-1}) exists. Then, one should search deeply for individual thresholds using the exE-S test. When the sample size is relatively small, extra caution in statistical inference is advised. The validation of the results with the use of other tests (i.e. Tsay test, Hansen and Seo test) concludes the process.

References

Aharoni, Y. (2014). *The Israeli Economy. Dreams and Realities*. Routledge. https://doi.org/10.4324/9781315863160.

Balke, N. S., & Fomby, T. B. (1997). Threshold cointegration. *International Economic Review*, *38*(3), 627–645. https://doi.org/10.2307/2527284.

Boehlke, J., Faldzinski, M., Galecki, M., & Osinska, M. (2017). Dynamics of Economic Growth in Ireland in 1980–2014. *St. Petersburg State Polytechnical University Journal. Economics*, *10*(2), 7–20. https://doi.org/10.18721/JE.10201.

Boehlke, J., Fałdziński, M., Gałecki, M., & Osińska, M. (2018). Economic growth in Ireland in 1980–2014. A threshold cointegration approach. *Argumenta Oeconomica*, (2), 157–188. https://doi.org/10.15611/aoe.2018.2.07.

Boehlke, J., Fałdziński, M., Gałecki, M., & Osińska, M. (2019). Econometric Analysis of Economic Miracles in Selected Economies Using TECM Approach. In M. Osińska (Ed.), *Economic Miracles in the European Economies* (pp. 175–230). Springer. https://doi.org/10.1007/978-3-030-05606-3_9.

Bruzda, J. (2007). *Procesy nieliniowe i zależności długookresowe w ekonomii. Analiza kointegracji nieliniowej*. Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics*, *21*(1), 520–533. https://doi.org/10.1214/aos/1176349040.

Chorev, S., & Anderson, A. R. (2006). Success in Israeli high-tech start-ups; Critical factors and process. *Technovation*, *26*(2), 162–174. https://doi.org/10.1016/j.technovation.2005.06.014.

Enders, W., & Siklos, P. L. (2001). Cointegration and Threshold Adjustment. *Journal of Business & Economic Statistics*, *19*(2), 166–176. https://doi.org/10.1198/073500101316970395.

Engle, R. F., & Granger, C. W. J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, *55*(2), 251–276. https://doi.org/10.2307/1913236.

Frey, G., & Manera, M. (2007). Econometric Models of Asymmetric Price Transmission. *Journal of Economic Surveys*, *21*(2), 349–415. https://doi.org/10.1111/j.1467-6419.2007.00507.x.

Gałecki, M., & Osińska, M. (2019). Threshold Error Correction Model: A Methodological Overview. In M. Osińska (Ed.), *Economic Miracles in the European Economies* (pp. 151–173). Springer. https://doi.org/10.1007/978-3-030-05606-3_8.

Ghassan, H. B., & Banerjee, P. K. (2015). A Threshold Cointegration Analysis of Asymmetric Adjustment of OPEC and non-OPEC Monthly Crude Oil Prices. *Empirical Economics*, *49*(1), 305–323. https://doi.org/10.1007/s00181-014-0848-0.

Gosińska, E., Leszkiewicz-Kędzior, K., & Welfe, A. (2020). Who is responsible for asymmetric fuel price adjustments? An application of the threshold cointegrated VAR model. *Baltic Journal of Economics*, *20*(1), 59–73. https://doi.org/10.1080/1406099X.2020.1746114.

Granger, C. W. J., & Lee, T. H. (1989). Investigation of Production, Sales and Inventory Relationships Using Multicointegration and Non-Symmetric Error Correction Models. *Journal of Applied Econometrics*, *4*(S1), 145–159. https://doi.org/10.1002/jae.3950040508.

Hansen, B. E., & Seo, B. (2002). Testing for two-regime threshold cointegration in vector error-correction models. *Journal of Econometrics, 110*(2), 293–318. https://doi.org/10.1016/S0304-4076(02)00097-0.

Hassouneh, I., Serra, T., Bojnec, Š., & Gil, J. M. (2017). Modelling price transmission and volatility spillover in the Slovenian wheat market. *Applied Economics*, *49*(41), 4116–4126. https://doi.org/10.1080/00036846.2016.1276273.

Hogg, R. V., McKean, J. W., & Craig, A. T. (2005). *Introduction to Mathematical Statistics* (6th edition). Upper Saddle River, Pearson Education.

Kapetanios, G., Shin, Y., & Snall, A. (2006). Testing for Cointegration in Nonlinear Smooth Transition Error Correction Models. *Econometric Theory*, *22*(2), 279–303. https://doi.org/10.1017/S0266466606060129.

Leszkiewicz-Kędzior, K., & Welfe, A. (2014). Asymmetric Price Adjustments in the Fuel Market. *Central European Journal of Economic Modelling and Econometrics*, (2), 105–127. https://doi.org/10.24425/cejeme.2014.119235.

Lichtenberg, F. R. (1995). The output contributions of computer equipment and personnel: A firm-level analysis. *Economics of Innovation and New Technology*, *3*(3–4), 201–218. https://doi.org/10.1080/10438599500000003.

Lloyd, C. J. (2006). Estimating test power adjusted for size. *Journal of Statistical Computation and Simulation*, *75*(11), 921–933. https://doi.org/10.1080/00949650412331321160.

Martens, M., Kofman, P., & Vorst, T. C. F. (1998). A threshold error-correction model for intraday futures and index returns. *Journal of Applied Econometrics*, *13*(3), 245–263. https://doi.org/10.1002/(SICI)1099-1255(199805/06)13:3%3C245::AID-JAE480%3E3.0.CO;2-E.

Piłatowska, M., & Włodarczyk, A. (2017). The Environmental Kuznets Curve in the CEE Countries – the Threshold Cointegration Approach. *Argumenta Oeconomica*, (2), 307–340. http://dx.doi.org/10.15611/aoe.2017.2.13.

Quandt, R. E. (1958). The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes. *Journal of the American Statistical Association, 53*(284), 873–880. https://doi.org/10.2307/2281957.

Stigler, M. (2010, January 10). *Threshold cointegration: overview and implementation in R.* https://cran.r-project.org/web/packages/tsDyn/vignettes/ThCointOverview.pdf.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press.

Trajtenberg, M. (2001). R&D Policy in Israel. In M. P. Feldman, & A. N. Link (Eds.), *Innovation policy in the knowledge-based economy* (pp. 409–454). Springer. https://doi.org/10.1007/978-1-4615-1689-7_18.

Tsay, R. S. (1998). Testing and Modeling Multivariate Threshold Models. *Journal of the American Statistical Association*, *93*(443), 1188–1202. https://doi.org/10.1080/01621459.1998.10473779.

Zilberfarb, B-Z. (2006). From Boom to Bust: The Israeli Economy 1990–2003. *Israel Affairs*, *12*(2), 221–233. https://doi.org/10.1080/13537120500535126.

Przegląd Statystyczny. Statistical Review, 2022, vol. 69, 1, 21–38 DOI: 10.5604/01.3001.0015.8791
Received: 2.06.2022, revised: 15.06.2022, accepted: 16.06.2022

REITs impact on typical investment portfolio – further evidence of the sector split importance

Jakub Pacholec[a]

Abstract. The REIT (Real Estate Investment Trust) returns demonstrate a time-varying linear correlation with various equity indexes, therefore they are fit for multi-asset portfolio enhancement. On the one hand, each REIT sector is characterised by a unique set of return properties, and on the other, companies within those sectors remain homogenous.

The aim of this research is twofold: firstly, to verify the earlier studies on how adding REITs to mixed equities/bonds portfolios affects their risk and return characteristics, and secondly, to contribute to these studies by examining the impact of adding different REIT sectors to such portfolios over a relatively long and more up-to-date sample, i.e. the period of 1990–2019.

The results indicate that, in contrast to what some previous studies suggested, adding the REIT index exposure leads to a limited portfolio enhancement only. More significant and consistent effects can be achieved by the inclusion of individual REIT sectors in an investment portfolio. Apartment REITs offered diversification benefits across the entire spectrum in all the periods, while Industrials were useful across the curve in 1990s and 2010s. Self-storage exposure, on the other hand, improved the investment portfolio performance in each of the studied decades. In general, it was enough for investors who strived for portfolio improvement over the three decades between 1990 and 2019 to have a small portion of their Value holdings replaced with the REIT sector exposure to obtain a positive impact on both the returns and the risk.

Keywords: REIT, real estate, portfolio

JEL: G11, R33, R39

1. Introduction

Real Estate Investment Trust (REIT) legislation has been to date introduced in 39 countries with 10 more currently in the planning or legislative process (Nareit, 2020). There were over 850 REITs worldwide at the end of 2019 (European Public Real Estate Association, 2019). It has become a popular investment tool in many developed economies – approximately 80 million Americans and 14 million Australians have exposure to REITs, which constitutes 24% and 57% of these countries' respective total populations (Nareit, 2020; Property Council of Australia, 2017). REITs account for approximately 2.7% of the global listed equity market cap (and even more for some local markets, e.g. over 7.2% in Australia). Although due to their systemic meaning, REIT risk and return characteristics have been widely studied in recent years, information on their standalone performance has only limited real-life investment application. This study focuses on the REIT behaviour in multi-asset portfolios,

[a] SGH Warsaw School of Economics, Institute of Econometrics, Financial Markets Modelling Unit, ul. Madalińskiego 6/8, 02-513 Warszawa, e-mail: jp50820@doktorant.sgh.waw.pl, ORCID: https://orcid.org/0000-0003-2144-6696.

namely in various portfolios consisting of bonds and equities. Given the fact that REIT returns show a time-varying correlation with equity indexes and bonds, they may be predisposed to serve as an effective tool in a diversified portfolio construction process. The 'uniqueness' of REIT characteristics is confirmed by studies of Liang and McIntosh (1998) or Chiang and Lee (2002), which point to a fact that replicating REIT returns with other asset classes is impossible.

There are numerous studies conducted by market practitioners and academics that emphasise a positive impact of including REITs in mixed-asset portfolios, vast majority of them implying that REITs are a valuable addition to multiple investment strategies, shifting efficient frontiers upward. Such conclusion may be vital for investment, pension funds and individual investors, as any investor looking for optimising their portfolio will find value in improving their investment efficiency. Individual REIT sectors expose unique price movement characteristics, and as such, they most probably could offer different benefits for portfolios at different risk levels, thus significantly expanding investors' set of tools.

The aim of this study is twofold: firstly, to verify earlier research on how adding REITs to a mixed-equities or mixed-bonds portfolio affects its risk and return characteristics. Secondly, to expand those studies by examining the impact of adding different REIT sectors to such portfolios over the period of the last 30 years. On the one hand, each REIT sector is characterised by a unique set of return properties, and on the other, companies within those sectors remain homogenous. The study is designed both to answer the question how the inclusion of REITs enhances the efficient frontier, and to help in choosing the best sectors for specific risk-tolerance strategies.

2. Literature review

The benefits from the diversification of a portfolio by including REITs have been investigated by numerous researchers. Kuhle (1987) examined monthly time series over the period of 1980–1985, concluding that adding REITs to an all-equity portfolio does not bring any benefits (although pure REIT portfolios are superior to pure equity portfolios in Markowitz's sense). Mueller et al. (1994), on the other hand, showed that the benefits depend on the time frame – REITs proved to be beneficial for the mixed-asset portfolio in the periods of 1976–1980 and 1990–1993, but not for the 1980–1990 period, which corresponds with the fact that REITs have undergone significant structural changes over the past decades.

Lee and Stevenson (2005) expanded previous studies by examining the impact of the duration of a holding period on the diversification benefits. They conclude, for example, that the attractiveness of REITs as diversification assets increases as the holding period is becoming longer. This is in line with studies showing that the correlation between REITs and the broad market is highest for short periods and decreases with the widening of the observation window. Interestingly, the results suggest that REITs not only enhance return at the lower end of the efficient frontier, but also reduce risk at the top end. Further, Lee (2010) concludes that the overall benefits of the inclusion of REITs in a portfolio depends both on the time frame and on individual portfolio constituents (asset classes).

Bhuyan et al. (2014) measured the impact of REITs on the optimal mixed-asset portfolio creation for investors with various degrees of risk-aversion, using a mean-variance utility function framework. Their findings suggest, among other things, that risk-averse investors should invest in REITs at the expense of bonds, and that the marginal effect of REIT returns on their optimal portfolio weights increases with risk tolerance.

Several authors dealt with the issue of the role of non-US REITs in the portfolio optimisation. Marzuki and Newell (2016) elaborated on the significance of the UK-REITs in a mixed-asset portfolio. Their study shows that over the 2007–2014 period, REITs delivered poor returns compared to the broad stock market (prior to the global financial crisis (GFC), after July 2009, the risk-adjusted return was strong), and generated virtually no diversification benefits for mixed-asset portfolio (both in the pre- and the post-GFC period). A study by Newell et al. (2013), on the other hand, found that French REITs (SIICs) contributed to the mixed-asset portfolio in the post-GFC era.

There are also numerous studies conducted by market practitioners, such as Wilshire (Nareit, 2016), Fidelity Investments (2016) or Morningstar (2016), which almost unanimously point to a significant contribution of REITs to the improvement of portfolio characteristics. It has to be borne in mind, however, that some of these studies was either sponsored by or written in collaboration with the REIT associations.

In a recent study, Ye and Song (2017) examined the diversification benefits of adding REITs to a portfolio, placing a special focus on the sector split (Hotel, Healthcare, Industrial, Retail, Diversified, Office, Self-storage, Manufactured Home, Apartments and Residential). Using daily, monthly and annual data from the period between 1998 and 2016, they examined the performance of the optimal mean-variance portfolios based on investors' risk aversion level and the access to different asset classes. The conclusions are threefold: Hotel and Self-Storage REITs have the highest potential for improving portfolio characteristics, Self-Storage, Manufactured and Residential REITs are becoming increasingly wealth-compensating as the risk aversion is increasing, and, finally, bonds downplay the importance of REITs for a moderate- or high-risk aversion.

This study's contribution to the literature is expanding the sector analysis. The mature US REIT market offers investors access to several REIT sectors with different risk and return characteristics, which should not be treated as one, uniform 'bag'. The sector focus became popular in the 1990s, so the time series were insufficient until recently, when studies including sector split started to appear.

Compared to the above-mentioned study by Ye and Song (2017), the author will add those REIT segments that play an important role in today's US REIT market, namely Tower, Data Centre and Timber. According to Green Street Advisors, non-traditional REIT sectors, as those mentioned above, constitute up to approximately 55% of total equity REIT market capitalisation. At the end of 2019, three out of five largest REITs in the US belonged to one of the added categories (NYSE:AMT, NYSE:CCI – Cell Tower and NASDAQ:EQIX – Data Centre). Additionally, high-tech sectors were among the best performing in the post-GFC period, so including them in the analysis can presumably yield interesting results in terms of the portfolio-performance metrics.

This study also updates and expands the sample base of earlier studies by using the data from the period between 1990 and 2019, presenting a coherent picture of the REITs' role in a mixed-asset portfolio over the last 30 years.

3. Data and methodology

This study has been based on the monthly total return data for US REITs over the period of 1990–2019. The sample consists of panel data of up to 91 REITs (depending on the availability) grouped into 14 equally-weighted sector indexes (Apartment, Data Centre, Healthcare, Hotel, Industrial, Mall, Manufactured Home, Office, Self-Storage, Single-Family Rental, Strip, Student Housing, Timber and Tower).

Given the fact that the idea behind this analysis is to introduce REITs from the perspective of an investor, the constituents were chosen using market capitalisation and liquidity criteria. Most importantly, a given stock should be investable and price dynamics should not be disturbed by low liquidity. An arbitrary threshold of market cap of USD 200 million and the average daily volume of trailing 30 days of USD 1 million at any given date were applied. FTSE NAREIT Equity REITs Index was used as a broad REIT market proxy.

The study was conducted from the standpoint of an American investor due to data availability, the elimination of currency risk exposure and, most notably, in order for the results to be comparable to other studies. For the purpose of stock market replication and in order to capture different styles, the author used four US domestic stock indexes: S&P 500 Growth and S&P500 Value for Large-Growth and Large-Value, respectively, and Russell 2000 Growth and Russell 2000 Value for Small-

Growth and Small-Value, respectively. Non-US domestic stocks were replicated by MSCI World ex-US Index (USD-denominated). The US government bonds of various maturities were used to address the fixed-income part of the portfolio. An arbitrary set of 2, 5 and 10–year maturities was chosen. All the equity data were the total returns coming from the Thomson Reuters Datastream, and the fixed income data came from the Federal Reserve Economic Data (Federal Reserve Bank of St. Louis).

There were several constraints imposed on the composition of portfolios. In order to stick to the mixed-asset framework, the share of each of the above-mentioned stock indexes should not exceed 70% of the portfolio, and the share of each bond maturity, as well as all of bonds combined, should not be larger than 50%. Due to the domestic preference, the MSCI World was capped at 30%. For the purpose of the analysis, the share of an individual REIT sector index was limited to 20% – this constraint was imposed to address the potential risk of concentrations. No short-selling strategies were applied.

The data sample was divided into three time periods: 1990–1999, 2000–2009 and 2010–2019. The 'Modern REIT Era' starts with the IPO boom in the 1990s. Prior to that time, the REIT market was in the development stage and was immature. Ott et al. (2005) differentiate between the old-REIT (pre-1992) and the new-REIT (post-1992) era. Their study reveals that the market grew significantly between the year 1992 and the time of their study. Moreover, capital management and the structure (including debt ratios), as well as internal business operations of REITs changed in the last three decades. What is also extremely important, in the 1990s, REITs became sector-oriented, substantially diverging from the diversified model.

The next period, 2000–2009, begins at the peak of the dot-com bubble and ends right after the start of the post-GFC rebound. The turn of the millennium was another mark in the REIT history. Sing et al. (2016) notes that the time-varying beta characteristics are fundamentally different prior to the 2000 and in the 2000s. Intriguingly, a strong downward trend in the equity REITs betas observed throughout the period of 1972–2000 seems to be substantially reversing after the year 2000.

The 2010–2019 period marks the post-GFC decade. In that period, REITs have expanded significantly both in terms of market cap and the number of entities. It was not only because the asset class experienced a meaningful market expansion, but also because REITs materially changed their capital structure (by lowering LTVs even more), and moved from the external to the internal management model.

A variance-covariance matrix is constructed for each period and an efficient (in Markowitz' sense) mixed-asset ex-REIT portfolio frontier is estimated (the *base*). Each *base* frontier consists of 10 efficient portfolios, ranging from the least volatile MVP (minimum variance portfolio) to the one producing largest returns (portfolio no. 10).

The returns and variance are calculated as follows:

$$r_p = w'r, \tag{1}$$

where

$$r = [r_1, \dots, r_n] \text{ and } w = [w_1, \dots, w_n], \tag{2}$$

and

$$\delta_p^2 = w'\Sigma w \tag{3}$$

given

$$r \sim N(\mu, \Sigma). \tag{4}$$

The *base* portfolios were estimated solving the optimisation problem:

$$\min_w \delta_p^2 = w'\Sigma w \tag{5}$$

under the conditions

$$\mu_p = w'\mu = \mu_i, \tag{6}$$

$$w'i = 1, \tag{7}$$

where i is the identity matrix. Using Lagrangian to solve constrained optimisation problem

$$\mathcal{L} = w'\Sigma w + \lambda_1(w'\mu - \mu_i) + \lambda_2(w'i - 1) \tag{8}$$

$$\nabla\mathcal{L} = 0, \tag{9}$$

so the first order conditions being:

$$\frac{\partial\mathcal{L}}{\partial w'} = 2\,\Sigma w + \lambda_1\mu + \lambda_2 i = 0, \tag{10}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = w'\mu - \mu_i = 0, \tag{11}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = w'i - 1 = 0 \tag{12}$$

with no $\mu_p = \mu_i$ condition for MVP and μ_i for portfolios no. 2–10 growing proportionally.

Then, similarly to Lee and Stevenson (2005), two sets of new efficient frontiers were calculated for each period. The reason behind it was to evaluate (a) if REITs as a whole contribute to portfolio diversification, (b) what REIT sectors would have been included in efficient portfolios, and (c) what is the impact of including specific REIT segments on the overall efficient portfolio characteristics. The first set was computed by fixing the standard deviation of the base portfolios, and then re-estimating the portfolios by adding the REIT index and 14 different sector indexes, one at a time. Such an approach made it possible to see if there was any impact of specific sector inclusion on portfolio returns, and if so, how significant it was.

The second set was produced by fixing the returns of the first efficient frontier and following the same procedure which, analogically to the above, serves the purpose of examining the potential of REITs in terms of the portfolio volatility reduction.

Apart from the three *base* (ex-REIT) frontiers, there were 30 frontiers computed for each time period, each consisting of 10 portfolios, accumulating up to 903 portfolios in total.

4. Results

Table 1 presents the main statistics of the first efficient frontier (years 1990–1999). The high end of the curve is dominated by growth stocks (as the period ends in the eve of the dot-com bubble burst). Interestingly, the small-cap growth stocks index did not make it to any of the portfolios, while the small-cap value stocks have a place at the lower end of the frontier (together with non-US index). All the data presented throughout the study are on a monthly basis.

Table 1. Efficient frontier (*base*) for the 1990–1999 period with portfolio holding weights

Portfolio	MVP	2	3	4	5	6	7	8	9	10
	%									
μ_p	0.78	0.86	0.93	1.00	1.08	1.15	1.23	1.30	1.37	1.45
δ_p	1.87	1.89	1.96	2.07	2.29	2.58	2.86	3.19	3.56	3.94
RUSSELL 2000 Value	16.8	13.2	9.6	4.7	0.0	0.0	0.0	0.0	0.0	0.0
RUSSELL 2000 Growth	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S&P 500 Value	11.0	9.5	8.0	3.0	0.0	0.0	0.0	7.5	18.6	30.0
S&P 500 Growth	9.6	19.7	29.6	42.3	53.4	60.6	67.7	70.0	70.0	70.0
MSCI WORLD ex-US	12.5	7.6	2.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
US 2 Y	50.0	50.0	50.0	50.0	46.6	39.4	32.3	22.5	11.4	0.0
US 5 Y	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
US 10 Y	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Source: author's calculations based on Thomson Reuters data.

Adding the REIT exposure to the efficient portfolios in 1990s had an impact on their overall performance. Tables 2 and 3 show the share of individual indexes in efficient portfolios (upper tables) and their impact on returns and volatility.

What is interesting is that NAREIT index found a place in efficient portfolios only at the lower end of the frontier at the expense of the Russell 2000 Value and the S&P Value indexes, and it was a risk-reduction scenario where the REITs' share was larger (e.g. 25.4% vs 11.8% for MVP). It can be stated that in the 1990s, REITs were a valuable addition to low-risk portfolios, successfully replacing value stocks.

The benefits of the REITs' inclusion, however, differ significantly depending on the sector exposure an investor could have had. In general, for the analysed period, REITs tend to be a valuable addition to a portfolio rather at the lower end of the spectrum – this is where the largest risk reduction and return enhancement is visible, and where individual REIT sectors replace both value indexes and bonds.

There are three notable exceptions among the classic real estate segments, namely the Office, Apartment and the Industrial segments. Each of them constituted a share of the efficient portfolio close or at the upper constraint (20%) across the whole curve. Those sectors also displayed the largest diversification benefits – although it is worth noting that in terms of the return enhancement, this was the case rather in low-risk portfolios (23–26 bps for MVP and 6–7 bps for portfolio no. 9), whereas the highest risk reduction could have been observed for high-volatility portfolios (40–52 bps for portfolio no. 9 and 28–31 bps for MVP). None of the REIT sectors improved the characteristics of the riskiest portfolio.

Healthcare and Timber REITs also had a slight, yet consistent impact on the efficient portfolio performance across the entire risk spectrum. Investors of various risk tolerance levels would have benefitted by dedicating approximately one tenth of their portfolios to those segments.

The overall effect of the Manufactured Housing was marginal and applied only to the lowest-yielding portfolios; however, a notice of short sample (only two years) has to be taken into account.

Table 2. Share of an individual index in the efficient portfolio (upper part) and the return improvement (lower part), 1990–1999

Portfolio	MVP	2	3	4	5	6	7	8	9	10
	%									
NAREIT	11.77	10.63	7.59	3.07	0.00	0.00	0.00	0.00	0.00	0.00
Apartment	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	12.71	0.00
Healthcare	12.06	11.70	7.42	7.41	8.35	9.56	10.75	16.34	18.92	0.00
Hotel	5.34	5.24	4.88	2.98	3.20	3.48	3.76	4.74	5.27	0.00
Industrial	15.00	15.19	15.72	16.60	18.31	20.00	20.00	20.00	12.10	0.00
Mall	7.03	6.39	6.59	6.91	7.53	8.32	9.11	10.25	10.72	0.00
Manufactured Home	7.37	6.75	5.00	2.11	0.00	0.00	0.00	0.00	0.00	0.00
Office	16.61	16.82	17.40	18.36	20.00	20.00	20.00	20.00	11.69	0.00
Self-Storage	11.31	11.24	7.44	5.71	6.31	7.09	7.87	12.42	13.68	0.00
Strip	7.59	6.88	6.00	6.29	6.85	7.57	8.30	10.07	10.71	0.00
Timber	8.54	7.59	6.70	6.98	7.52	8.21	8.91	12.11	12.41	0.00
NAREIT	0.16	0.10	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Apartment	0.26	0.19	0.14	0.10	0.09	0.10	0.10	0.11	0.07	0.00
Healthcare	0.18	0.12	0.07	0.03	0.01	0.01	0.02	0.03	0.03	0.00
Hotel	0.14	0.09	0.06	0.02	0.01	0.01	0.01	0.01	0.01	0.00
Industrial	0.23	0.17	0.11	0.07	0.06	0.07	0.07	0.08	0.06	0.00
Mall	0.19	0.12	0.07	0.02	0.01	0.01	0.01	0.01	0.01	0.00
Manufactured Home	0.14	0.08	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Office	0.24	0.17	0.12	0.08	0.07	0.07	0.08	0.09	0.06	0.00
Self-Storage	0.17	0.11	0.07	0.02	0.01	0.01	0.01	0.02	0.02	0.00
Strip	0.19	0.12	0.07	0.02	0.01	0.01	0.01	0.01	0.01	0.00
Timber	0.19	0.13	0.07	0.03	0.01	0.01	0.01	0.02	0.02	0.00

Source: author's calculations based on Thomson Reuters data.

Table 3. Share of an individual index in the efficient portfolio (upper part) and the volatility improvement (lower part), 1990–1999

Portfolio	1	2	3	4	5	6	7	8	9	10
	%									
NAREIT	25.4	21.9	13.5	4.6	0.0	0.0	0.0	0.0	0.0	0.0
Apartment	18.3	19.9	20.0	19.0	20.0	20.0	20.0	20.0	20.0	0.0
Healthcare	13.7	13.4	12.6	7.0	8.2	9.3	10.5	11.7	17.9	0.0
Hotel	5.5	5.4	5.3	4.7	3.2	3.5	3.7	4.7	5.2	0.0
Industrial	16.5	17.0	17.1	14.7	16.7	18.7	20.0	20.0	20.0	0.0
Mall	18.5	17.0	11.5	6.7	7.4	8.2	9.0	10.2	10.7	0.0
Manufactured Home	16.5	11.8	7.1	2.4	0.0	0.0	0.0	0.0	0.0	0.0
Office	17.7	18.7	19.1	16.1	18.2	20.0	20.0	20.0	20.0	0.0
Self-Storage	10.3	13.1	14.8	10.3	11.0	12.4	13.8	18.1	20.0	0.0
Strip	19.0	16.4	9.7	3.4	3.7	4.0	4.3	3.1	2.8	0.0
Timber	12.0	11.5	9.2	5.9	6.5	7.1	7.6	8.2	10.3	0.0
NAREIT	-0.24	-0.19	-0.13	-0.04	0.00	0.00	0.00	0.00	0.00	0.00
Apartment	-0.28	-0.33	-0.37	-0.33	-0.31	-0.34	-0.36	-0.42	-0.52	0.00
Healthcare	-0.17	-0.16	-0.15	-0.10	-0.04	-0.05	-0.06	-0.11	-0.15	0.00
Hotel	-0.10	-0.10	-0.09	-0.07	-0.03	-0.03	-0.03	-0.04	-0.05	0.00
Industrial	-0.30	-0.32	-0.33	-0.25	-0.21	-0.23	-0.26	-0.32	-0.40	0.00
Mall	-0.32	-0.30	-0.23	-0.09	-0.03	-0.03	-0.04	-0.04	-0.04	0.00
Manufactured Home	-0.34	-0.19	-0.08	-0.01	0.00	0.00	0.00	0.00	0.00	0.00
Office	-0.31	-0.34	-0.36	-0.27	-0.23	-0.26	-0.28	-0.35	-0.42	0.00
Self-Storage	-0.08	-0.10	-0.12	-0.09	-0.03	-0.03	-0.04	-0.08	-0.09	0.00
Strip	-0.33	-0.30	-0.23	-0.09	-0.03	-0.04	-0.04	-0.08	-0.09	0.00
Timber	-0.24	-0.26	-0.25	-0.12	-0.07	-0.07	-0.08	-0.13	-0.15	0.00

Source: author's calculations based on Thomson Reuters data.

As shown in Table 4, effective portfolios in the 2000s were dominated by a combination of Small Value stocks and bonds. It should be noted that the US stocks outperformed the ex-US ones, and Large Growth stocks found a place in low-risk portfolios. Unlike in the previous decade, the addition of NAREIT index played some role in the return enhancement only in the highest end of the curve. Interestingly, the aggregate REIT index again substituted Value stocks. It is worth noting, though, that the overall portfolio characteristics improvement was limited (0.05% of additional return and 0.29% reduction of standard deviation).

Table 4. Efficient frontier (*base*) for the 2000–2009 period with portfolio holding weights

Portfolio	MVP	2	3	4	5	6	7	8	9	10
	%									
μ_p	0.24	0.28	0.32	0.36	0.40	0.45	0.49	0.53	0.57	0.61
δ_p	2.20	2.21	2.23	2.27	2.32	2.39	2.47	2.56	2.68	3.88
RUSSELL 2000 Value	9.7	15.0	20.2	25.0	29.2	33.4	37.4	41.4	49.1	70.0
RUSSELL 2000 Growth	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S&P 500 Value	9.8	7.3	4.6	1.7	0.0	0.0	0.0	0.0	0.0	0.0
S&P 500 Growth	30.5	27.7	24.5	20.9	16.3	11.3	6.0	0.7	0.0	0.0
MSCI WORLD ex-US	0.0	0.0	0.7	2.4	4.5	5.3	6.6	7.9	0.9	0.0
US 2 Y	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
US 5 Y	49.1	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	0.0
US 10 Y	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	30.0

Source: author's calculations based on Thomson Reuters data.

An interesting picture arises from the analysis of the inclusion of REIT sectors in an investment portfolio. On the one hand, the share of individual REIT segments grow in efficient portfolios as we move upwards the risk spectrum, but on the other, their impact moves right the opposite way. The REIT characteristics were, once again, most valuable at the lower end of the frontier.

The 2000–2009 decade was another one when Office REITs contributed the largest benefits to the portfolio. Student Housing, Strip Centres and Apartments were also among the most effective segments across the entire frontier.

Unlike the REIT Index, several sectors would only have worked for more risk-averse investors. The Hotel, Industrial, Timber and Tower sectors found a place only at the lower end of the curve of the efficient portfolios, and the impact of the two latter segments was rather limited straight from the MVP. Those were predominantly a replacement for Large Growth stocks.

It is worth noting that in the analysed period, REITs played a larger role in the overall portfolio risk reduction than in 1990s, i.e. an average of 0.46% compared to 0.16% a decade earlier. REITs generally decreased the overall risk at the lower end of the spectrum, predominantly replacing riskier Growth stocks and, to some minor extent, 5Y Bonds. Additionally, compared to the previous decade, numerous REIT segments were present in portfolio no. 10, where they replaced Value stocks.

An impact of Data Centre and Tower REITs is muted. The burst of dot-com bubble and limited trust for any high-tech names afterwards contributed to their relatively poor performance.

Table 5. Share of an individual index in the efficient portfolio (upper part) and the return improvement (lower part), 2000–2009

Portfolio	MVP	2	3	4	5	6	7	8	9	10
	%									
NAREIT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	35.0
Apartment	14.6	14.7	15.0	15.4	15.8	16.4	17.1	20.0	20.0	20.0
Data Centre	0.4	0.3	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HC	9.0	9.6	11.1	13.0	15.0	17.1	19.2	20.0	20.0	20.0
Hotel	7.4	7.4	7.4	7.4	7.4	7.4	7.4	3.4	0.3	0.0
Industrial	12.5	12.5	12.6	12.7	11.1	7.6	4.9	2.6	0.3	0.0
Mall	9.9	10.0	10.2	10.6	11.0	11.6	14.8	17.9	20.0	20.0
Manufactured Home	14.7	14.9	15.2	15.7	16.4	17.2	20.0	20.0	20.0	20.0
Office	16.2	16.3	16.5	17.8	20.0	20.0	20.0	20.0	20.0	20.0
Self-Storage	6.2	6.4	7.0	7.6	8.3	9.1	10.0	10.8	12.4	20.0
Strip	17.2	17.3	17.6	19.1	20.0	20.0	20.0	20.0	20.0	20.0
Student Housing	17.9	17.9	18.2	18.5	18.9	19.4	20.0	20.0	20.0	20.0
Timber	11.0	11.0	11.1	11.0	11.0	11.0	6.5	3.2	0.3	0.0
Tower	1.9	1.9	2.0	2.1	2.2	2.4	2.5	2.6	0.4	0.0
NAREIT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05
Apartment	0.23	0.19	0.16	0.14	0.13	0.12	0.11	0.10	0.06	0.06
Data Centre	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HC	0.15	0.13	0.12	0.13	0.13	0.14	0.15	0.16	0.17	0.17
Hotel	0.18	0.15	0.12	0.10	0.08	0.07	0.06	0.04	0.00	0.00
Industrial	0.26	0.22	0.19	0.17	0.14	0.11	0.08	0.04	0.00	0.00
Mall	0.27	0.23	0.20	0.18	0.17	0.16	0.09	0.11	0.08	0.08
Manufactured Home	0.20	0.16	0.14	0.12	0.11	0.10	0.09	0.08	0.04	0.04
Office	0.34	0.30	0.27	0.25	0.21	0.17	0.14	0.10	0.06	0.06
Self-Storage	0.09	0.06	0.04	0.04	0.04	0.03	0.03	0.03	0.04	0.07
Strip	0.37	0.33	0.30	0.28	0.24	0.20	0.17	0.13	0.09	0.09
Student Housing	0.33	0.29	0.24	0.20	0.17	0.13	0.09	0.05	0.01	0.01
Timber	0.23	0.20	0.16	0.14	0.12	0.10	0.07	0.04	0.00	0.00
Tower	0.07	0.04	0.03	0.02	0.02	0.02	0.02	0.01	0.00	0.00

Source: author's calculations based on Thomson Reuters data

Table 6. Share of an individual index in the efficient portfolio (upper part) and the volatility reduction (lower part), 2000–2009

Portfolio	MVP	2	3	4	5	6	7	8	9	10
NAREIT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	29.1
Apartment	14.4	12.7	14.3	14.1	13.5	15.4	16.2	17.0	17.7	18.0
Data Centre	0.4	0.3	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HC	8.9	8.7	9.4	12.6	13.9	16.2	18.9	16.4	18.9	16.3
Hotel	7.3	6.4	6.1	7.2	6.7	6.6	6.2	3.2	0.3	0.0
Industrial	11.4	11.3	10.8	12.2	9.1	6.4	4.4	2.4	0.2	0.0
Mall	9.7	9.3	9.3	9.1	9.0	10.0	14.6	15.1	18.9	18.8
Manufactured Home	12.7	13.3	13.3	15.0	15.6	17.0	16.3	17.6	19.9	18.8
Office	13.1	14.0	13.4	17.2	17.6	18.2	18.5	19.0	17.8	17.2
Self-Storage	5.5	6.1	6.1	7.3	7.2	7.4	9.1	9.9	10.4	18.2
Strip	15.6	14.7	15.4	16.8	19.4	18.5	18.1	19.1	19.9	16.0
Student Housing	17.8	16.7	17.6	15.3	15.2	16.4	16.6	17.9	16.4	19.7
Timber	10.5	10.0	11.0	9.7	10.3	10.5	5.3	3.0	0.3	0.0
Tower	1.6	1.6	1.8	1.8	1.9	2.1	2.2	2.2	0.4	0.0
NAREIT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.29
Apartment	-1.07	-0.89	-0.75	-0.64	-0.56	-0.50	-0.46	-0.40	-0.25	-0.34
Data Centre	-0.22	-0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HC	-0.85	-0.68	-0.61	-0.59	-0.58	-0.58	-0.59	-0.61	-0.61	-0.83
Hotel	-0.96	-0.76	-0.60	-0.48	-0.39	-0.31	-0.26	-0.16	-0.02	0.00
Industrial	-1.14	-0.97	-0.83	-0.71	-0.61	-0.47	-0.33	-0.18	-0.02	0.00
Mall	-1.16	-0.99	-0.86	-0.76	-0.68	-0.63	-0.34	-0.45	-0.33	-0.46
Manufactured Home	-1.00	-0.81	-0.67	-0.57	-0.49	-0.44	-0.40	-0.34	-0.19	-0.25
Office	-1.29	-1.15	-1.02	-0.92	-0.80	-0.67	-0.54	-0.40	-0.26	-0.34
Self-Storage	-0.58	-0.37	-0.27	-0.22	-0.19	-0.17	-0.16	-0.16	-0.16	-0.40
Strip	-1.34	-1.20	-1.08	-0.99	-0.87	-0.75	-0.63	-0.50	-0.37	-0.49
Student Housing	-1.27	-1.11	-0.96	-0.82	-0.67	-0.53	-0.38	-0.22	-0.06	-0.05
Timber	-1.09	-0.91	-0.75	-0.63	-0.52	-0.44	-0.33	-0.18	-0.02	0.00
Tower	-0.50	-0.28	-0.18	-0.13	-0.10	-0.09	-0.08	-0.07	-0.02	0.00

Source: author's calculations based on Thomson Reuters data.

The efficient frontier in the last analysed period (depicted in Table 7) was dominated by the Large Value which was a part of the portfolios across the entire curve. Interestingly, global stocks were not present in any of the portfolios. Maximal return was achieved as a combination of the Large and Small Value, which is counterintuitive given the performance of tech-oriented growth stocks in recent years.

Table 7. Efficient frontier (base) for the 2010–2019 period with portfolio holding weights

Portfolio	MVP	2	3	4	5	6	7	8	9	10
μ_p	0.69	0.74	0.78	0.83	0.88	0.92	0.97	1.02	1.06	1.11
δ_p	1.55	1.60	1.76	1.95	2.17	2.41	2.71	3.02	3.36	3.92
RUSSELL 2000 Value	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	30.0
RUSSELL 2000 Growth	34.2	9.9	0.0	0.0	0.0	1.7	9.6	17.5	25.3	0.0
S&P 500 Value	15.8	40.1	53.5	59.5	65.4	70.0	70.0	70.0	70.0	70.0
S&P 500 Growth	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
MSCI WORLD ex-US	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
US 2 Y	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
US 5 Y	50.0	50.0	46.5	40.6	34.6	28.3	20.4	12.5	4.7	0.0
US 10 Y	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Source: author's calculations based on Thomson Reuters data.

As shown in Table 8 and Table 9, the addition of the aggregate REIT Index in the analysed period contributed the largest benefits to the upper parts of the frontier, apart from the highest-return portfolio. It is worth noting, however, that although the index becomes a part of the portfolios, the impact on return characteristics is marginal. Once again, individual REIT segments are a more valuable portfolio addition than the aggregate REIT index. In fact, for investors across the entire risk spectrum, any REIT sector had a stronger impact on overall performance that the REIT aggregate index.

It is worth noting that in the period of 2010–2019, for the third decade in a row, REITs replaced Value stocks in efficient portfolios (and to some extent they also replaced bonds at the lower end of the frontier). The REIT sectors gave the best results in the higher risk portfolios and their share averaged low-to-mid teens, depending on the segment. In other words, for the analysed period, a risk-tolerant investor should have replaced a portion of their Value stocks with individual REIT sector exposure.

The highest benefits for the portfolios were offered by high-tech REIT segments, such as Data Centres and Cell Towers, which were effective across the entire curve both in terms of the volatility reduction and the return improvement (yet with the exception of Towers not included in the riskiest portfolio in terms of the value enhancement, Self-Storage and Industrials).

A systematic shift in shopping conditions set a challenging environment for the Mall REITs performance, making them rather a redundant tool in the portfolio diversification. This stays in vivid contrast to the previous decades, when Malls offered stable benefits across the entire curve. Despite a good standalone performance, the addition of the Single Family segment would not result in any substantial portfolio improvement (or, if any, then possibly in MVPs).

The Office segment, which was a valuable addition in 1990s and 2000s, became not such a good choice in 2010s. The Apartment segment, on the other hand, was the only one that turned out to have a positive, consistent and meaningful impact on the portfolio characteristics throughout the entire sample, except for the return improvement in the most aggressive strategies.

Table 8. Share of an individual index in the efficient portfolio (upper part) and the return improvement (lower part), 2010–2019

Portfolio	1	2	3	4	5	6	7	8	9	10
NAREIT	0.0	0.0	0.0	0.0	0.0	1.9	8.9	13.9	19.0	0.0
Apartment	14.7	15.6	18.9	20.0	20.0	20.0	20.0	20.0	13.9	0.0
Data Centre	10.3	11.5	14.4	17.7	20.0	20.0	20.0	20.0	20.0	20.0
Healthcare	4.5	3.2	4.0	5.4	6.7	8.1	14.2	17.2	8.8	0.0
Hotel	5.2	5.4	6.0	6.7	7.4	8.2	11.8	13.5	14.1	0.0
Industrial	13.7	14.5	17.0	19.8	20.0	20.0	20.0	20.0	20.0	0.0
Mall	2.1	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Manufactured Home	11.7	13.5	16.9	20.0	20.0	20.0	20.0	20.0	20.0	20.0
Office	5.6	5.7	6.0	6.5	6.9	7.3	11.0	12.4	6.1	0.0
Self-Storage	14.3	15.4	18.9	20.0	20.0	20.0	20.0	20.0	20.0	0.0
Single-Family	4.3	3.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Strip	7.4	7.8	8.9	10.2	11.6	12.9	18.6	20.0	8.9	0.0
Student Housing	4.8	3.0	3.6	4.9	6.2	7.5	15.0	18.3	7.8	0.0
Timber	4.8	4.7	4.4	4.1	3.7	4.3	6.5	6.7	6.9	0.0
Tower	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0	15.0	0.0
NAREIT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00
Apartment	0.10	0.07	0.08	0.09	0.10	0.10	0.12	0.12	0.08	0.00
Data Centre	0.09	0.06	0.07	0.08	0.10	0.11	0.13	0.14	0.13	0.03
Healthcare	0.04	0.01	0.00	0.01	0.01	0.01	0.02	0.02	0.02	0.00
Hotel	0.06	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.00
Industrial	0.13	0.10	0.11	0.12	0.13	0.14	0.16	0.16	0.12	0.00
Mall	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Manufactured Home	0.08	0.05	0.07	0.09	0.11	0.12	0.14	0.15	0.15	0.04
Office	0.05	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00
Self-Storage	0.12	0.09	0.11	0.13	0.14	0.15	0.17	0.17	0.12	0.00
Single-Family	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Strip	0.05	0.02	0.02	0.02	0.02	0.02	0.03	0.04	0.02	0.00
Student Housing	0.04	0.01	0.00	0.00	0.00	0.01	0.01	0.02	0.01	0.00
Timber	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tower	0.14	0.12	0.12	0.12	0.12	0.13	0.14	0.14	0.10	0.00

Source: author's calculations based on Thomson Reuters data.

Table 9. Share of an individual index in the efficient portfolio (upper part) and the volatility reduction (lower part), 2010–2019

Portfolio	1	2	3	4	5	6	7	8	9	10
NAREIT	0.0	0.0	0.0	0.0	0.0	1.8	8.6	13.3	18.0	0.0
Apartment	13.5	12.6	14.4	17.0	20.0	20.0	20.0	20.0	20.0	20.0
Data Centre	8.1	8.4	10.7	13.3	15.8	18.3	20.0	20.0	20.0	20.0
Healthcare	5.9	4.0	3.9	5.2	6.5	7.8	12.4	15.8	18.5	0.0
Hotel	6.3	5.6	5.8	6.5	7.2	7.9	9.9	12.5	14.1	0.0
Industrial	10.4	10.7	12.4	14.2	16.4	18.6	20.0	20.0	20.0	20.0
Mall	3.7	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Manufactured Home	10.2	10.2	12.9	15.7	18.4	20.0	20.0	20.0	20.0	20.0
Office	8.1	5.6	6.0	6.0	6.8	7.2	10.7	12.1	8.6	0.0
Self-Storage	12.3	12.4	13.0	15.4	18.1	20.0	20.0	20.0	20.0	20.0
Single-Family	9.1	3.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Strip	9.0	7.2	8.4	9.7	11.0	12.3	13.5	19.3	17.5	0.0
Student Housing	7.0	4.2	3.5	4.8	6.1	7.3	14.1	17.1	14.1	0.0
Timber	8.6	4.8	4.5	4.1	3.7	3.4	6.4	6.7	6.9	0.0
Tower	15.1	15.5	17.2	19.9	20.0	20.0	20.0	20.0	20.0	20.0
NAREIT	0.00	0.00	0.00	0.00	0.00	0.00	-0.02	-0.04	-0.06	0.00
Apartment	-0.18	-0.18	-0.22	-0.29	-0.36	-0.44	-0.55	-0.66	-0.78	-1.08
Data Centre	-0.07	-0.10	-0.19	-0.26	-0.33	-0.42	-0.55	-0.69	-0.82	-1.17
Healthcare	-0.05	-0.02	-0.01	-0.02	-0.03	-0.05	-0.11	-0.15	-0.18	0.00
Hotel	-0.10	-0.07	-0.06	-0.07	-0.08	-0.10	-0.16	-0.19	-0.22	0.00
Industrial	-0.17	-0.21	-0.30	-0.38	-0.46	-0.54	-0.68	-0.82	-0.95	-1.30
Mall	-0.06	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Manufactured Home	-0.06	-0.09	-0.17	-0.25	-0.34	-0.44	-0.58	-0.72	-0.85	-1.20
Office	-0.13	-0.05	-0.03	-0.03	-0.03	-0.04	-0.06	-0.07	-0.07	0.00
Self-Storage	-0.14	-0.17	-0.26	-0.35	-0.45	-0.56	-0.70	-0.84	-0.98	-1.33
Single-Family	-0.13	-0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Strip	-0.14	-0.08	-0.07	-0.08	-0.09	-0.12	-0.19	-0.24	-0.27	0.00
Student Housing	-0.06	-0.02	-0.01	-0.01	-0.02	-0.03	-0.09	-0.12	-0.13	0.00
Timber	-0.11	-0.04	-0.02	-0.01	-0.01	-0.02	-0.02	-0.02	-0.02	0.00
Tower	-0.19	-0.23	-0.33	-0.42	-0.49	-0.57	-0.67	-0.78	-0.89	-1.22

Source: author's calculations based on Thomson Reuters data.

5. Conclusions

There are several interesting conclusions to be drawn. First of all, in contrast to some previous studies (Lee, 2010; Lee & Stevenson, 2005), the author's analysis demonstrated that the extent to which portfolio's properties were enhanced by a broad REIT index exposure was limited. In the 1990s, the aggregate REIT exposure provided benefits for low-risk portfolios. Between the year 2000 and 2019, the impact of the REIT index inclusion was marginal and visible only in the upper parts of the spectrum. This is a meaningful conclusion for the contemporary market practitioners – the REIT market exposure *per se* did not improve efficient portfolios in the last 20 years.

A larger, more meaningful and more consistent effects can be achieved by the inclusion of the individual REIT sectors in an investment portfolio. Apartment REITs offered diversification benefits across the entire spectrum and throughout all the periods. Industrials were useful across the curve in the 1990s and the 2010s, but in the 2000s worked well only in lower risk portfolios. Self-Storage exposure improved the performance in each decade, although its impact was slightly smaller in the 2000s than in the other analysed decades. The benefits of the inclusion of the Office segment were not visible in last decade to such an extent as the in the earlier periods. The last analysed period was skewed towards the high-tech REITs, namely the Data Centres and the Cell Towers. The Mall REITs offered no significant benefits. Also, in contrast to the observations by Ye and Song (2017), the Hotel REITs provided only moderate (yet consistent over all the three time periods) improvement of the portfolio characteristics.

The optimal share of REITs in a portfolio is still a matter of discussion. It hit the upper boundaries (20) numerous times throughout the study, but a high share of REITs did not always result in increased benefits for a portfolio. What is interesting is that REITs were a replacement for Value stocks in most of the cases. In general, if investors strived for a portfolio improvement over the past three decades, they should have replaced a small portion of their Value holdings with the REIT sector exposure, as it had a positive impact on both the returns and the risk. Further studies on the REIT portfolio inclusion, especially in the COVID-19–impacted environment (post-2020), are recommended.

References

Bhuyan, R., Kuhle, J., Ikromov, N., & Chiemeke, C. (2014). Optimal Portfolio Allocation among REITs, Stocks, and Long-Term Bonds: An Empirical Analysis of US Financial Markets. *Journal of Mathematical Finance*, *4*(2), 104–112. https://doi.org/10.4236/jmf.2014.42010.

Chiang, K. C. H., & Lee, M-L. (2002). REITs in the decentralized investment industry. *Journal of Property Investment & Finance*, *20*(6), 496–512. https://doi.org/10.1108/14635780210446496.

European Public Real Estate Association. (2019, July). *Global Real Estate Total Markets Table*. https://prodapp.epra.com/media/EPRA_Total_Markets_Table_-_Q2–2019_-_rev_1564490547486.pdf.

Fidelity Investments. (2016, April). *REIT Stocks: An Underutilized Portfolio Diversifier*. https://www.reit.com/sites/default/files/media/PDFs/Research/REITStocksAnUnderutilizedPortfolioDiversifier_Fidelity.PDF.

Kuhle, J. L. (1987). Portfolio Diversification and Return Benefit – Common Stocks vs. Real Estate Investment Trusts (REITs). *Journal of Real Estate Research*, *2*(2), 1–9. https://doi.org/10.1080/10835547.1987.12090535.

Lee, S. L. (2010). The Changing Benefit of REITs to the Mixed-Asset Portfolio. *The Journal of Real Estate Portfolio Management*, *16*(3), 201–216. https://doi.org/10.1080/10835547.2010.12089876.

Lee, S., & Stevenson, S. (2005). The Case for REITs in the Mixed-Asset Portfolio in the Short and Long Run. *Journal of Real Estate Portfolio Management*, *11*(1), 55–80. https://doi.org/10.1080/10835547.2005.12089711.

Liang, Y., & McIntosh, W. (1998). REIT Style and Performance. *Journal of Real Estate Portfolio Management*, *4*(1), 69–78. https://doi.org/10.1080/10835547.1998.12089552.

Marzuki, J., & Newell, G. (2016). The significance and performance of UK-REITs in a mixed-asset portfolio. *Journal of European Real Estate Research*, *9*(2), 171–182. https://doi.org/10.1108/JERER-08-2015-0032.

Morningstar. (2016). *The Role of REITs in a Portfolio. Potential to increase returns or reduce risk, 1972–2015*. https://www.reit.com/sites/default/files/media/PDFs/REITFactSheetDiversification1_16.pdf.

Mueller, G. R., Pauley, K. R., & Morrill, W. K. (1994). Should REITs be Included in a Mixed-Asset Portfolio?. *Real Estate Finance*, *11*(1), 23–28.

Nareit. (2016). *Wilshire Research: REITs Were A Key to Increasing Retirement Income*. Retrieved January 20, 2021, from https://www.reit.com/data-research/research/wilshire-research-reits-were-key-increasing-retirement-income.

Nareit. (2020). *Global Real Estate Investment*. Retrieved October 10, 2020, from https://www.reit.com/investing/global-real-estate-investment.

Newell, G., Adair, A., & Kim Nguyen, T. (2013). The significance and performance of French REITs (SIICs) in a mixed-asset portfolio. *Journal of Property Investment & Finance*, *31*(6), 575–588. https://doi.org/10.1108/JPIF-01–2011–0004.

Ott, S. H., Riddiough, T. J., & Yi, H.-C. (2005). Finance, Investment and Investment Performance: Evidence from the REIT Sector. *Real Estate Economics*, *33*(1), 203–235. https://doi.org/10.1111/j.1080-8620.2005.00117.x.

Property Council of Australia. (2017, April 19). *Stapled Structures Consultation Paper*. https://treasury.gov.au/sites/default/files/2019-03/c2017-t240634-Property-Council.pdf.

Sing, T. F., Tsai, I. C., & Chen, M. C. (2016). Time-Varying Betas of US REITs from 1972 to 2013. *The Journal of Real Estate Finance and Economics*, *52*(1), 50–72. https://doi.org/10.1007/s11146–015–9502–7.

Ye, Z., & Song, D. (2017). *Diversification Benefits of REITs in Portfolio Allocation by REIT Property Types*. https://econ.unc.edu/wp-content/uploads/sites/38/2017/09/Mayo_Ye_Song_2017.pdf.

Przegląd Statystyczny. Statistical Review, 2022, vol. 69, 1, 39–49 DOI: 10.5604/01.3001.0015.8792
Received: 15.06.2022, revised: 23.06.2022, accepted: 24.06.2022

Noise and bias – some controversies raised by the book 'Noise: A Flaw in Human Judgment', written by Daniel Kahneman, Olivier Sibony, Cass R. Sunstein[1]

Mirosław Szreder[a]

Abstract. The paper reviews and discusses the statistical aspects of the phenomenon called 'noise' which Daniel Kahneman, the Nobel Prize winning psychologist, and his colleagues present in their new book entitled 'Noise: A Flaw in Human Judgment'. Noise is understood by the authors as an unexpected and undesirable variation present in people's judgments. The variability of judgments influences decisions which are made on the basis of those judgments and, consequently, may have a negative impact on the operations of various institutions. This is the main concern presented and analyzed in this book.

The objective of this paper is to look at the relationship between bias and noise – the two major components of the mean squared error (MSE) – from a different perspective which is absent in the book. Although the author agrees that each of the two components contributes equally to MSE, he claims that in some circumstances a reduction of noise can make accurate inference not less, but more difficult. It is justified that the actual impact of noise cannot be accurately determined without considering both bias and noise simultaneously.

Keywords: noise, bias, mean squared error, statistical inference

JEL: C12, C13, C18, D80

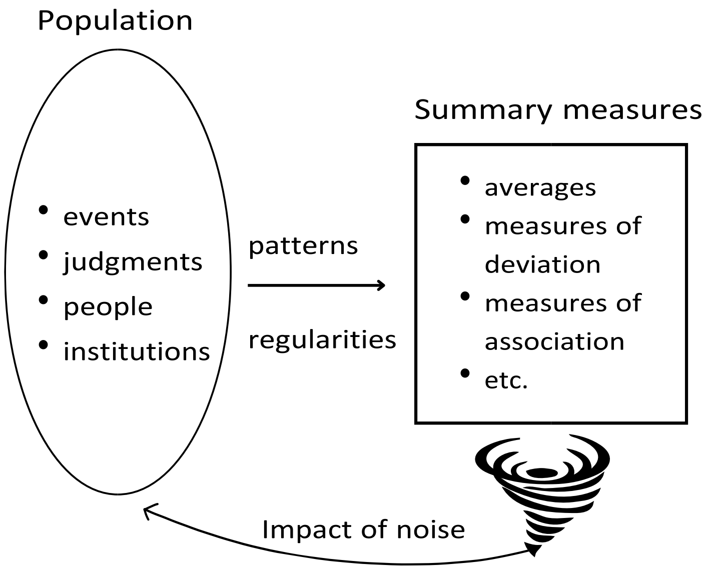
1. Introduction

Events, estimates or judgments which occur repetitively in large numbers are viewed by statisticians in the way that in simplified form is presented by Figure 1. They look at the set of population units (or sample units which represent the parent population) and try to identify patterns and regularities which will be described by summary measures (numbers) and their proper interpretation. Quantitative description of patterns derived from observed events or people's judgments is the main goal of statistics. Daniel Kahneman, Olivier Sibony i Cass R. Sunstein, authors of a new book related to flaws in human judgments (Kahneman et al., 2021) have decided to look further or – to be more precise – to seek possible implications of given statistical characteristics of the population on activities of institutions and behavior of people who constitute this population. They are particularly interested in consequences of variability among people's judgments, described by measures of deviation, which may be regarded as problems and challenges in certain institutions.

[1] Kahneman et al. (2021).

[a] University of Gdańsk, Faculty of Management, ul. Armii Krajowej 101, 81-824 Sopot, e-mail: miroslaw.szreder@ug.edu.pl, ORCID: https://orcid.org/0000-0002-7597-0816.

Figure 1. Noise in statistical thinking



Source: author's concept.

Daniel Kahneman, the Nobel Prize winning psychologist, a prominent and widely respected researcher of the human mind, author of the bestselling book 'Thinking, fast and slow' (Kahneman, 2011) has explained in a number of scientific papers typical cognitive errors which many people share, and errors which they make in judgments and decision-making. These are errors which can be observed both in professional and personal lives of people. One of the consequences of those cognitive errors is variation present in people's judgments which refer to the same event. Although a certain level of variation may not be disturbing to many people, it is likely to be more challenging if it increases. As the main reasons of variation among people's judgments have already been discovered and explained (see e.g., Kahneman, 2011; Kahneman et al., 1982; Morvan & Jenkins 2017), it seems reasonable to ask the question of the impact of variability of judgments on the activities of various institutions, in particular those in which consistency of judgments is expected and desirable. This question is raised in the presented book and investigated from different perspectives. The authors look at how decisions based on diverse judgments affect people's lives and reputation of institutions. Additionally, they suggest some ways which allow reducing the present level of deviation observed in judgments.

The title 'noise' is understood by the authors as unexpected and undesirable variation present in people's judgments, and they add that 'there is too much of it' (p. 361). Explaining motives and reasons for studying this problem the authors write:

'The surprises that motivated this book are the sheer magnitude of system noise and the amount of damage that it does' (p. 365). From the very beginning of this book a reader will find examples of negative consequences of noise – variability of judgments and estimates among professionals. Convincing examples relate to: divergent judgments of doctors in diagnosis of illness, variation in judicial judgments related to the same case, and also deviation in administrative decisions (e.g. related to asylum status), economic forecasts, decisions of patent offices, actuarial estimates of insurance premiums.

On the one hand, it is quite natural that people in their personal judgments are different. The basis for their judgments is formed not only by professional knowledge, whose range and quality may be different among individuals, but also by their overall experience, and usually unique way of combining and processing information, or in other words, unique way of thinking. However, judgment is a category which should be regarded as narrower than thinking. According to the authors: 'Judgment is a form of measurement in which the instrument is a human mind' (p. 361). The measure applied does not have to use numbers, it may employ other scales. If the dispersion of judgments is not large, it does not attract public attention, even if the differences occur among professionals. It could be argued that we all are used to the presence of certain level of noise in judgments related to political issues, economic or environmental events or processes. It is not surprising that they differ one from another, as long as the differences are not too large. Also, in academic communities there seems to be common acceptance for a certain level of diversity among judgments concerning students' achievements or research outputs of scientists.

On the other hand, it would be much more difficult to obtain acceptance for diversity (noise) in judgments and decisions that follow them, if they seriously affect people's lives or future careers. Noise tends to be considered unwanted in institutions representing judicial system, health care system, vocational advice system, and others. If doctors present a wide range of judgments related to the choice of treatment in a particular case, or judges announce vastly different views on the severity of punishment, it will inevitably lead to confusion and disorientation. Additionally, it may undermine the competences of those professionals and confidence in their expertise. These are the main reasons, why we should seek information about the sources and nature of noise, as well as accessible ways of monitoring and reducing noise. This is what the book is about.

2. Bias and noise in judgments

The initial claim presented by the authors of this book indicates that noise, as one of two components of the total error which accompany every judgment, attracts less attention than the other one – bias. Noise tends to be overlooked and neglected. 'This book is our attempt to redress the balance' – declare the authors (p. 6). Using statistical terms, this view could be expressed as follows: systematic errors and bias in estimates attract more interest than an equally or more crucial factor of inaccuracy – noise. This claim, however, seems disputable. No doubt that a high level of noise in judgments of doctors, judges and many decision-makers in public administration is harmful, and sometimes has painful consequences for people. But on the other hand, there are several reasons for which bias, not noise, ought to remain in the center of our concern. Two of the reasons are explained below, and the final one later, together with measures of errors.

Firstly, bias is a systematic error or tendency toward a distorted judgment, and may manifest itself in dangerous social phenomena, like various kinds of inequalities or discriminations. Bias is responsible for false judgments which form basis for racial, religious, gender or wage discrimination. All these kinds of prejudice and discrimination are not caused by noise (variation of judgments). It is an irreducible constant bias present in judgments of groups of people and representatives of institutions, and their decisions based on those judgments, that accounts for these phenomena. Alleged racial bias in police activities in some countries, gender gaps and biased (prior) assessments of productivity by age or sex in labor market are examples of such discrimination.

Secondly, unlike noise, which can be reduced by increasing the number of independent judgments or averaging them, bias does not exhibit that or any other similar property. One cannot reduce bias by simply increasing the number of judgments collected. Also, authors of the presented book confirm that bias is not a decreasing function of the number of judgments or evaluations. Statistical methods and techniques are less helpful in reducing systematic errors than noise. If a bias occurs in judgments, one of the most efficient ways to deal with it is to incorporate other relevant sources of information, which in practice may be difficult. To reduce noise seems to be an easier task. Readers may of course not share this view. And it seems that the authors of this excellent book do not, either.

3. In-depth analysis of noise

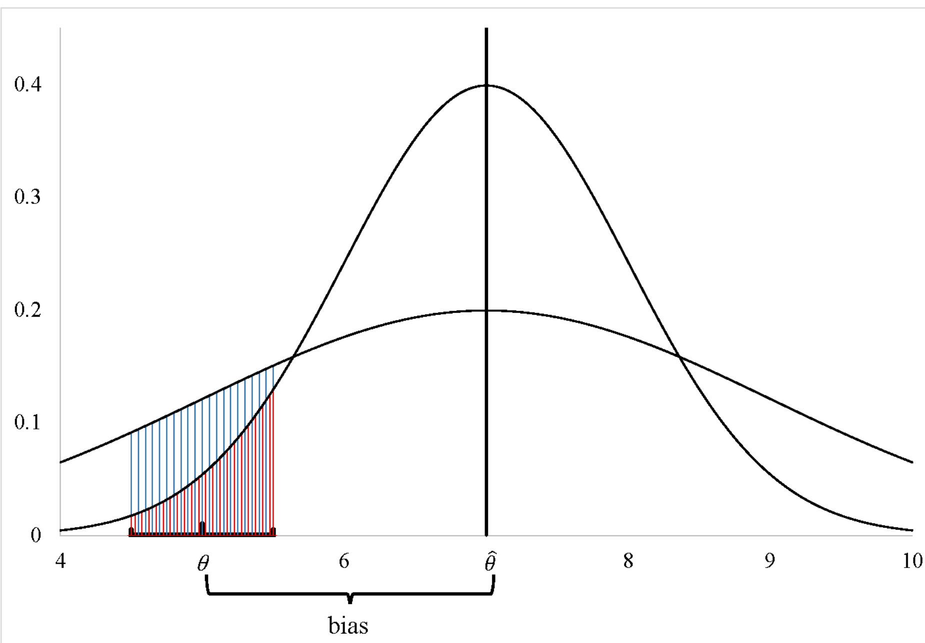
The content of the book is split into six parts, each consisting of three to eight chapters. Every chapter is accompanied by a short recapitulation and conclusions. The total number of chapters is 28. The last one is followed by 'Review and Conclusions: Taking Noise Seriously'. The final part of the book, called 'Epilogue' is entitled: 'A Less Noisy World'. It includes three appendices in which practical rules and procedures designed for dealing with noise, including audit of noise, are proposed and discussed. One of the advantages of the book worth pointing out is clear and precise language, and also a large number of examples which enable one to understand interesting and original considerations involved in all its parts.

The first part of the book which consists of three chapters begins with a number of persuasive examples of undesirable variations which can be found in judges' decisions in courts and in estimates and judgments of underwriters employed by insurance companies. Such variations may evoke a sense of injustice among people and additionally incur financial losses, sometimes of considerable volumes. In this part of the book the notion of 'system noise' appears for the first time. It is defined as undesirable variation existing among judgments of different people assessing the same case. In further parts of the book system noise is divided into 'level noise' and 'pattern noise'. An interesting point is presented by the authors in relation to noise which can exist, although it tends to be overlooked, in singular events or unique decisions. It is proposed that a unique decision should be regarded as a potentially recurrent decision, even if it is taken only once. The decision-maker should follow the same rules aimed at reducing bias and noise, which are applied in the case of repetitive events. In statistics one will find more analogies to this kind of logical approach.

One of the first stages of the approach designed to reduce undesirable variability of judgments involves measurement of variation or measurement of noise. 'Your Mind Is a Measuring Instrument' is the title of the second part of the book. Evaluation of judgments in order to improve ways of making them does not seem to be easy. Especially, if the judgments cannot be verified with regard to their accuracy and precision, for example in hypothetical scenarios, or long-term forecasts. Therefore, the authors propose to look both at the accuracy of judgments, when it is possible, and simultaneously at the process of formulating judgments. In other words, it is recommended to compare *ex post* judgments with actual outcomes, if possible, and additionally to assess the quality of the process of making judgments.

The problem of measuring two principal components of the total error: bias and noise is extensively discussed in Chapter 5. Recalling the well-known formula in statistics for mean squared error (MSE) which can be expressed as the sum of squared bias and squared standard deviation, the authors claim that each of the two components contributes equally to MSE. They call this formula 'the error equation' and recognize it as 'the intellectual foundation of this book' (p. 66). The authors emphasize that a given change (increase or decrease) in bias or in noise has the same impact on MSE. They write: 'Reducing noise or reducing bias by the same amount has the same effect on MSE' (p. 65). This statement can be regarded disputable or controversial. The influence of bias and noise on MSE are described correctly, however the problem is not unambiguous, if other measures of quality of judgments are taken into account. It would be reasonable, for instance, to consider consequences of reducing noise in the context of how frequently less dispersed judgments will be equal to the true value or fall close to it. In other words, it may be important to look at how the probability that the next judgment will be close to the true parameter changes, if the noise is reduced, given that a certain level of bias is involved in all judgments. Figure 2 presents two distributions of judgments which have the same bias (the difference between the mean of judgments – $\hat{\theta}$, and the true value – θ) but different values of noise (different standard deviations). It is clear that smaller dispersion, although desired in other circumstances, results in smaller probabilities of obtaining judgments which fall to the unit interval around true parameter θ. If one could reduce noise even further, the corresponding probability would get smaller and smaller. This means that the probability of obtaining judgments close to the true value will under this assumption approach zero. This is the real consequence of bias.

Figure 2. Bias and noise in the distributions of judgments vs. probability of getting judgments close to the true value θ



Source: author's idea and calculations.

A reduction of noise, if bias is present in judgments, makes it less likely that the true value is going to be discovered. This is because biased judgments will absorb increasingly large amount of probability. And ultimately, probabilities around the true value will be smaller compared to the distribution that has larger dispersion (larger noise). Whether or not it happens in a particular case depends mainly on the mutual relation between the size of bias and the size of noise. They need to be considered simultaneously. Otherwise, a reduction of noise may have positive or negative impact on the quality of judgments. Although the authors are aware of this problem, they do not explain it in the context suggested above. They confine themselves to claiming that 'Reducing noise would be less of a priority if bias were much larger than noise' (p. 66). It is difficult to agree that this is just a question of priorities. It is a problem of benefits or lack of benefits that one could expect, given different quantities of bias present in judgments. Of course, benefits are expressed in terms of probabilities of getting judgments close to the true value assessed. Emphasizing (more than once in this book) that bias and noise contribute equally to MSE, which is true, and paying little attention to various consequences of mutual relation between the two components, may be misleading.

All the other chapters of the second part of the book present discussion of the factors which generate noise and further classifications of noise. Psychological background which can often be identified behind a given kind of noise is widely and interestingly discussed. Perhaps most commonly 'occasion noise' is observed by people in their everyday social activities. However, the authors argue that this is not the most important source of system noise. Regardless of the factors which influence noise in individual assessments and judgments, it should not be assumed that an efficient way to reduce noise is always a group discussion which searches consensus. The discussion and its output may be affected, the authors prove, by a kind of social influence. Group polarization can be recognized to be a special case of such influence and simultaneously a source of noise. After a thorough discussion of these issues and the results accompanying the experiments, the authors conclude that groups of discussants looking for a consensus need to be properly managed in order to avoid high level of noise. Lack of management may cause that noise present in some individual judgments could be amplified.

The third part of the book covers various issues of noise which can arise in predictive judgments. The initial assumption stated and justified by the authors is that in forecasting human abilities are inferior to statistical models, including correlation and regression models and additionally artificial intelligence (AI). They indicate that one of the main factors which accounts for this is noise which commonly affects people's judgments. Even simple rules may be in these circumstances superior to human judgments. 'The combination of personal patterns and occasion noise weighs so heavily on the quality of human judgment that simplicity and noiselessness are sizable advantages' (p. 133). Having access to the same information, models and algorithms tend to be more efficient and more accurate than humans. In further chapters a reader will find descriptions of rules (algorithms) free from noise.

The authors prefer using the term ‘rules’, which in their interpretation has broader meaning than models and algorithms. Unlike people, rules are not overconfident, which is one of the reasons why they are generally more reliable in prediction. Moreover, some people tend to deny their own lack of knowledge: ‘The denial of ignorance is all the more tempting when ignorance is vast’ (p. 145). Assuming that a causal mechanism of a phenomenon of interest has been discovered, it remains difficult to predict accurately its future development. People tend to neglect not only uncertainty but also noise. ‘The future seems as predictable as the past. And noise is neither heard nor seen’ (p. 158) write the authors in a slightly metaphorical mode at the end of the third part of the book.

The content of the fourth part is well reflected by its title: ‘How Noise Happens’. The authors focus on psychological aspects of noise in human judgments. This is an area of science in which Professor Daniel Kahneman is an internationally renowned and respected expert. Readers of his previous book (Kahneman, 2011) will find once again a clear presentation and examples of some common cognitive errors and biases which people tend to make. They are in particular: heuristics of substitution, conclusion biases, and excessive coherence (i.e. forming coherent views quickly and slowly changing them). Each of them can generate noise and additionally bias. The analysis of sources of noise is complemented by discussion of the statistical and psychological aspects of the formulation of judgments. Concentrating on statistical issues the authors stress how important it is to adopt proper scales in order to help avoid noise in predictive judgments. Some psychological circumstances are discussed even more extensively. Special attention is given to sources of ‘what may be the most intriguing type of noise: the patterns of responses that different people have to different cases’ (p. 159).

Pattern noise, defined by the authors as ‘an error in an individual’s judgment of a case that cannot be explained by the sum of the separate effects of the case and the judge’ (p. 203) is studied in this book thoroughly. It is disaggregated into two components – occasion noise and stable pattern noise. Both are explained with regard to their sources and specific features. In the last Chapter of the fourth part of the book one will find a list of all components of noise presented earlier and ways of measuring each of them. However, one thing may be found slightly unclear in this chapter. While graphical interpretation of MSE in Figure 16 raises no ambiguities, as it refers directly to the analytical formula for MSE (the sum of squared bias and standard deviation), the decomposition of stable pattern noise depicted in Figure 15 does not seem clear. This applies also to the following formula which precedes the graph:

$$(\textit{Pattern Noise})^2 = (\textit{Stable Pattern Noise})^2 + (\textit{Occasion Noise})^2.$$

Squaring all elements in this equation may be perceived as not satisfactorily explained. It is intuitively meaningful; however, any analytical justification of this relation may not be straightforward.

Organizations which face the problem of actual or potential consequences of noise in their activities may want to take measures in order to improve judgments which account for the noise. How to do this is outlined in detail in the fifth part of the book. A part of this presentation is an original procedure of the audit of noise (Appendix A) and a checklist for the decision observer – a person who is in charge of searching for symptoms of cognitive errors in organizations (Appendix B). Sometimes, undesirable variability in judgments (noise) may be the result of lack of knowledge or insufficient expertise of employees. However, very often it is not just one reason but a combination of different reasons, which require some specially designed strategies to reduce the noise. They include for instance, aggregation of independent judgments. Such strategies may be effective in reducing particular cognitive errors, and consequently also noise, but will not be very useful in establishing which of the errors account mostly for the observed noise. 'Noise is an invisible enemy' – say the authors (p. 244).

The elimination of cognitive errors is not easy. They tend to be overlooked by a person who commits them, in spite of his/her ability to recognize them in others. To illustrate this phenomenon (blind spot) the authors refer to the survey of 400 professional forensic scientists from 21 countries, in which 71% agreed that cognitive bias is a cause for concern in the forensic sciences as a whole, but only 26% said that their own judgments are influenced by cognitive errors.

Criminology is one of several areas for which the authors propose strategies and methods for improving judgments and reducing noise. Other areas include: anticipation, judgment formulation and decision-making in healthcare, assessment of staff performance ('rank but not force', p. 294), structuring complex judgments in processes of recruitments. A more developed approach to the issue of recruitment is a procedure which the authors have called MAP – Mediating Assessments Protocol. Details of the protocol and examples of its implementation based both on the authors experiences and their research are discussed in the last Chapter of the fifth part of the book (Chapter 25).

4. Do we need certain noise in judgments?

The content of the fifth part but also other chapters of the book may suggest that perhaps the authors seek ways not only to reduce but to eliminate noise in all institutions which face its consequences. It could eventually be achieved by employing artificial intelligence (AI) or algorithms which would produce judgments for making decisions in judicial institutions, hospitals, centers of vocational counselling and others. Rationality of neural networks and various other representatives of AI may be perceived as a remedy for human's wavering and lack of certainty. It should be stressed, however, that the authors are aware of major risks which such a substitution would probably generate. Some of them have been identified and well described by O'Neil (2017). And the risks are not confined to the possible deepening of discrimination of various backgrounds, which is convincingly demonstrated by O'Neil (2017). The risks seem to be of more fundamental nature – they consist in profound standardization of all those characteristics of units of interest (people, events) that are unknown to algorithms. For example, many of us would presumably feel awkward, if a doctor's diagnosis were based solely on algorithms, excluding medical experience and conclusions of the patient interview. Similarly, professional judges and court jurors are commonly expected to take into account not only the output of algorithms but also other personal or social circumstances of the defendant. In other words, like the authors of this book, people are ready to accept certain level of noise in judgments of professionals.

What is the accepted level of noise and whether optimal noise exists is explored in the final part of the book. One will find there several objections to efforts aimed at reducing or eliminating noise. The major ones are: prohibitive costs or even lack of feasibility of such efforts, the risk that the reduction of noise can introduce other errors, the difficulty of adopting new values in a society where there is no room for noise. Moreover, some noise-reduction strategies might squelch people's creativity, point out the authors. The last risk is well exemplified by algorithms which in principle tend to replicate patterns from the past and do not indicate the need for change. All these objections and arguments are interestingly discussed in this part of the book. The discussion leaves room for the reader's own reflections and views.

Apart from 'Review and Conclusion' (18 pages) the authors have decided to include the main findings and key messages of the book in a separate one-page long part called 'Epilogue: A Less Noisy World'. They argue that the world less affected by noise would bring large savings of money, improve public safety and health, increase fairness, and prevent many errors. In the process of transition to such a world they see a role for AI to play.

5. Summary and conclusions

The book constitutes a broad view on human's judgments affected by noise analyzed from various perspectives. It is also a competent and interesting discussion of the human weaknesses that underlie the mistakes we make when forming judgments. The authors did their best to be objective in evaluating the consequences of noise for institutions and society, as well as in discussing the reasons for reducing or eliminating noise. Conclusions presented in this book are particularly important nowadays, when AI offers increasing support in decision-making and gradually replaces human judgments with its own assessments.

The only thing that may be considered unclear and ambiguous in the book is the aforementioned issue of the relationship between bias and noise. Particularly, in the context of frequencies of wrong decisions based on judgments affected by both these errors. The actual consequences of noise cannot be accurately determined without considering both bias and noise simultaneously. Noise reduction, despite its positive effect on MSE values, should not always be considered beneficial. If we all say the same, and it will turn out that we were all wrong, some may correctly conclude that if more dispersed views were available, they would suggest that the truth might be different.

References

Morvan, C., & Jenkins, W. J. (2017). *An Analysis of Amos Tversky and Daniel Kahneman's Judgment under Uncertainty. Heuristics and Biases*. Macat Library.

O'Neil, C. (2017). *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*. Penguin Books.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D., Sibony, O., & Sunstein C. R. (2021). *Noise. A Flaw in Human Judgment*. William Collins.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809477.

Przegląd Statystyczny. Statistical Review, 2022, vol. 69, 1, 50–52
Received: 19.04.2022, revised: 11.05.2022, accepted: 16.05.2022

The 39th Multivariate Statistical Analysis MSA 2021. Conference Report

Marta Małecka,[a] Artur Mikulec[b]

The 39th Multivariate Statistical Analysis MSA 2021 international scientific conference was held on November 8–10, 2021 in Łódź, at the Faculty of Economics and Sociology of the University of Łódź. The conference was organised by the Department of Statistical Methods of the University of Łódź in collaboration with the Institute of Statistics and Demography of the University of Łódź, the Committee of Statistics and Econometrics of the Polish Academy of Sciences and the Łódź branch of the Polish Statistical Association. The honorary patronage over the conference was taken by Elżbieta Żądzińska, Rector of the University of Łódź, and Dominik Rozkrut, President of Statistics Poland.

The conference was organised in co-operation with the MASEP (Measurement and Assessment of Social and Economic Phenomena) conference, arranged by the Department of Economic and Social Statistics of the University of Łódź. The conference received financial support from the Ministry of Education and Science (MEiN) as part of the 'Excellent Science' programme (DNK/SP/515427/2021). Educational activities related to the conference were supported by the National Bank of Poland (NBP) as part of its educational project (NBP-DEW-WPE-AB-0662-0226-2021) which corresponds to NBP's priority areas of economic education – 'New horizons of economic thought'. StatSoft Polska Sp. z o.o. was also the content partner of the conference. Prof. Czesław Domański was the chairman of the Scientific Committee, while the Organising Committee was chaired by Alina Jędrzejczak, Assoc. Prof. of the University of Łódź.

The main goal of the MSA 2021 conference was to provide an international forum for the discussion and exchange of ideas and views on the development of statistics as a science. The specific objectives included:

- the presentation of the latest achievements in the field of multidimensional statistical analysis;
- disseminating knowledge in the field of data analysis and the application of statistical methods in other scientific disciplines, especially in economics, sociology and finance, as well as the exchange of experiences;
- creating a bridge between science (statistics) and research practice (individual users, business and administration).

[a] University of Łódź, Department of Statistical Methods, ul. Rewolucji 1905 r. nr 41, 90-214 Łódź, e-mail: marta.malecka@uni.lodz.pl, ORCID: https://orcid.org/0000-0003-4465-9811.
[b] University of Łódź, Department of Statistical Methods, ul. Rewolucji 1905 r. nr 41, 90-214 Łódź, e-mail: artur.mikulec@uni.lodz.pl, ORCID: https://orcid.org/0000-0001-8249-2296.

The MSA 2021 conference was held in hybrid form: in-person and online. 84 participants from various academic centres in Poland attended the event, including representatives from Gdańsk, Katowice, Kraków, Lublin, Łódź, Poznań, Radom, Szczecin, Warsaw and Wrocław, as well as representatives of Statistics Poland and the Statistical Office in Łódź. The conference also hosted members of the academic society from abroad: the Czech Republic, India, Iran and Italy. The conference consisted of 15 sessions (plenary and parallel), with 63 papers presented.

The conference was opened by the Chairwoman of the Organising Committee, Alina Jędrzejczak. On behalf of Elżbieta Żądzińska, Rector of the University of Łódź, the conference participants were welcomed by Agnieszka Kurczewska, Vice-Rector for External Relations. Subsequently, short welcome speeches were given by Ewa Kusideł, Vice-Dean for Science (Faculty of Economics and Sociology, University of Łódź) and the Chairman of the Scientific Committee, Czesław Domański (University of Łódź).

According to the tradition of the MSA conference, the first plenary session was devoted to prominent representatives of the historical statistical thought and to the memory of statisticians who have recently passed away. This session was chaired by Bronisław Ceranka (Poznań University of Life Sciences). The first lecture, devoted to the life and scientific work of Antoni Łomnicki (1881–1941), a probabilist and statistician, was given by Mirosław Krzyśko (Adam Mickiewicz University in Poznań). The next speaker was Czesław Domański (University of Łódź), who presented the images of Józef Kleczyński (1841–1900), the precursor of population estimation between censuses, and Kazimierz Władysław Kumaniecki (1880–1941), the initiator of the Polish Statistical Society and the first Statistical Yearbook – Polish Statistics. In this session, the profiles of three famous Polish statisticians who passed away last year were recalled:

- ‘Dominik Szynal – creator of the probabilistic environment in Lublin’ – by Mariusz Bieniek (Maria Curie-Skłodowska University);
- ‘Daniel Kosiorowski – an outstanding Krakow statistician’ – by Józef Pociecha (University of Economics in Krakow);
- ‘Ryszard Walkowiak – statistician and naturalist’ – by Małgorzata Graczyk (University of Life Sciences in Poznań).

During the conference, there were four open lectures delivered by the invited speakers:

- ‘Harnessing the power of Earth observation for official statistics’ – Dominik Rozkrut (Statistics Poland);
- ‘About the sampling plans depending on the position statistics of the auxiliary variable’ – Janusz Wywiał (University of Economics in Katowice);
- ‘The appearance of the Rawlsian Paradox when neglecting income dependence of the random equivalence scales’ – Stanisław Maciej Kot (Gdańsk University of Technology);

- 'Graphical and computational tools to guide parameter choice for robust clusterwise regression' – Francesca Greselin (University of Milan).

A broad range of topics related to the theory and application of mathematical statistics was covered during parallel sessions which focused in particular on the following groups of issues:

1. The theory of statistical methods. The papers presented at the conference related to both estimation and statistical inference. Studies from the area of taxonomic issues were also presented. Outliers, fuzzy numbers, Big Data, bootstrapping techniques and text recognition were among the variety of the discussed topics.
2. Macroeconomic applications. This thematic group included issues related to macroeconomic interventions, inflation and the use of modern data collection methods such as scanner data on web-scraped data.
3. Demographic and social issues. A large group of papers concerned the labour market. There were debates on many social issues relating to disabled or older people, retirement benefits and the quantification of poverty. In the area of demography, topics such as birth dynamics and cities demography were discussed. A number of papers dealt with social issues related to the COVID-19 pandemic.
4. Sustainable development. Among social topics, a special place was taken by discussions on sustainable development, relating in particular to the impact of economic activity on the environment, air quality, water demand depending on the weather conditions or the transformation of cities.
5. Business applications. The conference discussions covered the following areas within business applications: energy consumption forecasts, logistics, duration of business entities, micro-enterprise statistics, investment potential of voivodships, industrial transformation, identification of bid rigging, organisational security culture.
6. Financial market. A separate group of topics within statistical applications was the use of statistical methods in financial market analysis. The presented papers dealt with issues such as: the relationship between the COVID-19 pandemic and exchange rates, the cryptocurrency market, banking scoring models, financial efficiency of insurance companies and the modelling of systemic risk in the insurance sector.

A detailed list of topics is available at https://sites.google.com/view/msa2021pl/archiwum/msa-2021.

The next Multivariate Statistical Analysis MSA conference will be held at the University of Łódź on November 7–9, 2022.

Thanks to our Reviewers

The Editorial Team of *Przegląd Statystyczny. Statistical Review* would like to thank the following Reviewers of articles submitted for publication in 2021:

Marcin Anholcer
Tomasz Brodzicki
Marek A. Dąbrowski
Joanna Górka
Teresa Kamińska
Roman Kiedrowski
Dariusz Klimek
Tomasz Kuszewski
Ewa Lechman
Blanka Łęt
Grzegorz Mentel
Witold Orzeszko
Agnieszka Przybylska-Mazur
Ewa Roszkowska
Honorata Sosnowska
Józef Stawicki
Paweł Strzelecki
Jadwiga Suchecka
Ewa Syczewska
Elżbieta Szulc
Małgorzata Tarczyńska-Łuniewska
Krzysztof S. Targiel
Andrzej Torój
Piotr Wdowiński
Janusz L. Wywiał