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

Przegląd Statystyczny
Statistical Review

Vol. 67 | No. 2 | 2020

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 sciences 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. 67 No. 2 2020

ADRES REDAKCJI / EDITORIAL OFFICE'S ADDRESS

Uniwersytet Gdański, ul. Armii Krajowej 101, 81-824 Sopot

Redakcja językowa / Language editing: Wydział Czasopism Naukowych, Główny Urząd Statystyczny
Redakcja techniczna, skład i łamanie: Zakład Wydawnictw Statystycznych – zespół pod kierunkiem Wojciecha Szuchty
Technical editing and typesetting: Statistical Publishing Establishment – team supervised by Wojciech Szuchta

Zakład Wydawnictw Statystycznych

Druk i oprawa / Printed and bound:
Zakład Wydawnictw Statystycznych / Statistical Publishing Establishment
al. Niepodległości 208, 00-925 Warszawa, zws.stat.gov.pl

Strona internetowa / Website: ps.stat.gov.pl

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

Informacje w sprawie nabywania czasopism / Information on purchasing of the journal:
Zakład Wydawnictw Statystycznych / Statistical Publishing Establishment
tel./phone +48 22 608 32 10, +48 22 608 38 10

Zam. 418/2020 – nakład 220 egz.

CONTENTS

Przegląd Statystyczny. Statistical Review, 2020, vol. 67, 2, 97–113 DOI: 10.5604/01.3001.0014.5725

Modelling cyclical variation in the cost pass-through: a regime-dependent approach

Karolina Konopczak[a]

Abstract. In this study a regime-dependent ARDL model is developed in order to investigate how labour costs feed through into prices conditional on the business cycle position. Its estimates enable inference on the cyclical behaviour of markups. The proposed methodology is applied to the Polish industrial sectors. The obtained estimates point to procyclicality as the prevailing pattern of markup adjustment. Thus, overall markups in the Polish industry seem to have a mitigating effect on business cycle fluctuations. The degree of procyclicality seems, however, to be positively correlated with the degree of the industry's competitiveness.

Keywords: non-linear cointegration, regime-dependence, cost pass-through, markup cyclicality
JEL: Classification: C22, E31, E32

1. Introduction

Wage rigidity is commonly thought to be the cause of unemployment in the wake of adverse shocks, thus increasing the depth and prolonging the duration of a downturn. Following the same line of thought, wage flexibility is often perceived as an absorption mechanism, with wage concessions in economic slack hypothesised to facilitate job protection, boost international competitiveness (and exports) and, consequently, contribute to the containment of negative shocks. This belief, widely held in policy-making circles, hinges upon a classical assumption of the interchangeability between price and quantity adjustments of labour force, with either wages or employment bearing the brunt of the shock. However, as argued in recent literature (see Gali, 2013 and Galí & Monacelli, 2016), wage concessions affect labour demand and, hence, employment, only if they affect prices and induce monetary policy response in the form of interest rate cuts, thus stimulating the demand for goods. The effectiveness of downward wage adjustments in containing adverse shocks is, as demonstrated, conditional upon the degree of price rigidity. In particular, if falling wages do not reduce prices, wage flexibility may have little or no effect on the output and, consequently, employment outcomes. In such circumstances wage decreases may spur contractionary effects. It is then the interrelation between the wage- and price-flexibility that is central to the mechanism of business cycle propagation, rather than the wage flexibility alone. If prices are set up as a markup over marginal costs, it is the cyclical behaviour of the markup that determines the shock-absorption capacity of wage adjustments.

[a] Warsaw School of Economics, 162 Niepodleglosci Av., 02-554 Warsaw, Poland,
e-mail: karolina.konopczak@sgh.waw.pl, ORCID: https://orcid.org/0000-0002-6677-5269.

Empirical evidence on markup cyclicality is abundant, yet notoriously unrobust. Extracting the markup series is one of the most challenging empirical issues in macroeconomics (Nekarda & Ramey, 2013). Theoretically, markups can be derived by comparing prices and marginal costs. The latter, however, are not observable, leading to a number of approximations having been proposed in the literature, e.g. taking account of the evolution of the Solow residual (Hall, 1986, 1988; Roeger, 1995), the labour share (Bils, 1987), inventories (Bils & Kahn, 2000), advertising spending (Hall, 2012) or through adjusting average costs series (Galí et al., 2007; Martins & Scarpetta, 2002; Rotemberg & Woodford, 1991, 1999). The results obtained for the U.S. industrial sectors using the above-mentiond techniques are suggestive of the pro- (e.g. Chirinko & Fazzari, 1994; Domowitz et al., 1986, 1988; Hall, 2012; Nekarda & Ramey, 2013) and counter-cyclicality (e.g. Bils, 1987; Bils & Kahn, 2000; Martins & Scarpetta, 2002; Rotemberg & Woodford, 1999) of markups.

Since the conclusions on the markup behaviour depend heavily on the estimation method, in this study we bypass the estimation of markups and instead propose to investigate how labour costs feed through into prices conditional on the business cycle position. For this purpose we develop a regime-dependent ARDL model of cost pass-through, extending the asymmetric ARDL model by Shin et al. (2014). The proposed methodology does not allow for the derivation of markup series but instead enables the capture of the interrelation between wage and price adjustments over the business cycle, i.e. the degree of pass-through. Nonetheless, a large body of literature (i.a. Atkeson & Burstein, 2008; Goldberg & Hellerstein, 2013; Gopinath et al., 2010; Hellerstein, 2008; Nakamura, 2008; Nakamura & Zerom, 2010) identifies time-varying markups as one of the most important determinants of the pass-through variation.[1] Thus, the estimation results allow us to assess whether markup behaviour has a mitigating or amplifying effect on business cycle fluctuations. On this basis, conclusions can be drawn on whether wage flexibility and moderation constitute an appropriate policy prescription for the economic stabilisation. The Polish industry serves as an application example.

The paper is organised in the following way: Section 2 gives a theoretical background, Section 3 outlines the methodology employed in the study and discusses the empirical strategy, i.e. our approach to investigating business cycle dependence in the cost pass-through, and Section 4 presents the empirical results. The last section summarises our findings.

[1] It should be borne in mind that when comparing the trajectories of labour costs and prices, we do not control for other costs, in particular the cost of intermediate inputs and capital. Therefore, precisely speaking, our conclusions pertain to 'wage markups'.

2. Theoretical notes

The behaviour of markups over the business cycle is an unresolved issue in theoretical economics. Depending on the underlying assumptions, theoretical models predict different outcomes regarding markup cyclicality. The Phelps & Winter model (1970) predicts procyclicality by assuming that when firms anticipate higher demand in the future, they lower prices in order to expand their consumer base. In the Green and Porter model (1984), firms cannot observe the reason behind falling market demand and, thus, misinterpret economic slack as other firms' cheating. It is, therefore, harder to sustain collusion in recessions, which leads to procyclical markups. In the model proposed by Rotemberg and Saloner (1986), the changing ability of firms to collude is also the main driver of cyclical variation in markups, but the assumption that the benefits of cheating are proportional to the current demand renders collusion harder to sustain in economic upturns than downturns. Thus, the model predicts countercyclicality of markups. Growing competition during economic booms is also the driving force behind procyclical markups in the Rotemberg and Woodford (1992) model. In Bils (1989), Klemperer (1995), Okun (1981) and Stiglitz (1984), markups are predicted to rise in recessions due to lower price elasticity of the demand and, thus, higher pricing power of firms. Additionally, Stiglitz (1984) suggests that by lowering the markup during economic booms, incumbent firms deter others from entering the market. In turn, Chevalier and Scharfstein (1996), Gilchrist et al. (2017), Gottfries (1991) and Greenwald et al. (1984) attribute countercyclicality of markups to capital market imperfections that constrain the ability of firms to obtain external financing, especially during recessions. The subsequent liquidity squeezes force firms to raise profit margins.

The explanation to this lack of robustness in theoretical perditions can be provided by the recent advances in the pass-through literature. As derived by Weyl and Fabinger (2013), a general formula for the cost-price pass-through (ρ), applicable to a wide range of market settings (perfect competition, monopoly, symmetric imperfect competition) takes the following form:

$$\rho = \frac{1}{1 + \frac{\varepsilon_D}{\varepsilon_S} - \frac{\theta}{\varepsilon_S} + \theta\varepsilon_\theta + \frac{\theta}{\varepsilon_{ms}}}, \tag{1}$$

where:
ε_D is the elasticity of demand,
ε_S is the elasticity of supply,

ε_{ms} is the elasticity of marginal consumer surplus, measuring the curvature of demand,
θ is a conduct parameter, ranging from 0 for perfect competition to 1 for monopoly (see Genesove & Mullin, 1998),
$\varepsilon_{\theta} = \frac{\partial \theta}{\partial q}\frac{q}{\theta}$ is the elasticity of the conduct parameter with respect to quantity (q).

The pass-through depends, therefore, on the shape of the demand and supply curves as well as on the intensity of competition. Under perfect competition $(\theta = 0)$ the pass-through rate hinges solely upon the relative slopes of demand and supply. *Ceteris paribus*, the steeper the demand curve (the less responsive the demand to changes in prices) or the flatter the supply curve (the more responsive the output to changes in prices), the higher the degree of pass-through. Under oligopolistic and monopolistic settings not only the slope, but also the curvature of the demand function plays a role. *Ceteris paribus*, the pass-through will be higher if the demand is log-convex (i.e. $\frac{1}{\varepsilon_{ms}} < 0$).

The role played by the intensity of competition in determining the pass-through rate is less straightforward, since it depends on the shape of the demand and supply functions. All else being equal, the pass-through increases with the intensity of competition, providing that the demand is log-concave and decreases in the case of log-convex demand. The impact of changing competitive conduct on firms' ability to pass through costs depends also upon the shape of the cost function. In the case of increasing returns to scale, growing intensity of competition provides cost-absorption, whereas under decreasing returns it amplifies the cost changes. Therefore, the degree of pass-through diminishes with growing competition in the case of downward sloping, while increases in the case of upward-sloping marginal costs function. Additionally, the pass-through may be dampened or amplified by the way the competitive conditions change in response to demand fluctuations (ε_{θ}). If higher demand leads to firm entry (i.e. strengthens competitive conduct), then the initial impact of cost hikes on prices becomes partially absorbed, ultimately resulting in a lower degree of pass-through.

Given the complex and interactive way the degree of pass-through depends on its determinants, its cyclical behaviour cannot be easily inferred from the cyclical properties of demand, supply and competition. For instance, it is well established in the literature (e.g. Clementi & Palazzo, 2016; Lee & Mukoyama, 2015; Tian, 2018) that the economic expansion, leading to increasing profit opportunities in relation to entry costs, renders firm entry procyclical. Combined with counter- or acyclical firm exit, this suggests more competitive conduct in economic upturns. However, the resulting pass-through dynamics is not straightforward. In industries facing log-concave demand (and/or upward-sloping costs) this translates into procyclicality of the pass-through, whereas for sectors experiencing log-convex demand (and/or downward-sloping costs) it leads to countercyclicality. The question of cyclicality of the pass-through (as well as the markup, being the key driver of the pass-through variation[2]) is, as demonstrated, industry-specific and, ultimately, empirical.

3. Empirical framework

3.1. Regime-dependence in the ARDL model

In order to capture cyclical variation in the cost pass-through, we develop a regime-dependent ARDL model. For this purpose, we utilize and expand the non-linear cointegration analysis proposed by Shin et al. (2014), building upon Pesaran et al. (2001) and Pesaran and Shin (1999). In the 2-dimensional case, the non-linear cointegration equation takes the following form:

$$x_t = \delta_0 + \delta_1^+ y_t^+ + \delta_1^- y_t^- + \varepsilon_t, \tag{2}$$

where y_t^+ and y_t^- are partial sums of changes in y_t, so that $y_t = y_0 + y_t^+ + y_t^-$. In Shin et al. (2014), the non-linearity takes the form of asymmetry with y_t decomposed into y_t^+ and y_t^- around the threshold value of Δy_t. The threshold can be exogenously imposed (often set at zero) or endogenously determined (e.g. *via* the grid search). In the case of a zero threshold, the relation becomes asymmetric with respect to the sign of changes in y_t, with parameter δ_1^+ capturing the long-run response of x_t to an increase in y_t, and δ_1^- the long-run response to a decrease.

In order to capture regime-dependence (in this case, the dependence on the business cycle position), we propose the extension to the Shin's et al. (2014) framework by making the decomposition in y_t conditional on the behaviour of a transition variable (z_t). In this approach, y_t is partitioned according to the threshold value of $\Delta z_t(\tau)$, with partial sums defined as $y_t^- = \sum_{i=1}^{T} \Delta y_i \mathbb{I}_{\{\Delta z_i \leq \tau\}}$ and $y_t^+ = \sum_{i=1}^{T} \Delta y_i \mathbb{I}_{\{\Delta z_i > \tau\}}$, where $\mathbb{I}_{\{.\}}$ is an indicator function taking the value of one if the condition in the bracket is met, and zero otherwise.

[2] The empirical literature on the pass-through determination is almost entirely devoted to the exchange rate pass-through, in the case of which usually the non-traded costs contribute the most to its variation, followed by markup adjustments. The role of nominal rigidities ('menu costs') is universally considered negligible. Therefore, it can be hypothesised that it is markup adjustments that are the driving force in the context of the wage pass-through.

Following Shin et al. (2014), the estimation of short- and long-run elasticities as well as testing for the existence of the cointegration relationship is performed within the non-linear ARDL model:

$$x_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i x_{t-i} + \sum_{i=0}^{q} (\beta_i^+ y_{t-i}^+ + \beta_i^- y_{t-i}^-) + \vartheta_t. \tag{3}$$

After reparametrisation, the model is estimated in the unrestricted error correction form:

$$\Delta x_t = \alpha_0 + \gamma x_{t-1} + \beta^+ y_{t-1}^+ + \beta^- y_{t-1}^- + \sum_{i=1}^{p-1} \alpha_i \, \Delta x_{t-i} + \sum_{i=0}^{q-1} (\beta_i^+ \Delta y_{t-i}^+ + \\ + \beta_i^- \Delta y_{t-i}^-) + \vartheta_t, \tag{4}$$

where $\gamma = -(1 - \sum_{i=1}^{p} \alpha_i)$, $\beta^+ = \sum_{i=0}^{q} \beta_i^+$ and $\beta^- = \sum_{i=0}^{q} \beta_i^-$.

The existence of a long-run relationship is established using the bounds-testing approach proposed by Pesaran and Shin (1999). It involves testing the null hypothesis of $\gamma = \beta_1^+ = \beta_1^- = 0$. The framework is applicable for both I(1) and I(0) regressors. Therefore, there are two asymptotic critical values: one under the assumption that all regressors are I(1), and the other assuming their stationarity. If the test statistics falls outside the critical value bounds, the null of no level relationship can be rejected. If it falls within the bounds, the inference is inconclusive. The relevant critical values are tabulated in Pesaran et al. (2001).

In order to recover the long-run parameters, the restricted error correction model can be derived as follows:

$$\Delta x_t = \alpha_0 + \gamma \left(x_{t-1} + \frac{\beta^+}{\gamma} y_{t-1}^+ + \frac{\beta^-}{\gamma} y_{t-1}^- \right) + \sum_{i=1}^{p-1} \alpha_i \, \Delta x_{t-i} + \\ + \sum_{i=0}^{q-1} (\beta_i^+ \Delta y_{t-i}^+ + \beta_i^- \Delta y_{t-i}^-) + \vartheta_t, \tag{5}$$

where $-\frac{\beta^+}{\gamma}$ and $-\frac{\beta^-}{\gamma}$ are the long-run elasticities, δ_1^+ and δ_1^- respectively, and γ is the error correction coefficient. The symmetry in the short-run ($\beta_i^+ = \beta_i^-$) and long-run ($\delta_1^+ = \delta_1^-$) responses can be tested by applying the Wald statistics. If, however, the threshold is estimated, the statistics follows a nonstandard asymptotic distribution (Davies, 1977). For this reason, the approximate critical values should be obtained by means of a bootstrap procedure proposed in Hansen (1996, 2000).

3.2. The data

The data on the Polish industry comes from Eurostat and Statistics Poland. Unit labour cost, price and demand series were obtained from the short-term business statistics (STS) database (Eurostat). The sample covers the years 2000 through 2016 and is of quarterly frequency. The data is both seasonally- and calendar-adjusted. The Herfindahl-Hirschman index, as a measure of the industry's degree of concentration, comes from Statistics Poland (Statistical Yearbook of Industry).

Unit labour costs are defined as productivity-adjusted wages and the demand faced by the industry is proxied by its turnover (for the definition of variables see Table 1).

Table 1. Definition of variables[a]

Variable	Symbol	Definition
prices	p_t	producer price index (PPI)
unit labour costs	ulc_t	gross wages and salaries over PPI-deflated output
demand	$demand_t$	volume of sales (i.e. total turnover in industry deflated by PPI)

a All variables are in natural logarithms.
Source: Eurostat.

The sectoral coverage includes NACE rev. 2 sections B (*mining and quarrying*), C (*manufacturing*), D (*electricity, gas, steam and air conditioning*) and E (*water supply, sewerage, waste management*), i.e. the industry. The manufacturing section consists of 23 divisions (see Table 2 for basic characteristics of the sectors).

Table 2. Sectoral characteristics[a]

Sectoral classification	NACE code	Production (% of total industry)	Employment (% of total industry)	Herfindahl-Hirschman index
Manufacture of:				
food	C10	14.4	13.6	0.004
beverages	C11	2.2	0.9	0.062
tobacco	C12	0.8	0.2	0.228
textiles	C13	0.9	1.8	0.036

a Data come from Eurostat and Statistics Poland and cover the year 2015.

Table 2. Sectoral characteristics[a] (cont.)

Sectoral classification	NACE code	Production (% of total industry)	Employment (% of total industry)	Herfindahl--Hirschman index
Manufacture of (cont.):				
wearing apparel	C14	0.6	3.1	0.004
leather and related products	C15	0.4	0.9	0.066
wood, cork, straw and wicker products	C16	2.5	4.2	0.013
paper and paper products	C17	2.6	2.0	0.020
printing and reproduction	C18	1.0	1.7	0.021
coke and refined petroleum products	C19	7.9	0.5	0.367
chemicals and chemical products	C20	4.6	2.7	0.018
pharmaceutical products	C21	1.1	0.8	0.109
rubber and plastic products	C22	5.7	6.4	0.006
other non-metallic mineral products	C23	3.6	4.5	0.010
basic metals	C24	3.5	2.2	0.081
metal products	C25	6.3	10.5	0.003
computer, electronic and optical products	C26	2.8	2.1	0.061
electrical equipment	C27	3.8	3.5	0.030
machinery and equipment n.e.c.	C28	3.1	4.2	0.011
motor vehicles, trailers and semi-trailers	C29	9.1	6.0	0.028
other transport equipment	C30	1.4	1.5	0.031
furniture	C31	2.7	5.6	0.019
other products	C32	0.9	2.0	0.016
Mining and quarrying	B	4.3	5.7	0.148
Electricity, gas, steam and air conditioning	D	9.3	4.3	0.071
Water supply; sewerage, waste management	E	2.5	4.8	0.005

a Data come from Eurostat and Statistics Poland and cover the year 2015.

3.3. Empirical strategy

We investigate the pass-through of unit labour costs (ULC) to prices with the aim to make an inference on markup variation over the business cycle. To this end, we combine asymmetry and regime-dependence in the cointegration relation, by decomposing unit labour costs series into four partial sums conditional upon the business cycle position ('good' and 'bad' times in terms of the demand faced by the industry) and the direction of changes in the ULC:

$$ulc_t^{--} = \sum_{i=1}^{T} \Delta ulc_i \mathbb{I}_{\{\Delta demand_i \leq \tau \ \wedge \ \Delta ulc_i \leq 0\}},$$

$$ulc_t^{-+} = \sum_{i=1}^{T} \Delta ulc_i \mathbb{I}_{\{\Delta demand_i \leq \tau \ \wedge \ \Delta ulc_i > 0\}},$$

$$ulc_t^{++} = \sum_{i=1}^{T} \Delta ulc_i \mathbb{I}_{\{\Delta demand_i > \tau \ \wedge \ \Delta ulc_i \leq 0\}},$$

$$ulc_t^{+-} = \sum_{i=1}^{T} \Delta ulc_i \mathbb{I}_{\{\Delta demand_i > \tau \ \wedge \ \Delta ulc_i > 0\}}.$$

Under such specification, the cointegration equation takes the following form:

$$p_t = \delta_0 + \delta_1^{--} ulc_t^{--} + \delta_1^{-+} ulc_t^{-+} + \delta_1^{++} ulc_t^{++} + \delta_1^{+-} ulc_t^{+-} + \varepsilon_t, \tag{6}$$

where δ_1^{--} and δ_1^{-+} are the long-run responses of prices (p_t) to, respectively, falling and rising labour costs in 'bad' times, whereas δ_1^{++} and δ_1^{+-} constitute the corresponding responses in 'good' times. The error correction model correspondent to (6) can be expressed as:

$$\Delta p_t = \alpha_0 + \gamma\left(p_{t-1} - \delta_1^{--} ulc_{t-1}^{--} - \delta_1^{\mp} ulc_{t-1}^{\mp} - \delta_1^{++} ulc_{t-1}^{++} - \delta_1^{\pm} ulc_{t-1}^{\pm}\right) +$$
$$+ \sum_{i=1}^{p-1} \alpha_i \Delta p_{t-i} + \sum_{i=0}^{q-1} (\beta_i^{--} \Delta ulc_{t-i}^{--} + \beta_i^{\mp} \Delta ulc_{t-i}^{\mp} + \beta_i^{++} \Delta ulc_{t-i}^{++} + \tag{7}$$
$$+ \beta_i^{+-} \Delta ulc_{t-i}^{+-}) + \vartheta_t.$$

The threshold value for 'good' and 'bad' times (τ) is estimated by means of a grid search, so as to minimise the sum of squared residuals (Q) from (7):

$$\hat{\tau} = \underset{\tau \in D}{\operatorname{argmin}} \, \mathrm{Q}(\tau), \tag{8}$$

where the domain D of percentage changes in the demand faced by the industry is set by trimming extreme observations at the 25th and 75th percentile. The lag structure of ARDL models is established using the 'general-to-specific' approach and controlling for serial correlation of residuals.

The ARDL methodology – as a single equation approach – can produce biased estimates if variables are endogenously determined. Such endogeneity can be expected in the wage-price system. In our case, however, the sectoral structure of the data allows the unambiguous determination of the direction of causality (prices in a particular sector – unlike the overall price level – do not influence sectoral wages), which justifies the utilisation of a univariate analysis.

Table 3. Unit root tests[a]

Sectoral classification	Prices		Unit labour costs		Demand	
	I(1)	I(2)	I(1)	I(2)	I(1)	I(2)
Manufacturing of:						
food	−0.83	−4.26***	−0.43	−7.13***	−1.20	−6.17***
beverages	−2.31	−6.34***	−0.49	−11.32***	−2.12	−8.90***
tobacco	−1.42	−6.62***	−2.37	−3.92***	−2.55	−6.93***
textiles	−2.59	−4.84***	−1.38	−6.68***	0.83	−6.56***
wearing apparel	−0.87	−6.36***	−0.77	−5.69***	−1.83	−7.56***
leather and related products	0.09	−6.67***	−3.42	−7.35***	−0.54	−6.33***
wood, cork, straw and wicker products	−1.44	−4.71***	−1.70	−8.06***	−1.07	−7.41***
paper and paper products	−0.98	−4.98***	−1.41	−5.51***	−0.14	−6.13***
printing and reproduction	−2.11	−6.45***	−2.50	−4.28***	−0.29	−6.18***
coke and refined petroleum products	−1.70	−5.74***	−2.01	−8.22***	−2.30	−5.81***
chemicals and chemical products	−1.07	−5.20***	−1.29	−6.66***	−1.51	−6.80***
pharmaceutical products	0.92	−3.72***	−0.83	−8.59***	−1.74	−7.46***
rubber and plastic products	−1.13	−5.41***	−0.77	−6.64***	−1.55	−6.60***
other non-metallic mineral products	−2.04	−4.14***	0.09	−8.41***	−2.56	−5.75***
basic metals	−1.93	−4.72***	−2.08	−5.85***	−2.53	−4.83***
metal products	−1.96	−4.87***	−1.69	−5.92***	−1.40	−4.58***
computer, electronic and optical products	−2.52	−5.20***	−1.67	−6.13***	−2.90	−5.69***
electrical equipment	−1.44	−6.52***	−2.42	−3.13**	−2.97	−7.01***
machinery and equipment n.e.c.	−2.17	−5.07***	−0.22	−8.15***	−2.51	−8.04***
motor vehicles, trailers and semi-trailers	−2.07	−5.72***	−1.10	−5.93***	−1.04	−7.24***
other transport equipment	−0.73	−7.27***	−0.01	−9.75***	−0.74	−11.00***
furniture	−1.96	−5.06***	−1.76	−7.53***	−0.21	−7.29***
other products	−1.80	−6.40***	−2.91*	−8.96***	−1.62	−2.89**
Mining and quarrying	−1.88	−4.93***	−0.76	−5.46***	−2.05	−5.96***
Electricity, gas, steam and air conditioning	−2.37	−5.73***	−1.71	−6.76***	−1.78	−6.36***
Water supply; sewerage, waste management	−1.66	−4.14***	−2.16	−7.51***	−0.30	−6.63***

a The table presents the **ADF statistics** computed using regressions with an intercept, intercept and deterministic trend or without deterministic terms based on the visual inspection. One, two and three asterisks indicate statistical significance at the level of 10%, 5% and 1%, respectively.

Source: author's calculations.

Table 4. Estimation results[a,b,c]

Sectoral classification	Test for cointegration[a]	Test for cyclical variation[a]	$\hat{\delta}_1^{--}$	$\hat{\delta}_1^{-+}$	Symmetry: 'bad' times[a]	$\hat{\delta}_1^{+-}$	$\hat{\delta}_1^{++}$	Symmetry: 'good' times[a]
Manufacture of:								
food	48.18***	27.29***	1.36**	−0.83***	6.47**	−2.08**	2.90***	11.39***
beverages	40.18***	33.76***	0.35***	−0.08	7.08**	−0.46***	0.72***	22.18***
tobacco	48.56***	45.16***	−0.34***	0.22***	65.13***	0.54**	0.27**	1.84
textiles	18.86***	0.59	0.24	0.12	0.08	0.15	0.28	0.47
wearing apparel	13.02**	16.87***	−0.17	0.16***	13.25***	0.57***	−0.25	12.97***
leather and related products	25.95***	25.94***	0.55**	−0.36**	8.09**	−0.92***	0.95***	15.54***
wood, cork, straw and wicker products	17.30**	13.08***	−0.29*	−0.05	3.50*	0.01	0.05	0.01
paper and paper products	59.55***	53.52***	0.85***	0.24*	3.95*	−0.35**	−0.01	1.57
printing and reproduction	32.14***	12.49***	−0.52**	−0.28**	1.32	0.13	0.15	0.08
coke and refined petroleum products	24.96***	9.90**	0.56	1.39***	6.50**	0.74***	0.40**	6.50**
chemicals and chemical products	34.46***	27.34***	0.41**	0.13	3.91*	−0.42***	1.11***	33.51***
pharmaceutical products	27.55***	17.55***	0.01	−0.75	0.66	0.23*	0.67***	8.31**
rubber and plastic products	31.66***	28.09***	2.11***	0.35	4.01**	−0.60***	0.33	9.59***
other non-metallic mineral products	43.12***	31.53***	0.08	−0.40**	14.40***	−1.10**	0.39***	9.88***
basic metals	30.57***	5.88*	−0.02	−0.27**	6.42**	0.11	1.20**	0.02
metal products	45.25***	21.45***	0.85**	−0.33**	6.94**	−0.54***	−0.08	7.61**
computer, electronic and optical products	20.78***	8.36**	0.88***	0.86***	0.01	0.14	0.19	0.05
electrical equipment	31.41***	15.65***	0.53***	0.34**	12.05***	−0.02	−0.14	0.43
machinery and equipment n.e.c.	22.16***	7.36*	1.10*	−0.37	2.22*	0.01	1.05*	2.40
motor vehicles, trailers and semi-trailers	48.61***	7.19*	1.49**	0.75**	4.02*	0.00	0.14	0.25
other transport equipment	48.21***	38.09***	0.05	0.05	0.02	−0.16***	−0.05	38.07***
furniture	35.32***	26.23***	0.45***	0.18***	7.33**	−0.09**	0.18***	13.45***
other products	24.72***	11.17**	−0.20	−0.05	1.04	0.44***	0.60***	2.92*
Mining and quarrying	13.23**	8.68**	−0.40*	0.70**	9.39***	1.67***	−0.15	13.16***
Electricity, gas, steam and air conditioning	15.40**	2.49	1.08	0.17	0.67	0.12	0.91***	4.57**
Water supply; sewerage, waste management	12.23***	10.05**	1.40**	0.40***	3.34*	−0.64	1.46***	12.31***

a The table presents the Wald statistics. b One, two and three asterisks indicate statistical significance at the level of 10%, 5% and 1%, respectively. c Other estimation results are available on demand.

Source: author's calculations.

4. Empirical findings

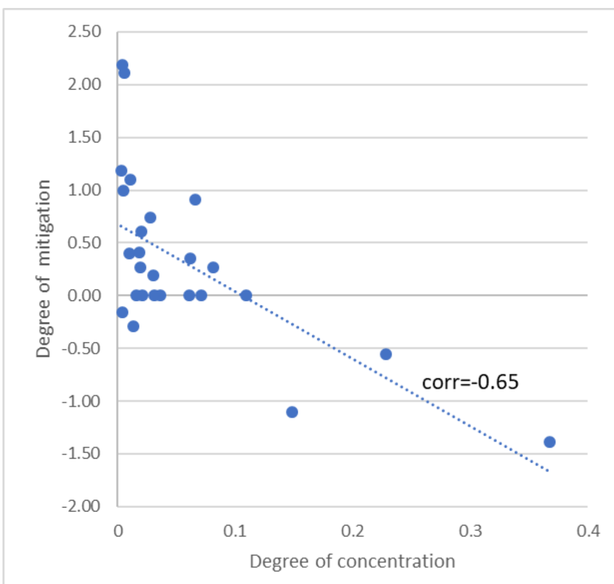
Cointegration analysis within the ARDL model as proposed by Pesaran et al. (2001) and Pesaran and Shin (1999) can be used for a mixture of $I(0)$ and $I(1)$ series, but not for variables of a higher degree of integration. For this reason, the $I(2)$-ness of the series has to be excluded. The results of unit root tests universally indicate integration of order 1 (see Table 3), allowing for the application of the ARDL methodology.

The existence of the long-run relationship is verified by means of the bounds test proposed by Pesaran et al. (2001) with the null hypothesis of the non-significant both the error correction parameter and the long-run elasticities. In all cases the null hypothesis is rejected, and in most cases the relation is non-degenerate (both the error correction parameter and at least one of the long-run elasticities is significantly different from zero), implying the existence of a meaningful long-run relationship between unit labour costs and prices (see Table 4).

In most sectors the test for cyclical variation is positive, i.e. the null hypothesis of symmetrical price responses to changing costs in 'good' and 'bad' times ($\delta_1^{--} = = \delta_1^{-+} = \delta_1^{++} = \delta_1^{+-}$) is rejected (Table 4). Thus, the pass-through of unit labour costs to prices in Polish industry is conditional upon the business cycle position, implying cyclical variation in markups. In the majority of industries, the degree of pass-through in 'good' times is significantly higher in response to an increase in unit labour costs than to a decrease, suggesting an amplifying impact of markup adjustments on prices. In many sectors the elasticities of prices with respect to falling unit labour costs are even negative. Therefore, in favourable demand conditions prices are raised even in the face of falling costs, thereby increasing markups. In 'bad' times the opposite pattern seems to prevail, with decreases in unit labour costs feeding through into prices to a significantly greater extent than increases. This implies a mitigating role of markup adjustments in economic slack. Only in a few sectors the opposite pattern can be observed, i.e. a mitigating behaviour of markups during cyclical upturns and amplifying during downturns. This is especially pronounced in the case of manufacturing of *tobacco*, *coke and refined petroleum products*, as well as *mining and quarrying*, all of which are characterised by a high degree of concentration as defined by the Herfindahl-Hirschman index (see Table 2). In several sectors no clear-cut pattern of pass-through variation emerges from the estimation results.

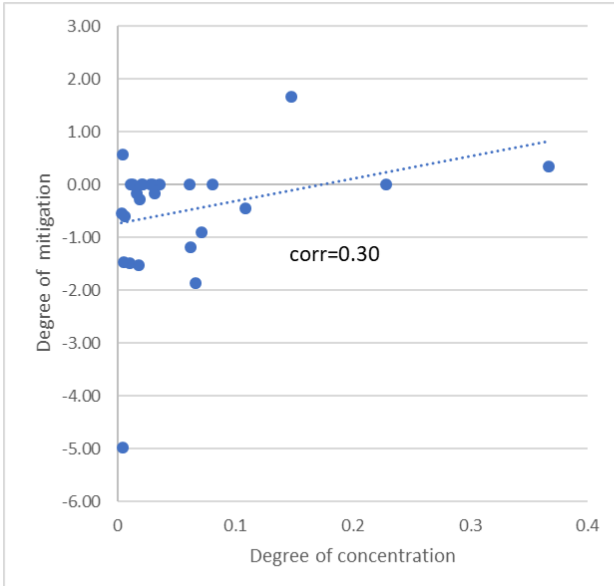
The obtained estimates, indicating in most sectors a mitigating impact of markups on prices in 'bad' times together with an amplifying effect in 'good' times, suggest the prevalence of markup procyclicality in the Polish industry. Nonetheless, the sectors are characterised by various degrees of mitigation/amplification, and some of them exhibit a different pattern of adjustment. In order to shed some light on the factors behind this heterogeneity, we tabulated each industry's degree of mitigation (defined as a difference between price response to a decrease and to an increase in costs, with non-significant differences imputed with zero) against its level of concentration (approximated by the Herfindahl-Hirschman index). There seems to be a significant, albeit moderate, relationship between the industry's degree of concentration and the adjustment pattern it exhibits (see Figure 1 and 2) with Pearson's correlation coefficient equal to 0.30 in 'good' times and −0.65 in 'bad' times (significant at the level of 0.05 and 0.01, respectively). In 'good' times, it seems that the more concentrated the industry, the more mitigation (less amplification) provided by the pass-through, i.e. the less the cost hikes feed through into prices relative to the cost drops. In 'bad' times, on the other hand, less concentrated sectors exhibit more mitigating behaviour. Higher degree of competition seems, therefore, preferable for the sake of shock-absorption in economic downturns.

Figure 1. The degree of mitigation[a] as a function of an industry's concentration[b] in 'bad' times.



a Degree of mitigation defined as a difference between price response to a decrease and to an increase in costs. b Degree of concentration is approximated by the Herfindahl–Hirschman index.
Source: author's calculations.

Figure 2. The degree of mitigation[a] as a function of an industry's concentration[b] in 'good' times



a Degree of mitigation defined as a difference between price response to a decrease and to an increase in costs. b Degree of concentration is approximated by the Herfindahl-Hirschman index.
Source: author's calculations.

5. Conclusions

This study aims at estimating a cyclical pattern in the cost pass-through. To this end, a regime-dependent framework is proposed, allowing the estimation of the pass-through parameters separately in cyclical upturns and downturns. The methodology is applied to the Polish industrial sectors.

The obtained results point to the prevalence of markup procyclicality in the Polish industry, since the impact of markups on prices is mitigating in 'bad' times and amplifying in 'good' times. In some industries, markup adjustments can be directly inferred upon, given that the response to increasing (decreasing) unit labour costs in 'bad' ('good') times entails lowering (raising) prices, reflective of negative (positive) changes in markups. In a few cases, however, the estimated pattern of adjustments is suggestive of markup counter- or acyclicality. The degree of procyclicality seems to be positively correlated with the level of competition, corroborating a large body of evidence dating back to the Domowitz et al. (1986, 1988), thus validating the proposed methodology of assessing the behaviour of markups based on the cyclicality of the cost pass-through.

Thus, in the majority of industries the estimates support the hypothesis of a mitigating effect of markups on business cycle fluctuations (markups boost prices in economic upturns and alleviate the pressure on them during downswings, thus, respectively, curbing and stimulating the demand). Polish industrial firms do not seem to take advantage of wage concessions in economic slack in order to boost their profits. In most industries wage flexibility seems, therefore, to be an appropriate policy prescription for economic stabilisation.

References

Atkeson, A., Burstein, A. (2008). Pricing-to-Market, Trade Costs, and International Relative Prices. *American Economic Review*, *98*(5), 1998–2031. https://doi.org/10.1257/aer.98.5.1998

Bils, M. (1987). The Cyclical Behavior of Marginal Cost and Price. *American Economic Review*, *77*(5), 838–855.

Bils, M. (1989). Pricing in a Customer Market. *Quarterly Journal of Economics*, *104*(4), 699–718. https://doi.org/10.2307/2937863

Bils, M., Kahn, J. A. (2000). What Inventory Behavior Tells Us About Business Cycles. *American Economic Review*, *90*(3), 458–481. https://doi.org/10.1257/aer.90.3.458

Chevalier, J. A., Sharfstein, D. S. (1996). Capital-Market Imperfections and Countercyclical Markups: Theory and Evidence. *American Economic Review*, *86*(4), 703–725. https://scholar.harvard.edu/files/davidscharfstein/files/capital_market_imperfections_and_countercyclical_markups.pdf

Chirinko, R. S., Fazzari, S. M. (1994). Economic Fluctuations, Market Power, and Returns to Scale, Evidence from Firm-Level Data. *Journal of Applied Econometrics*, *9*(1), 47–69. https://doi.org/10.1002/jae.3950090105

Clementi, G. L., Palazzo, B. (2016). Entry, Exit, Firm Dynamics, and Aggregate Fluctuations. *American Economic Journal: Macroeconomics*, *8*(3), 1–41. https://doi.org/10.1257/mac.20150017

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, *64*(2), 247–254. https://doi.org/10.1093/biomet/64.2.247

Domowitz, I., Hubbard, R. G., Petersen, B. C. (1986). Business cycles and the relationship between concentration and price-cost margins. *The RAND Journal of Economics*, *17*(1), 1–17. https://doi.org/10.2307/2555624

Domowitz, I., Hubbard, R. G., Petersen, B. C. (1988). Market Structure and Cyclical Fluctuations in U.S. Manufacturing. *The Review of Economics and Statistics*, *70*(1), 55–66. https://doi.org/10.2307/1928150

Gali, J. (2013). Notes for a new guide to Keynes (I): wages, aggregate demand, and employment. *Journal of the European Economic Association*, *11*(5), 973–1003. https://doi.org/10.1111/jeea.12032

Galí, J., Gertler, M., López-Salido, J. D. (2007). Markups, Gaps, and the Welfare Costs of Business Fluctuations. *Review of Economics and Statistics*, *89*(1), 44–59. https://doi.org/10.1162/rest.89.1.44

Galí, J., Monacelli, T. (2016). Understanding the Gains from Wage Flexibility: The Exchange Rate Connection. *American Economic Review*, *106*(12), 3829–3868. https://doi.org/10.1257/aer.20131658

Genesove, D., Mullin, W. P. (1998). Testing static oligopoly models: conduct and cost in the sugar industry. 1890–1914. *Rand Journal of Economics*, *29*(2), 355–377. https://doi.org/10.2307/2555893

Gilchrist, S., Schoenle, R., Sim, J., Zakrajšek, E. (2017). Inflation Dynamics during the Financial Crisis. *American Economic Review*, *107*(3), 785–823. https://doi.org/10.1257/aer.20150248

Goldberg, P. K., Hellerstein, R. (2013). A Structural Approach to Identifying the Sources of Local Currency Price Stability. *The Review of Economic Studies*, *80*(1), 175–210.

Gopinath, G., Itskhoki, O., Rigobon, R. (2010). Currency Choice and Exchange Rate Pass-Through. *American Economic Review*, *100*(1), 304–336. https://doi.org/10.1257/aer.100.1.304

Gottfries, N. (1991). Customer Markets, Credit Market Imperfections and Real Price Rigidity. *Economica*, *58*(231), 317–323. https://doi.org/10.2307/2554819

Green, E. J., Porter, E. H. (1984). Noncooperative Collusion under Imperfect Price Competition. *Econometrica*, *52*(1), 87–100. https://doi.org/10.2307/1911462

Greenwald, B., Stiglitz, J. E., Weiss, A. (1984). Information Imperfections in the Capital Market and Macroeconomic Fluctuations. *American Economic Review*, *74*(2), 194–199.

Hall, R. E. (1986). Market Structure and Macroeconomic Fluctuations. *Brookings Papers on Economic Activity*, (2), 285–322. https://doi.org/10.2307/2534476

Hall, R. E. (1988). The Relation between Price and Marginal Cost in U.S. Industry. *Journal of Political Economy*, *96*(5), 921–947. https://doi.org/10.1086/261570

Hall, R. E. (2012). *The Cyclical Response of Advertising Refutes Counter-Cyclical Profit Margins in Favor of Product Market Frictions* (NBER Working Paper No. 18370). https://doi.org/10.3386/w18370

Hansen, B. E. (1996). Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis. *Econometrica*, *64*(2), 413–430. https://doi.org/10.2307/2171789

Hansen, B. E. (2000). Sample Splitting and Threshold Estimation. *Econometrica, 68*(3), 575–603. https://doi.org/10.1111/1468-0262.00124

Hellerstein, R. (2008). Who bears the cost of a change in the exchange rate? Pass-through accounting for the case of beer. *Journal of International Economics*, *76*(1), 14–32. https://doi.org/10.1016/j.jinteco.2008.03.007

Klemperer, P. (1995). Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade. *The Review of Economic Studies*, *62*(4), 515–539. https://doi.org/10.2307/2298075

Lee, Y., Mukoyama, T. (2015). Entry and Exit of Manufacturing Plants over the Business Cycle. *European Economic Review*, *77*, 20–27. https://doi.org/10.1016/j.euroecorev.2015.03.011

Martins, J. O., Scarpetta, S. (2002). Estimation of the Cyclical Behaviour of Mark-ups: A Technical Note. *OECD Economic Studies*, *34*(1), 173–188. https://dx.doi.org/10.2139/ssrn.335543

Nakamura, E. (2008). Pass-Through in Retail and Wholesale. *American Economic Review, 98*(2), 430–437. https://doi.org/10.1257/aer.98.2.430

Nakamura, E., Zerom, D. (2010). Accounting for Incomplete Pass-Through. *The Review of Economic Studies*, *77*(3), 1192–1230. https://doi.org/10.1111/j.1467-937X.2009.589.x

Nekarda, Ch. J., Ramey, V. A. (2013). *The Cyclical Behavior of the Price-Cost Markup* (NBER Working Paper No. 19099). https://doi.org/10.3386/w19099

Okun, A. M. (1981). *Price and Quantities: A Macroeconomic Analysis*. Oxford: Basil Blakwell.

Pesaran, M. H., Shin, Y. (1999). An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis. In S. Strøm (Ed.), *Econometrics and Economic Theory in the 20th Century:*

The Ragnar Frisch Centennial Symposium (pp. 371–413). Cambridge: Cambridge University Press. https://doi.org/10.1017/CCOL521633230.011

Pesaran, M. H., Shin, Y., Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics*, *16*(3), 289–326. https://doi.org/10.1002/jae.616

Phelps, E. S., Winter, S. G. (1970). Optimal Price Policy under Atomistic Competition. In E. S. Phelps (Ed.), *Microeconomic Foundations of Employment and Inflation Theory* (pp. 309–337). New York: Norton.

Roeger, W. (1995). Can Imperfect Competition Explain the difference between Primal and Dual Productivity Measures? Estimates for U.S. Manufacturing. *Journal of Political Economy*, *103*(2), 316–330. https://doi.org/10.1086/261985

Rotemberg, J. J., Saloner, G. (1986). A Supergame-Theoretic Model of Price Wars during Booms. *American Economic Review*, *76*(3), 390–407. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=1DEC95EC5284A7DA2BE74B15ADF2F49D?doi=10.1.1.676.1323&rep=rep1&type=pdf

Rotemberg, J. J., Woodford, M. (1991). Markups and the Business Cycle. In O. J. Blanchard, S. Fisher (Eds.), *NBER Macroeconomics Annual 1991, Volume 6*, (pp. 63–129). Cambridge, MA: MIT Press. https://doi.org/10.1086/654159

Rotemberg, J. J., Woodford, M. (1992). Oligopolistic Pricing and Effects of Aggregate Demand on Economic Activity. *Journal of Political Economy*, *100*(6), 1153–1205. https://doi.org/10.1086/261857

Rotemberg, J. J., Woodford, M. (1999). The Cyclical Behavior of Prices and Costs. In J. B. Taylor, M. Woodford (Eds.), *Handbook of Macroeconomics, Volume 1B* (pp. 1051–1135). Amsterdam: Elsevier B.V. https://doi.org/10.1016/S1574-0048(99)10024-7

Shin, Y., Yu, B., Greenwood-Nimmo, M. (2014). Modelling Asymmetric Cointegration and Dynamic Multipliers in a Nonlinear ARDL Framework. In W. C. Horrace, R. C. Sickles (Eds.), *Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications* (pp. 281–314). New York: Springer. https://dx.doi.org/10.2139/ssrn.1807745

Stiglitz, J. E. (1984). Price Rigidities and Market Structure. *American Economic Review*, *74*(2), 350–355. https://doi.org/10.7916/D8FN1H7B

Tian, C. (2018). Firm-level entry and exit dynamics over the business cycles. *European Economic Review*, *102*, 298–326. https://doi.org/10.1016/j.euroecorev.2017.12.011

Weyl, E. G., Fabinger, M. (2013). Pass-Through as an Economic Tool: Principle of Incidence under Imperfect Competition. *Journal of Political Economy*, *121*(3), 528–583. https://doi.org/10.1086/670401

Przegląd Statystyczny. Statistical Review, 2020, vol. 67, 2, 114–151 DOI: 10.5604/01.3001.0014.5726

Assessment of the size of VaR backtests for small samples

Daniel Kaszyński,[a] Bogumił Kamiński,[b] Bartosz Pankratz[c]

Abstract. The market risk management process includes the quantification of the risk connected with defined portfolios of assets and the diagnostics of the risk model. Value at Risk (VaR) is one of the most common market risk measures. Since the distributions of the daily P&L of financial instruments are unobservable, literature presents a broad range of backtests for VaR diagnostics. In this paper, we propose a new methodological approach to the assessment of the size of VaR backtests, and use it to evaluate the size of the most distinctive and popular backtests. The focus of the paper is directed towards the evaluation of the size of the backtests for small-sample cases – a typical situation faced during VaR backtesting in banking practice. The results indicate significant differences between tests in terms of the *p*-value distribution. In particular, frequency-based tests exhibit significantly greater discretisation effects than duration-based tests. This difference is especially apparent in the case of small samples. Our findings prove that from among the considered tests, the Kupiec TUFF and the Haas Discrete Weibull have the best properties. On the other hand, backtests which are very popular in banking practice, that is the Kupiec POF and Christoffersen's Conditional Coverage, show significant discretisation, hence deviations from the theoretical size.

Keywords: Value at Risk, market risk management, backtesting, empirical size assessment.

JEL: C00, C12, C15, D81, G32

1. Introduction

In 2009, the Basel Committee on Banking Supervision has introduced the Basel II Accord, which includes recommendations for banks as well as for regulators operating in the EU (Basle Committee on Banking Supervision [BCBS], 2009). Within the Basel II framework, financial institutions, in particular, are recommended to ensure capital buffers against market risks – this recommendation is also sustained in Basel III, which will be implemented (and come into force) in January 2022. The market risk management process carried out by financial institutions includes the quantification of the risk connected with defined portfolios of assets. One of the most commonly used risk measures that has gained significant attention is Value at Risk (VaR). Among the consequences of implementing the Basel Accord is that banks are required to perform proper diagnostics, i.e. backtests of their VaR models.

[a] SGH Warsaw School of Economics, Institute of Econometrics, Decision Analysis and Support Unit, e-mail: dkaszy@sgh.waw.pl (corresponding author), ORCID: https://orcid.org/0000-0002-0865-0732.

[b] SGH Warsaw School of Economics, Institute of Econometrics, Decision Analysis and Support Unit, e-mail: bkamins@sgh.waw.pl, ORCID: https://orcid.org/0000-0002-0678-282X.

[c] SGH Warsaw School of Economics, Institute of Econometrics, Decision Analysis and Support Unit, e-mail: bpankra@sgh.waw.pl, ORCID: https://orcid.org/0000-0001-7618-9119.

A standard approach to backtesting a predictive model involves the comparison of *ex-post* realisations with the *ex-ante* forecasts of interest values (Hurlin & Tokpavi, 2006). This process is straightforward if the *ex-post* realisations (observations) of the forecasted values are measurable (i.e. observable). In the case of VaR backtesting, this approach is not applicable since the VaR is a quantile of the distribution of a random variable. It means that one can only observe the realisation of this random variable (Jorion, 2010), and not its distribution. Therefore, VaR backtesting is a non-trivial task, and significant research has been devoted to the development of appropriate test procedures, c.f. Berkowitz et al. (2011), Hurlin (2013) or Nieto and Ruiz (2016).

A natural approach to the assessment of *ex-ante* VaR forecast is to base it on *ex-post* observed series of times when the VaR is violated. Such a series should possess two essential properties (Hurlin & Tokpavi, 2006):

- *unconditional coverage*, i.e. the probability of a violation in a given period should be equal to the VaR level;
- *independence of violations*, i.e. the probability of violation in a given period should not depend on the occurrence of violations in the past.

Based on these two properties, a broad range of statistical tests for the VaR model evaluation have been proposed in literature. Hurlin (2013) classifies the VaR backtests into one of the following types:

- *Frequency-based* tests, which are based on the number of observed VaR violations, i.e. observations for which the daily P&L is below the calculated VaR, and the expected number of violations.
- *Independence-based* tests, which measure the dependency of VaR violations between consecutive days; these tests validate whether the probability of VaR violations depends on the occurrence of previous VaR violations.
- *Duration-based* tests that use the fact that, assuming the correctness of the VaR model, the periods between consecutive violations should follow the geometric distribution. Duration-based tests validate the latter.
- *Magnitude-based* tests, which are based not only on the number of VaR violations, but also on the severity of the violation: the bigger the difference between the P&L and the corresponding forecasted VaR during the occurrence of a violation, the more severe the violation.
- *Multivariate-based* tests, which evaluate the risk model based on more than one level of the VaR; these tests measure the correctness of VaR predictions based on joint tests for multiple VaR levels, e.g. 1% and 5% jointly.

In this paper, we argue that during the application of VaR backtesting procedures in practice, the samples of *ex-post* data are small (i.e. involve short time series) relative to the VaR level, i.e. the number of observations of VaR violations is scarce.

Within this perspective, we review the current approaches to VaR backtesting. Due to a large body of literature on this subject, we focus on backtests which consider series of violations for a fixed VaR level, further denoted by α. Technically, this class of tests is designed to check if a sequence of 0 and 1 values (non-violation and violation observations, respectively) is generated as IID Bernoulli variables with the probability of success equal to α. We have presented VaR backtesting results based on an independently developed library containing a set of the most popular backtests, allowing an efficient, intuitive simulation and straightness to benchmark. Given the typology of VaR backtests mentioned earlier, we focus on frequency-based, independence-based and duration-based tests.

Several reviews of backtesting procedures have been recently presented in literature. One of the first texts that compare different VaR backtesting procedures is Campbell (2006). This article describes the Kupiec (1995) proportion of failures test, the Christoffersen (1998) independence and joint tests, tests based on multi-level VaR, the Lopez (1998) loss function-based test and the Pearson Q test for goodness of fit.

Nieto and Ruiz (2016) provide a recent review of methodological and empirical achievements in VaR estimation and backtesting. In terms of VaR backtests, this 2016 study describes the most popular tests which are based on the binary hit variable for single and multiple α levels. The authors also present an approach based on the loss function proposed by Lopez (1998).

Zhang and Nadarajah's (2017) paper focuses solely on VaR backtesting. The authors provide descriptions of different procedures, referring to source papers for further details on power and size evaluations. The research presents the most popular backtest approaches and 28 different tests.

The above-mentioned studies provide mainly qualitative descriptions of backtesting procedures and refer readers to source articles for an evaluation of their statistical properties. Evers and Rohde's (2014) article additionally presents the results of a quantitative size evaluation of selected backtesting procedures. The scope of the analysed tests covers the Kupiec (1995) proportion of failures test, the Christoffersen (1998) conditional coverage (with a division into independence and joint tests), the Escanciano and Olmo (2011) test, the Christoffersen and Pelletier (2004) duration test, and Candelon et al. (2011). As pointed out by the authors, most of the evaluated tests present problems relating to heavy-size distortions for small samples. This finding is consistent with conclusions presented in some other research papers (e.g. Escanciano and Olmo (2011), and indicates that the proposed univariate backtests display size-related issues in small samples. It needs to be pointed out that some research (Małecka, 2014) shows that the empirical size for large samples is greater than for small samples (which is also presented in our research – see Fig. 7). Nevertheless, the current studies on the subject do not present a coherent approach comparing different VaR backtests – moreover, the cited papers consider only the most popular backtests. Therefore, in our opinion, there is a need for the unification of the backtests' size evaluation methodology.

There are two criteria that can be used to assess a backtest procedure (and, in fact, any statistical test), namely *size* and *power* of the test (Everitt, 2006).

The size of the test is defined as the probability of rejecting the H_0 when it is met. The size of the test is also called Type I error. The power of the test is defined as the probability of rejecting the null hypothesis H_0 when the alternative H_1 is true. The power of the test strictly corresponds to Type II error (i.e. not rejecting the null hypothesis when it is false). The power of the test is one minus the probability of Type II error (Altman, 1991). In this text, we propose a new methodology for the assessment of the test size in the case of small *ex-post* sample size and apply it to the VaR backtesting procedures proposed in literature. The motivation for this work is threefold.

Firstly, the VaR backtesting literature mostly refers the readers to source papers (i.e. papers introducing particular backtests) when discussing test sizes. In the study presented in this paper, we develop a unified framework consistently applied to all considered tests, which enabled us to obtain results of test size analysis which are directly comparable.

Secondly, when the *ex-post* sample size is small, many VaR tests exhibit high discretisation of test statistics (i.e. they take only a small number of possible values with significant probabilities). This means that the evaluation of the size of a given test for a fixed p-value can be misleading, as one cannot easily assess if the distribution of the test statistics has a large jump near the p-value threshold or not. Therefore we adopted a test-size visualisation and assessment procedure that enables us to check by how much the distribution of p-values of the test diverges from the uniform distribution over a [0,1] interval (a p-value of an ideal test should have such a distribution), after Murdoch et al. (2008).

Thirdly, the recent literature regarding backtesting has expanded, but our study focuses on tests whose size has not been analysed in earlier publications. An additional benefit of this unified approach is that for the purpose of the analyses presented in the article, we have implemented backtesting procedures reviewed within one software package. The library is available free of charge to everyone at https://github.com/dkaszynski/VVaR. One particular feature of the implemented procedures is that corner cases of all the considered statistical tests are carefully managed, which is often not the case, even in source papers introducing them. For instance, in relation to small samples and low values of α, an important issue to be appropriately dealt with is the case of no violations of VaR in an *ex-post* data set.

To sum up, the study presented in this paper contributes to VaR backtesting research in the following ways: 1) it provides a systematic evaluation and comparison of a wide range of VaR backtest procedures, including the ones most recently proposed in literature, that has been carried out for the first time; 2) it proposes a new method of analysing the size of VaR backtests evaluated on small samples; 3) it carefully reviews the specifications of all the analysed tests in order to properly manage corner cases, and offers a software package implementing them.

The paper has the following structure: Section 2 provides a formal definition of VaR and the proposed methodology for the procedure of verifying the VaR backtest sizes. Section 3 presents a comprehensive review of VaR backtesting procedures. In Section 4 the results of numerical simulations of the considered backtesting procedures are discussed. The fifth section consists of conclusions and remarks for future studies.

2. Methodology

In this section, formal definitions of Value at Risk (VaR) and the backtesting procedure (also referred to as backtest) are provided.

2.1. Value at risk notation

Let $VaR_\alpha(X)$ be a VaR of a random variable X with a tolerance level of α. The formal notation is as follows:

$$VaR_\alpha(X) = -inf\{x \in \boldsymbol{R}: Pr(X \leq x) > \alpha\}. \tag{1}$$

Therefore, if X is a continuous random variable, we receive the following:

$$Pr(X \leq -VaR_\alpha(X)) = \alpha. \tag{2}$$

If X is not assumed to be continuous, we have in general:

$$Pr(X \leq -VaR_\alpha(X)) \geq \alpha \tag{3}$$

and $\lim\limits_{x \to VaR_\alpha(X)^+} Pr(X \leq -x\,) \leq \alpha$. In the further parts of this paper we assume that X is continuous, unless explicitly stated otherwise.

Given those definitions, we will consider VaR forecasts in discrete time $t \in \boldsymbol{N}$. In this article, time units are assumed to be days.

Let us consider an asset whose daily returns are denoted as r_t. By $R_{t|t'}$, we denote a random variable describing the r_t distribution, which takes into account all the information available at time t'. Clearly when $t' \geq t$, then $R_{t|t'}$ is constant with $\Pr\left(R_{t|t'} = r_t\right) = 1$. Most of the time we will assume that $t' = t - 1$ and, therefore, we will use the notation $R_t \coloneqq R_{t|t-1}$.

Having assumed the above, we receive a formally defined value $VaR_\alpha(R_{t|t'})$, where $t' < t$, which is a true and unknown value of Value at Risk at time t assessed at time t' with an α tolerance level.

2.2. Backtesting – definition

Now consider that we are given a forecast for $VaR_\alpha(R_{t|t'})$ in time t', which we will denote as $VaR_\alpha^{t|t'}$. As in the case of the definition of R_t, we write VaR_α^t when $t_0 = t - 1$.

Since $VaR_\alpha(R_t)$ is not observable if we want to assess the quality of VaR_α^t, we can only test it against the observed values of r_t. Let us denote a random function, which indicates if value x was less than or equal to v, by $S(v, x) = 1_{[-inf,v]}(x)$. Using this notation, $S(VaR_\alpha^t, r_t)$ takes the value of 1 if the observed r_t was less than or equal to the value of the prediction of a VaR, or otherwise 0. Additionally, $S_\alpha^t = S(VaR_\alpha^t, R_t)$ is a sequence of random variables and $s_\alpha^t = S(VaR_\alpha^t, r_t)$ is a sequence of their realisations. We will call the sequence of forecasts VaR_α^t unbiased if $VaR_\alpha^t = VaR_\alpha(R_t)$.

Since it is not possible to directly verify this condition, we will check the implied properties of S_α^t. Formally, if a sequence of forecasts is unbiased, then we have $P_r(S_\alpha^t = 1) = E(S_\alpha^t) = \alpha$. This is a condition that can be verified. Observe that S_α^t is defined as subject to information available until time $t - 1$. In particular, this means that S_α^t is a sequence of independent Bernoulli random variables with an α probability of success. On the other hand, if $VaR_\alpha^t \neq VaR_\alpha(R_t)$ for at least time moment t, then the sequence S_α^t does not display this property.

In order to validate the assumption that VaR forecasts are unbiased at tolerance level α, we can use tests which check if the sequence s_α^t was sampled from a process generating independent Bernoulli random values with an α probability of success.

Less formally, backtesting, also referred to as *reality check* (Jorion, 2007), is a statistical framework of techniques for verifying the accuracy of risk models (including VaR models) and a part of a broader *model validation process* (Jorion, 2007). In essence, VaR backtesting refers to the comparison of P&L results with risk measures generated by the Value at Risk model. As stated by BCBS (1996), a backtest consists of a periodic comparison of daily Value at Risk measures to the subsequent daily P&L. The Value at Risk measures are intended to be under $1-\alpha\%$ trading outcomes.

2.3. Notation

Now let us assume that we have a sequence s_α^t sampled for time points from 1 to n. In order to simplify the notation, we add two virtual values s_α^0 and s_α^{n+1}, both equal to 1. We denote an increasing sequence of time points for which $s_\alpha^{v_i}$ equals 1 by v_i. Note that the length l of this sequence is at least two and at most $n+2$ elements. Based on this sequence, we can define inter-event times $d_i = v_{i+1} - v_i - 1$ for $i \in \{1, \dots, l-1\}$. Now observe that if we sample the sequence s_α^t as independent Bernoulli random values with a probability of success α, then the random variable D representing the value of d_i uniformly selected from the set $\{d_1, \dots, d_{l-1}\}$ has censored the geometric distribution (let us stress here that we consider the distribution of D before sampling s_α^t). Formally, the notation is as follows:

$$Pr(D=i) = \begin{cases} \alpha(1-\alpha)^i, & if\ 0 \le i \le n-1 \\ (1-\alpha)^n, & if\ i=n \\ 0, & otherwise \end{cases}. \tag{4}$$

Observe that if T is a random variable with geometric distribution with success probability α that is independent from the random variable D, then the variable $\widetilde{D}$ is defined as

$$\widetilde{D} = \begin{cases} D, & if\ D < n \\ D+T, & if\ D = n \end{cases} \tag{5}$$

and displays a geometric distribution with success probability α. This fact is utilised in duration-based tests, i.e. tests evaluating whether the duration between VaR violations are drawn from a geometric distribution.

2.4. Size evaluation methodology

Consider a statistical test with significance level p. By q we will denote the size of this test, i.e. the probability of the rejection of H_0 under H_0. We say that the test has a proper size at the significance level p if $p=q$. Additionally, we will say that it has a strictly proper size if it has a proper size for all $p \in [0,1]$.

We can state that the test is oversized (rejects H_0 too often) at the significance level p if $q > p$, and undersized (rejects H_0 too rarely) if $q < p$.

We define oversize frequency as a measure of the set $T_O = \{p \in [0,1]: q > p\}$ and the average oversize as $A_O = \int_{T_O} (q - p)dp / \int_{T_O} dp$. By the same token, we define the undersize frequency as a measure of the set $T_U = \{p \in [0,1] : q < p\}$ and the average undersize as $A_U = \int_{T_U} (p - q)dp / \int_{T_U} dp$.

Observe that in finite samples it is impossible for a test to have a uniformly proper size, because typically the set of possible values of q over all values of $p \in [0,1]$ is finite. We will denote this set by Q. Therefore, we will say that the test has a *weakly proper size* if it has a proper size for all p that belong to set Q. In practice, this property is realised when a function $q(p)$ has a property $q(p^-) < p \leq q(p)$ for all $p \in Q$, or, equivalently, a function $p(q)$ has a property $p \in p(\{q\})$.

For each analysed test, we will discuss the given VaR level α and sample size n if it has a weakly proper size, and report:

- T_O, i.e. oversize frequency (if for all $p \in [0,1]$ the test does not exhibit a proper size, then $T_O + T_U = 1$);
- T_U, i.e. undersize frequency;
- A_O, i.e. average oversize value;
- A_U, i.e. average undersize value;
- A, i.e. average deviation from the correct size.

3. Evaluated backtests

This section provides a detailed description of the tests that have been assessed in terms of size. For convenience, we define $h_i = v_i - v_i - 1 = d_i + 1$, where $i \in 1, \dots, l - 1$, which may be interpreted, c.f. Małecka (2014), as the period of time between two consecutive VaR violations; in this manner, we denote the time until the first VaR violation by h_1, and the number of days after the last 1 in the hit sequence by h_{l-1}.

3.1. Kupiec 1995 – Proportion of failures

The proportion of failures – POF, also referred to as the *Unconditional coverage test*, examines how many times a VaR is violated over a given time span (Kupiec, 1995). The null hypothesis assumes that the observed violation rate equals the expected number of VaR violations. This test belongs to the category of the frequency-based ones, as presented in Section 1. The statistic of the test takes the following form:

$$LR_{POF}(\alpha, n, s) = -2\,log\left(\frac{(1-\alpha)^{n-s}\alpha^s}{(1-\hat{\alpha})^{n-s}\,\hat{\alpha}^s}\right) \overset{asy}{\sim} \chi^2\,(1), \tag{6}$$

where $s = \sum_{t=1}^{n} s_t^{\alpha}$, and $\hat{\alpha} = \frac{s}{n}$.

Observe that when $s = 0$ and $s = n$, this formula is undefined. In those cases, the limit of the LR_{POF} expression in 0^+ and n^-, respectively, can be used, because they exist and are finite, namely $LR_{POF}(\alpha, n, 0) = -2n\log(1 - \alpha)$ and $LR_{POF}(\alpha, n, n) = -2nlog(\alpha)$.

3.2. Binomial test

An alternative approach to Kupiec's POF test is the one presented by Jorion (2007). Under the null hypothesis, the number of VaR violations follows the Bernoulli distribution, and by assuming that n is large, one can use the central limit theorem and approximate the binomial distribution with a normal distribution, i.e. Wald's statistics:

$$f(\alpha, n, s) = \frac{s - \alpha n}{\sqrt{\alpha(1-\alpha)n}} \overset{asy}{\sim} N(0,1). \tag{7}$$

In contrast to Kupiec's POF test, the $f(\alpha, n, s)$ statistic is well-defined also when no violation is observed. The possibility that there was no violation of VaR in the case of small-sample time series (i.e. financial backtesting), especially for a small α, is not trivial (Campbell, 2006). The Binomial test is also a frequency-based test.

3.3. Christoffersen 1998 tests

The previously-mentioned unconditional coverage tests are based solely on the proportion of VaR violations. Alternatively, Christoffersen (1998) proposed a very influential and popular conditional coverage test, where the null hypothesis assumes that $E[s_t^\alpha | s_{t-1}^\alpha] = \alpha$. This test verifies the frequency of the VaR violation occurrence as well as its independence. In terms of the independence property, it is evaluated using the following:

$$LR_{IND}(s) = -2\, log \left(\frac{\pi_{\cdot 0}^{n_{00}+n_{10}} \pi_{\cdot 1}^{n_{01}+n_{11}}}{\pi_{00}^{n_{00}} \pi_{01}^{n_{01}} \pi_{10}^{n_{10}} \pi_{11}^{n_{11}}} \right) \overset{asy}{\sim} \chi^2(1), \tag{8}$$

where n_{ij} is the number of observations, s_t^α stands for i and s_{t+1}^α for j, $\pi_{ij} = n_{ij} n_{ij} / \sum_j n_{ij}$, and $\pi_{\cdot j} = \sum_i n_{ij} / \sum_{i,j} n_{ij}$.

The likelihood ratio of conditional coverage test which takes into account Kupiec's unconditional test likelihood and independence likelihood results is as follows:

$$LR_{CC}(\alpha, n, s) + LR_{IND}(s) + LR_{POF}(\alpha, n, s) \overset{asy}{\sim} \chi^2(2). \tag{9}$$

Note that the LR_{CC} tests only the first order autocorrelation of the VaR violations – the process generating VaR violations in H_0 of the independence test is assumed to be a first-order Markov chain with independence of violation / non-violation state transitions.

3.4. Kupiec 1995 – Time until first failure

Kupiec (1995) also presents an alternative approach to examining the proportion of VaR violations – the time until the first failure (TUFF) test. The null hypothesis assumes that the random variable denoting the number of days until the first VaR violation is geometrically distributed – note that the definition of geometric distribution may include two distinct cases: the series $1, 2, \ldots$ and the series $0, 1, \ldots$; in the case of Kupiec's TUFF test, we refer to the former.

$$LR_{TUFF}(\alpha, d_1) = -2\,log\left(\frac{\alpha(1-\alpha)^{h_1-1}}{\frac{1}{h_1}\left(1-\frac{1}{h_1}\right)^{h_1-1}}\right) \overset{asy}{\sim} \chi^2\,(1), \tag{10}$$

where h_1 denotes the time until the first failure occurs, as defined earlier.

As indicated by Dowd (1998), Evers and Rohde (2014) or Haas (2001), the TUFF test has a low power to discriminate among alternative hypotheses and, therefore, it may be difficult to observe whether the VaR model is biased or not. The TUFF test is best applied as a preliminary procedure for the frequency of excessive losses tests and may be utilised whenever the VaR violation is observed (Dowd, 1998), or there is not enough data available to perform more sophisticated tests.

3.5. Haas 2001 – Time Between Failures

Based on the intuition of the TUFF and independence tests, Haas (2001) extended the TUFF approach by including not only the time until the first failure but also an entire distribution of a time interval between VaR violations. Modelling the independence of VaR violations in the framework of the time between failures (TBF) test has the following likelihood ratio:

$$LR_{IND}^{TBF}(\alpha, s) = \sum_{i=1}^{l-1}\left(-2\,log\left(\frac{\alpha(1-\alpha)^{h_1-1}}{\frac{1}{h_1}\left(1-\frac{1}{h_1}\right)^{h_1-1}}\right)\right) \overset{asy}{\sim} \chi^2\,(l-1), \tag{11}$$

where h_i is defined as above. Note that the last duration time is being neglected, i.e. the TBF test does not take into account the time span after the last VaR violation.

When combining the likelihood ratio of Kupiec's POF test with the likelihood ratio of the TBF test, we obtain the 'Mixed Kupiec's test' with the following likelihood ratio:

$$LR_{MIX}(\alpha,n,s) + LR_{IND}^{TBF}(\alpha,s) + LR_{POF}(\alpha,n,s) \overset{asy}{\sim} \chi^2(l). \tag{12}$$

The TUFF and TBF tests are both duration-based tests, as the time interval between failures, i.e. the duration, is utilised.

3.6. Christoffersen and Pelletier 2004 – Continuous Weibull

Christoffersen and Pelletier (2004) present an alternative approach to the backtest VaR which is based on the analysis of the time between consecutive VaR violations. As defined earlier, let h_i, $i = 1, \dots, l$ represent time spans between all observable VaR violations which should be IID, because VaR violations should be independent from each other. Under the null hypothesis of the test, the VaR violation sequence process has no memory property and, thus, the no-hit distribution follows the formula:

$$f_{EXP}(h_i;\lambda) = \lambda\, exp(-\lambda h_i). \tag{13}$$

Alternatively, if the process contains the property of memory, the distribution of no-hit durations may follow the Continuous Weibull distribution:

$$f_{CW}(h_i,a,b) = a^b b h_i^{b-1} exp\,(-(ah_i)^b. \tag{14}$$

Note that $f_{CW}(h_i,a,b)|_{b=1,a=p} = f_{GAMMA}(h_i,a,b) = f_{EXP}(h_i,p)$.

The duration between VaR violations should be IID. The test is based on the fitting of the continuous Weibull distribution (alternatively the Gamma distribution) to empirical data of durations between VaR violations. The null hypothesis of the test is $H_0\colon b = 1$.

Because the $\{h_i\}_{i=1}^{l}$ may be censored ($s_1^\alpha \neq 1\ or\ s_n^\alpha \neq 1$), along with creating a duration sequence h_i, $i = 1, \dots, l$, one has to also create a flag variable denoted as c_i, $i = 1, \dots, l$, which indicates whether h_i is censored. Except the first and the last duration ($h_1$and h_l), all durations h_i are uncensored ($c_i = 0,\ i = 2, \dots, l-1$). When $s_1^\alpha = 0$ ($s_n^\alpha = 0$), then $c_0 = 1$ ($c_l = 1$).

The log-likelihood is as follows:

$$LR_{CW}(\alpha, l, \{h_i\}_{i=1}^{l}, \{c_i\}_{i=1}^{l} c_1 \log\big(1 - F_{CW}(h_1)\big) + (1 - c_1) \log\big(f_{CW}(h_1)\big) + \\ + c_l \log\big(1 - F_{CW}(h_l)\big) + (1 - c_1) \log\big(f_{CW}(h_l)\big) + \sum_{i=2}^{l-1} \log\big(f_{CW}(h_i)\big), \tag{15}$$

where $F_{CW}(\cdot)$ takes on the continuous Weibull cumulative distribution function.

3.7. Haas 2005 – Discrete Weibull

On the basis of the previous duration-based test, Haas (2005) suggests using the discrete Weibull distribution to backtest d_i, $i = 1, \dots, l-1$ instead of applying the continuous one by Christoffersen and Pelletier (2004). Since the support of time between VaR violations are natural numbers, Haas (2005) argued that the duration between violations follows the discrete Weibull distribution

$$f_{DW}(d_i, a, b) = exp[-a^b\,(d_i - 1)^b] - ex\,p\big(-a^b d_i^b\big), \tag{16}$$

where $d_i = 1$ is the time between i and $i+1$ VaR violation and $b > 0$. The null hypothesis of the correct conditional probability α corresponds to $b = 1$ and $a = -log(1-\alpha)$. The null hypotheses of independence corresponds to $b = 1$. These hypotheses can be tested by means of the likelihood ratio test.

As shown by Candelon et al. (2011), the discrete distribution test exhibits higher power than its continuous competitor test. Moreover, the discrete distribution has a more intuitive interpretation in the context of modelling integer time durations.

3.8. Krämer and Wied 2015 – the Gini coefficient

Another duration-type approach to the backtesting of Value at Risk, proposed by Krämer and Wied (2015), is based on the inequality measure of d_i (Gini-coefficient):

$$g(d_1, \dots, d_l) = l^{-2} \frac{\sum_{i,j=1}^{l} (d_i - d_j)}{2\bar{d}}, \tag{17}$$

where: $\bar{d}$ is the arithmetic average of $\{d_i\}_{i=1}^{l}$. For the geometrically distributed d_i, the Gini coefficient is $g(d) = \frac{1-\alpha}{2-\alpha}$,where $0 \le g(d) \le \frac{1}{2}$. This test rejects the independence assumption, when $g(d_1, \dots, d_l)$ becomes too large. The test statistic is as follows:

$$T = \sqrt{n}\left(l^{-2}\frac{\sum_{i,j=1}^{l}\left(d_i - d_j\right)}{2\bar{d}} - \frac{1-\frac{1}{n}}{2-\frac{l}{n}} \right). \tag{18}$$

Critical values of the statistics can be obtained by a simulation, which is an approach preferred by the authors. This observation is also confirmed by our study.

3.9. Engle and Manganelli 2004 – DQ

Engle and Manganelli (2004) introduced a test that utilises the linear regression model and links the violation in t to all past violations. This test falls into the category of independence-based tests. For the purpose of the test, the following term is constructed:

$$\widetilde{Hit_t}(\alpha) = \begin{cases} 1-\alpha, & if\ r_t < VaR_{t|t-1}(\alpha) \\ -\alpha, & if\ r_t \geq VaR_{t|t-1}(\alpha) \end{cases} \begin{cases} 1-\alpha, & if\ r_t < VaR_{t|t-1}(\alpha) \\ -\alpha, & if\ r_t \geq VaR_{t|t-1}(\alpha) \end{cases}. \tag{19}$$

Based on the above-defined $Hit_t(\alpha)$, Engle and Manganelli (2004) proposed the following linear regression model:

$$Hit_t(\alpha) = \sigma + \sum_{k=1}^{K} \beta_k Hit_{t-k}(\alpha) + \epsilon_t. \tag{20}$$

The test specification usually includes also other variables from the available information set (e.g. past returns, square of past returns, the values of VaR forecasts). Whatever the chosen specification, the null hypothesis test of conditional efficiency corresponds to testing joint nullity of coefficients β_k and σ:

$$H_0 : \sigma = \beta_k = 0, \quad \forall k = 1, \ldots, K. \tag{21}$$

The Wald statistic is used to test the nullity of these coefficients simultaneously. We denote the vector of the $K+1$ parameters in the model by $\Psi = [\sigma, \beta_1, \ldots, \beta_K]'$. Let Z be a matrix of the explanatory variables of the model. The Wald statistic (noted as DQ_{CC}) is as follows:

$$DQ_{CC} = \frac{\widehat{\Psi}' Z' Z \widehat{\Psi}}{\alpha(1-\alpha)} \overset{asy}{\sim} \chi^2(K+1). \tag{22}$$

3.10. Berkowitz 2005 – Ljung-Box

The author of another approach points out that for a practical financial setup, i.e. short time series and low percentile (e.g. within one year of observations and $\alpha = 0.01$), the duration test can be computed only in 6 out of 10 cases.

Berkowitz et al. (2011) proposed a test of spectral density of the $Hit(\alpha)$ process and also on the univariate Ljung-Box test, which makes it possible to test the absence of autocorrelation in the $Hit(\alpha)$ sequence:

$$LB(K) = T(T+2)\sum_{k=1}^{K}\frac{\hat{\rho}_k^2}{T-k} \overset{asy}{\sim} \chi^2(K), \tag{23}$$

where $\hat{\rho}_k^2$ is the empirical autocorrelation coefficient of order k of the $Hit(\alpha)$ process. It should be recalled here, as the authors emphasise, that a test has good properties when $K > 1$; in their Monte Carlo simulations $K \in \{1,5\}$.

3.11. Candelon 2011 – GMM test

The test introduced by Candelon et al. (Candelon et al., 2011) is the last one to be discussed in this study. The authors use the GMM test framework proposed by Bontemps (2008) to evaluate the assumptions of the geometric distributional in the case of the VaR forecasts backtesting. The method is based on the *J*-statistic utilising the moments defined by the orthonormal polynomials connected with the geometric distribution. From the practical point of view, this test is simple to implement, as it consists of a simple GMM moment condition test. The orthonormal polynomials of the geometric distribution are defined as follows:

$$M_{j+1}(h,\beta) = \frac{(1-\beta)(2j+1)+\beta(j-h+1)}{(j+1)\sqrt{1-\beta}} M_j(h,\beta) - \frac{j}{j+1} M_{j-1}(h,\beta), \tag{24}$$

where $M_{-1}(h,\beta) = 0$ and $M_0(h,\beta) = 1$, and h is the vector representing times between consecutive VaR violations (i.e., $h_i = v_i - v_i - 1$ defined as before). The p is the hyperparameter of this test and refers to the number of orthogonal conditions (i.e. $M(h_i,\beta)$ is the $(p,1)$ vector representing all of the $M_j(h_i,\beta)$ orthogonal conditions).

The Unconditional Coverage test statistic is as follows:

$$J_{UC}(p) = \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} M(h_i, \beta) \right)^2 \overset{asy}{\sim} \chi^2(1). \tag{25}$$

The Conditional Coverage test statistic is as follows:

$$J_{CC}(p) = \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} M(h_i, \beta) \right)^T \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} M(h_i, \beta) \right) \overset{asy}{\sim} \chi^2(p). \tag{26}$$

3.12. Other notable approaches

Several authors argued that the final conclusions on the superiority of a particular VaR model over the others largely depend on the particular quantile that is being forecasted. Considering the VaR forecasts, some authors believe that VaR should be tested on several quantiles jointly.

The literature of VaR backtests is extensive and a number of the proposed tests are significant. The other notable approaches that were not described in this paper include: Berkowitz (2001); Clements and Taylor (2003); Dumitrescu et al. (2012); Escanciano and Olmo (2011); Pajhede (2015); Pelletier and Wei (2016), and Ziggel et al. (2014).

4. Test size evaluation

This section provides the results of the size assessment of backtests described in Section 3, using the simulation and methodological framework proposed in Subsection 2.4.

The simulation analyses were based on a simulation of 10,000 violation series, each of the length equal to either 250, 500 or 1000, i.e. corresponding to one year, two years and four years, respectively, of VaR violation observations. Each simulation for a particular sample size is denoted as an instance of the problem and follows the Bernoulli distribution (as we simulated the series of violations that follow the true H_0). For each of the tests described in Section 3, based on simulated instances of the problem, we have calculated test statistics and checked whether the H_0 is rejected, assuming that the Bernoulli distribution should be rejected with the p-value threshold probability. Having obtained the empirical rejection of H_0 frequency and a theoretical rejection probability (the threshold of the p-value), we arrived at information that can be utilised in the proposed size evaluation framework described in Subsection 2.4.

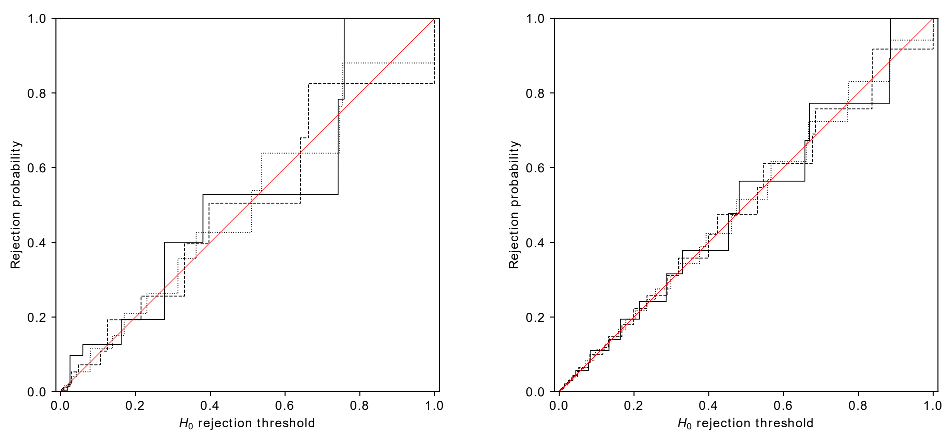
For each of the backtests, we present plots of empirical frequencies of H_0 rejections vs. theoretical rejection probabilities. The plots present the entire distribution of the p-value of the test, i.e. from 0 to 1. Usually the p-value thresholds are set to be small, e.g. 0.01 or 0.05. Those plots are easy to obtain by means of the library provided along with this article (see https://github.com/dkaszynski/VVaR).

The presented plots indicate the discrete feature of backtests for small samples. One of the findings of this study is that even though backtests may be of unbiased sizes, due to the fact that the tests' statistics can take discrete values, the comparison of the size of VaR backtesting procedures should be based on the distribution of empirical p-values.

4.1. Kupiec 1995 – Proportion of failures

The Kupiec POF test exhibits high discretisation – since the test statistic takes only a few values, the empirical rejection frequencies resemble a *step-chart*. Due to the fact that the variance of the number of VaR violations depends on the number of observations, i.e. $n\alpha(1-\alpha)$, then as the number of observations grows, the discretisation slowly decreases. Discretisation of the empirical rejection frequencies is a common issue relating to VaR backtests.

Figure 1. Size analysis for the Kupiec POF test for $\boldsymbol{\alpha} = \mathbf{0.01}$ (left plot) and $\boldsymbol{\alpha} = \mathbf{0.05}$ (right plot). Key: black $\boldsymbol{n} = \mathbf{250}$, black dashed $\boldsymbol{n} = \mathbf{500}$, grey dashed $\boldsymbol{n} = \mathbf{1000}$. The red line represents a correct-size test



Source: authors' calculation.

According to the method presented in Section 2, we have also calculated the test's size statistics – see table below.

Table 1. Size evaluation statistics – the Kupiec POF test

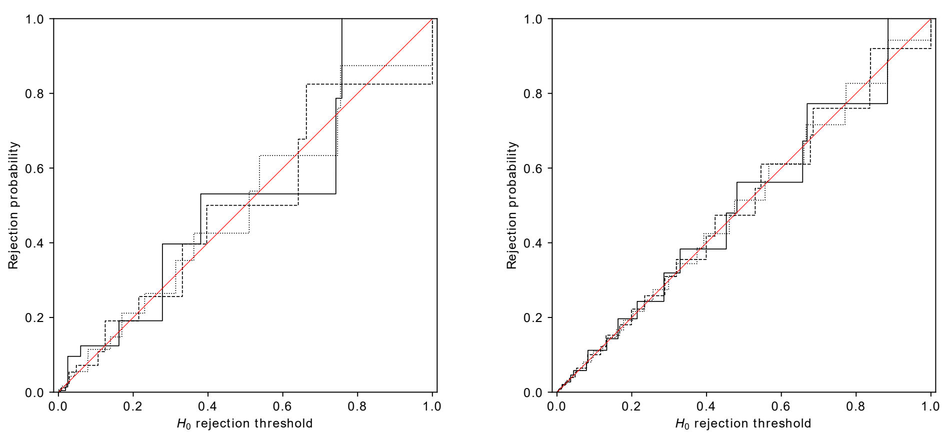
Test name	α	n	T_O	T_U	A_O	A_U	A
Kupiec-POF	0.01	250	0.64	0.36	0.08	0.08	0.08
Kupiec-POF	0.01	500	0.52	0.48	0.05	0.06	0.05
Kupiec-POF	0.01	1000	0.53	0.47	0.04	0.04	0.04
Kupiec-POF	0.05	250	0.55	0.45	0.03	0.03	0.03
Kupiec-POF	0.05	500	0.51	0.49	0.02	0.03	0.02
Kupiec-POF	0.05	1000	0.54	0.46	0.02	0.02	0.02

Source: authors' calculation.

The Kupiec POF test, as presented in the table above, exhibits small size-related issues (i.e., $A \leq 0.05$, which compared to other tests is relatively small), and along with the larger n and α, the average miss-size, measured with A, becomes smaller (see example of $\alpha = 0.05$ and $n = 1000$). To sum up, the Kupiec's POF test does not exhibit any significant size issues. In particular, it does not show any directional bias, i.e. over- or undersize features.

It is worth emphasising that the assumption relating to the number of simulations (i.e. 10,000) has been made according to the authors' expert judgment and an additional analysis of the confidence intervals. We have also recalculated the backtest for the Kupiec POF test, based on 100,000 simulations (for details, see figure below). The results demonstrate a similar shape to the baseline simulations, indicating that the discretisation problem is related to the backtest specification.

Figure 2. Additional size analysis for the Kupiec POF test for $\boldsymbol{\alpha} = \mathbf{0.01}$ (left plot) and $\boldsymbol{\alpha} = \mathbf{0.05}$ (right plot) of 100,000 simulations. Key: black $\boldsymbol{n} = \mathbf{250}$, black dashed $\boldsymbol{n} = \mathbf{500}$, grey dashed $\boldsymbol{n} = \mathbf{1000}$. The red line represents a correct-size test

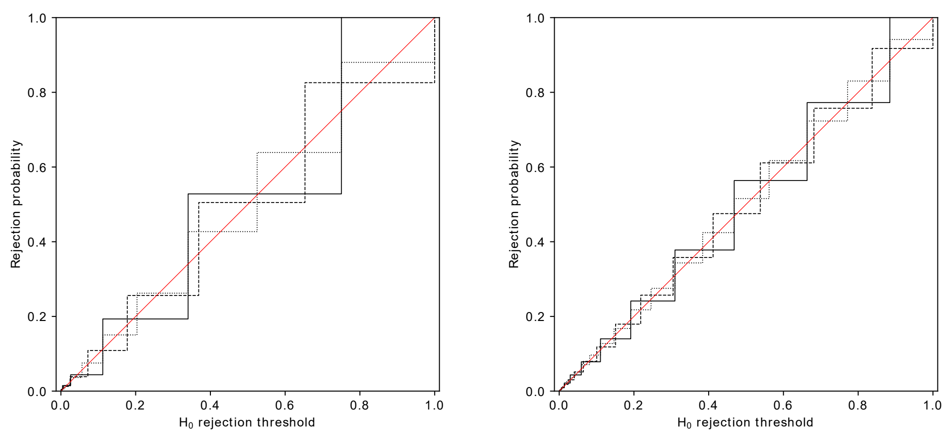


Source: authors' calculation.

4.2. Binomial test

The Binomial test, as presented in the figure below, exhibits similar or even higher discretisation issues, especially for small α and n, than the Kupiec POF test. As in the case of the test statistic taking only a few values, the empirical rejection frequencies resemble a *step-chart*. Also, due to the variance of the number of VaR, violation depends on the number of observations – as the number of observations and α grow, the discretisation gradually decreases.

Figure 3. Size analysis for the Binomial POF test for $\boldsymbol{\alpha = 0.01}$ (left plot) and $\boldsymbol{\alpha = 0.05}$ (right plot). Key: black $\boldsymbol{n = 250}$, black dashed $\boldsymbol{n = 500}$, grey dashed $\boldsymbol{n = 1000}$. The red line represents a correct-size test



Source: authors' calculation.

Table 2. Size evaluation statistics – the Binomial POF test

Test name	α	n	T_O	T_U	A_O	A_U	A
Binomial-POF	0.01	250	0.55	0.45	0.09	0.08	0.09
Binomial-POF	0.01	500	0.45	0.55	0.06	0.06	0.06
Binomial-POF	0.01	1000	0.46	0.54	0.05	0.04	0.04
Binomial-POF	0.05	250	0.51	0.49	0.04	0.04	0.04
Binomial-POF	0.05	500	0.47	0.53	0.03	0.03	0.03
Binomial-POF	0.05	1000	0.50	0.50	0.02	0.02	0.02

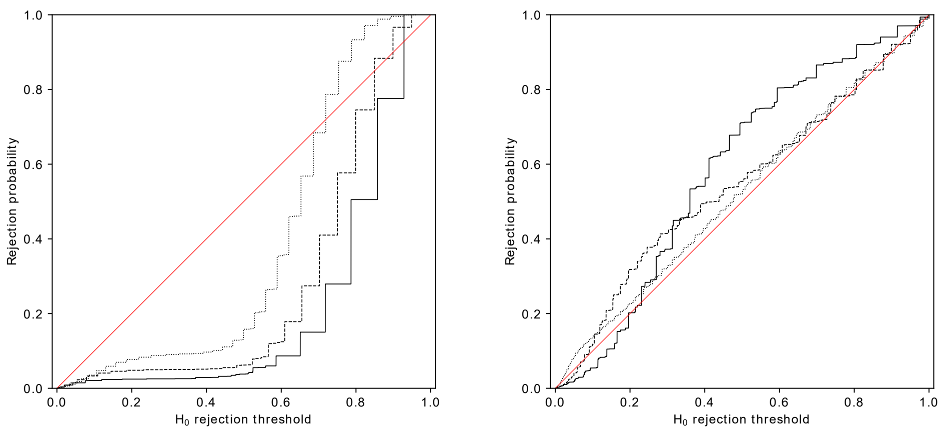
Source: authors' calculation.

The Binomial POF test, as presented in the table above, demonstrates small size-related issues (but still bigger than the Kupiec POF test), and along with the growth of n and α, the average miss-size, measured with A, becomes smaller (see example of $\alpha = 0.05$ and $n = 1000$). The above indicates that the Binomial POF test does not exhibit any significant size issues. More specifically, there is no trace of a significant directional bias, i.e. over- or undersize features.

4.3. Christoffersen 1998 tests

The Christoffersen Independence test – one of the most popular of all the backtests presented in this study – verifies whether the VaR violations tend to cluster. The p-value of the test is highly discrete, as the number of possible outcomes is finite and small. In fact, this test measures the number of cases where one VaR violation is strictly followed by another violation, which is a very rare situation in the case of small samples.

Figure 4. Size analysis for the Christoffersen Independence Coverage test for $\boldsymbol{\alpha} = \mathbf{0.01}$ (left plot) and $\boldsymbol{\alpha} = \mathbf{0.05}$ (right plot). Key: black $\boldsymbol{n} = \mathbf{250}$, black dashed $\boldsymbol{n} = \mathbf{500}$, grey dashed $\boldsymbol{n} = \mathbf{1000}$. The red line represents a correct-size test



Source: authors' calculation.

Table 3. Size evaluation statistics – Christoffersen Independence Coverage test

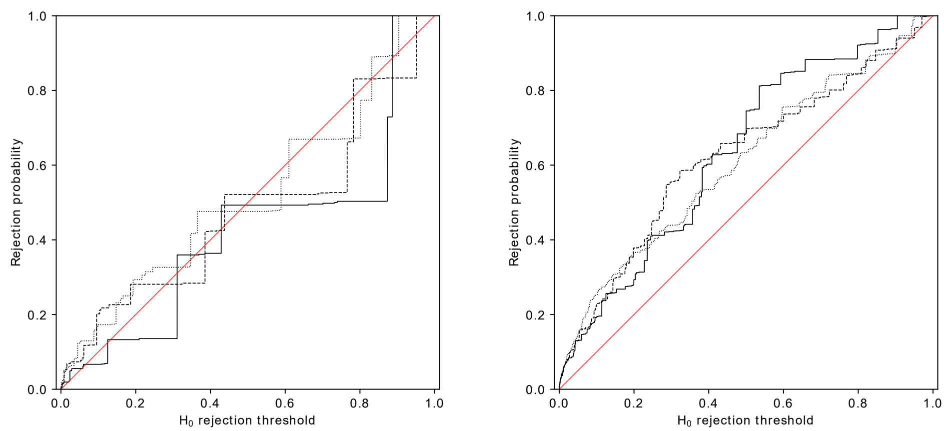
Test name	α	n	T_O	T_U	A_O	A_U	A
Christoffersen-Ind.	0.01	250	0.07	0.93	0.04	0.31	0.29
Christoffersen-Ind.	0.01	500	0.14	0.86	0.03	0.26	0.23
Christoffersen-Ind.	0.01	1000	0.28	0.72	0.09	0.19	0.16
Christoffersen-Ind.	0.05	250	0.77	0.23	0.12	0.03	0.10
Christoffersen-Ind.	0.05	500	0.81	0.19	0.06	0.01	0.05
Christoffersen-Ind.	0.05	1000	0.91	0.09	0.02	0.01	0.02

Source: authors' calculation.

The size of the test improves significantly with the increase of α. In this case, the backtest demonstrates a significantly improved distribution of the p-value.

As regards the combined test, i.e. the conditional coverage, devised by Christoffersen (1998), the results are presented below.

Figure 5. Size analysis for the Christoffersen Conditional Coverage test for $\boldsymbol{\alpha} = \mathbf{0.01}$ (left plot) and $\boldsymbol{\alpha} = \mathbf{0.05}$ (right plot). Key: black $\boldsymbol{n} = \mathbf{250}$, black dashed $\boldsymbol{n} = \mathbf{500}$, grey dashed $\boldsymbol{n} = \mathbf{1000}$. The red line represents a correct-size test



Source: authors' calculation.

Table 4. Size evaluation statistics – the Christoffersen Conditional Coverage test

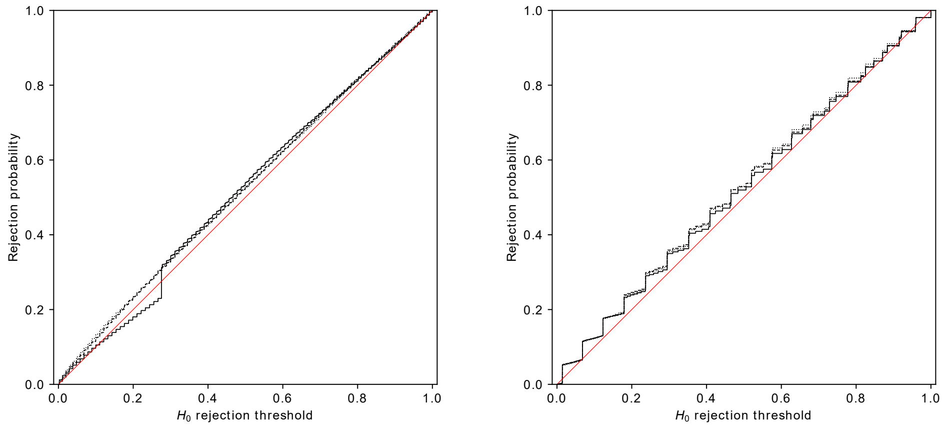
Test name	α	n	T_O	T_U	A_O	A_U	A
Christoffersen-CCoverage	0.01	250	0.29	0.71	0.04	0.13	0.10
Christoffersen-CCoverage	0.01	500	0.50	0.50	0.04	0.09	0.07
Christoffersen-CCoverage	0.01	1000	0.67	0.33	0.05	0.05	0.05
Christoffersen-CCoverage	0.05	250	1.00	0.00	0.14	0.00	0.14
Christoffersen-CCoverage	0.05	500	0.99	0.01	0.12	0.00	0.12
Christoffersen-CCoverage	0.05	1000	1.00	0.00	0.11	0.00	0.11

Source: authors' calculation.

4.4. Kupiec 1995 – Time until first failure

The Kupiec TUFF test, which, due to a significantly higher number of possible outcomes (i.e. the distribution of the possible outcome is much wider than in the POF test), exhibits less severe discretisation issues than the Kupiec POF test. Moreover, this test does not show any significant deviation, e.g. in terms of the maximal measure, from the uniform distribution, i.e. the black/grey lines lie *close* to the red line. This finding – a better size of the duration test – will be further discussed along with other examples of VaR backtests of this kind.

Figure 6. Size analysis for the Kupiec TUFF test for $\boldsymbol{\alpha} = \mathbf{0.01}$ (left plot) and $\boldsymbol{\alpha} = \mathbf{0.05}$ (right plot). Key: black $\boldsymbol{n} = \mathbf{250}$, black dashed $\boldsymbol{n} = \mathbf{500}$, grey dashed $\boldsymbol{n} = \mathbf{1000}$. The red line represents a correct-size test



Source: authors' calculation.

According to the method presented in Section 2, the test's size statistics have also been calculated (for details see the table below).

Table 5. Size evaluation statistics – the Kupiec TUFF test

Test name	α	n	T_O	T_U	A_O	A_U	A
Kupiec-TUFF	0.01	250	0.78	0.22	0.02	0.02	0.02
Kupiec-TUFF	0.01	500	0.98	0.02	0.02	0.00	0.02
Kupiec-TUFF	0.01	1000	0.99	0.01	0.02	0.00	0.02
Kupiec-TUFF	0.05	250	0.92	0.08	0.02	0.01	0.02
Kupiec-TUFF	0.05	500	0.93	0.07	0.03	0.01	0.03
Kupiec-TUFF	0.05	1000	0.94	0.06	0.03	0.01	0.03

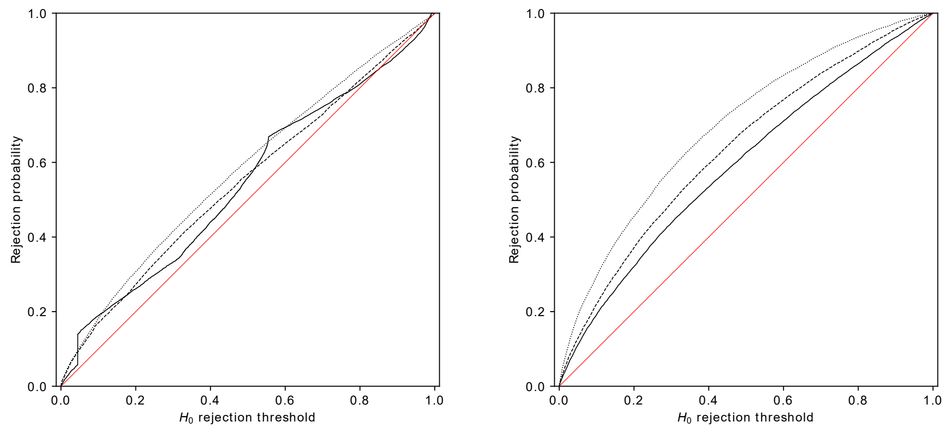
Source: authors' calculation.

The Kupiec TUFF test (which is an example of a duration test), as presented in the table above, demonstrates small size-related issues. The size of the test shows small improvement along with the increase in α and n.

4.5. Haas 2001 – Time Between Failures

The Haas's TBF test is another example of a duration approach towards VaR evaluation. As in the Kupiec TUFF, the distribution of the p-value is less discrete than it was in the case of the POF tests. Although, intuitively, the observation of more VaR violations should improve the test specification, the results suggest oversize-related issues. This is not surprising, though, as the test statistic assumes the independence of aggregated random variables, while – especially for small samples (as in our tests) – they are in fact dependent; e.g. if we observe that first-time failure is very extensive, then clearly in the subsequent instances it must be small, as we have a short test horizon.

Figure 7. Size analysis for Haas's TBF test for $\boldsymbol{\alpha} = \mathbf{0.01}$ (left plot) and $\boldsymbol{\alpha} = \mathbf{0.05}$ (right plot). Key: black $\boldsymbol{n} = \mathbf{250}$, black dashed $\boldsymbol{n} = \mathbf{500}$, grey dashed $\boldsymbol{n} = \mathbf{1000}$. The red line represents a correct-size test



Source: authors' calculation.

Table 6. Size evaluation statistics – Haas's TBF test

Test name	α	n	T_O	T_U	A_O	A_U	A
Haas-TBF	0.01	250	0.86	0.14	0.05	0.01	0.05
Haas-TBF	0.01	500	1.00	0.00	0.05	0.00	0.05
Haas-TBF	0.01	1000	1.00	0.00	0.08	0.00	0.08
Haas-TBF	0.05	250	1.00	0.00	0.09	0.00	0.09
Haas-TBF	0.05	500	1.00	0.00	0.14	0.00	0.14
Haas-TBF	0.05	1000	1.00	0.00	0.20	0.00	0.20

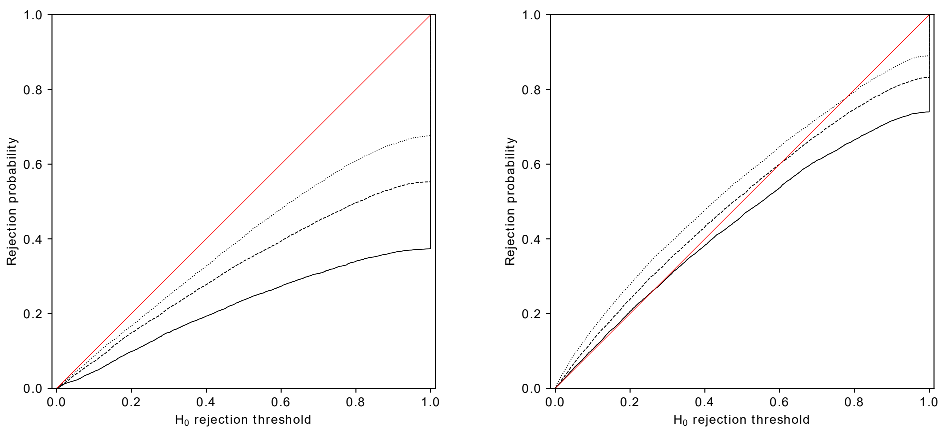
Source: authors' calculation.

As presented in the table above, the Haas TBF test exhibits relatively small size-related issues. However, with the larger α and n the test exhibits significant oversize issues.

4.6. Christoffersen and Pelletier 2004 – Continuous Weibull

The Christoffersen Continuous Weibull test is yet another instance of a duration approach. Unlike the previous examples, however, this test assumes the distribution of a duration between VaR violations, thus it falls within the category of analytical-based approaches. In terms of small VaR violation cases (e.g. α = 0.01), the p-value distribution of the tests indicated significant deviations from the uniform distribution. The distinctive *jump* on the right-hand side of the plots (in both $\alpha = 0.01$ and $\alpha = 0.05$) is caused by problems with convergence of numerical optimisation methods – in this example, the Weibull distribution parameters were calibrated using only a few examples.

Figure 8. Size analysis for the Christoffersen Continuous Weibull test for $\boldsymbol{\alpha = 0.01}$ (left plot) and $\boldsymbol{\alpha = 0.05}$ (right plot). Key: black $\boldsymbol{n = 250}$, black dashed $\boldsymbol{n = 500}$, grey dashed $\boldsymbol{n = 1000}$. The red line represents a correct size-test



Source: authors' calculation.

Table 7. Size evaluation statistics – the Christoffersen Continuous Weibull test

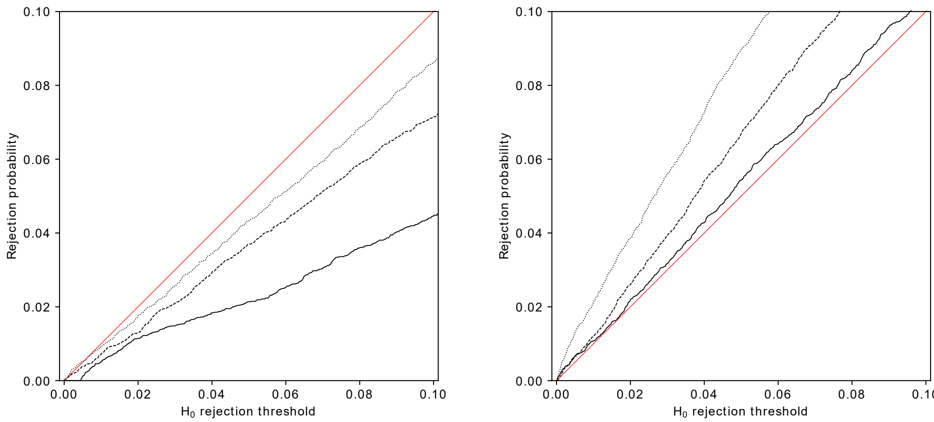
Test name	α	n	T_O	T_U	A_O	A_U	A
Christoffersen-CWeibull	0.01	250	0.00	1.00	0.00	0.28	0.28
Christoffersen-CWeibull	0.01	500	0.00	1.00	0.00	0.18	0.18
Christoffersen-CWeibull	0.01	1000	0.00	1.00	0.00	0.11	0.11
Christoffersen-CWeibull	0.05	250	0.26	0.74	0.00	0.09	0.07
Christoffersen-CWeibull	0.05	500	0.60	0.40	0.03	0.06	0.04
Christoffersen-CWeibull	0.05	1000	0.78	0.22	0.06	0.04	0.05

Source: authors' calculation.

The Christoffersen Continuous Weibull test, as duration tests in general, size of the test improves as the number of VaR violations increases.

Even though the tests appear to depart from the perfect size (i.e. red line on the plot) throughout the entire range of rejection thresholds, as mentioned earlier, the thresholds of statistical tests are usually small. In the case of the Christoffersen Continuous Weibull, the figure on the smaller range, i.e. $0 - 0.1$, is presented below. As regards the figure, the test on the threshold usually applied (for the $\alpha = 0.05$), appears to be more adequate.

Figure 9. Size analysis for the Christoffersen Continuous Weibull test (smaller range) for $\boldsymbol{\alpha} = \mathbf{0.01}$ (left plot) and $\boldsymbol{\alpha} = \mathbf{0.05}$ (right plot). Key: black $\boldsymbol{n} = \mathbf{250}$, black dashed $\boldsymbol{n} = \mathbf{500}$, grey dashed $\boldsymbol{n} = \mathbf{1000}$. The red line represents a correct-size test

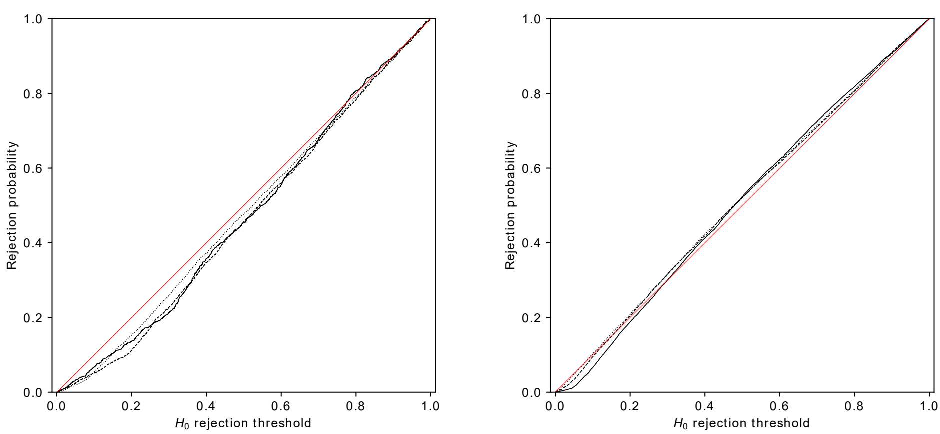


Source: authors' calculation.

4.7. Haas 2005 – Discrete Weibull

The discussion of duration approaches to the VaR evaluation concludes with the Haas Discrete Weibull test. As far as the low VaR violation cases (small α and n) are concerned, the p-value of this test is highly discrete. For the larger VaR violation cases, this test demonstrates a small deviation from the correct size.

Figure 10. Size analysis for the Haas Discrete Weibull test for $\boldsymbol{\alpha} = \mathbf{0.01}$ (left plot) and $\boldsymbol{\alpha} = \mathbf{0.05}$ (right plot). Key: black $\boldsymbol{n} = \mathbf{250}$, black dashed $\boldsymbol{n} = \mathbf{500}$, grey dashed $\boldsymbol{n} = \mathbf{1000}$. The red line represents a correct-size test



Source: authors' calculation.

Table 8. Size evaluation statistics – the Haas Discrete Weibull test

Test name	α	n	T_O	T_U	A_O	A_U	A
Haas-DWeibull	0.01	250	0.17	0.83	0.00	0.04	0.03
Haas-DWeibull	0.01	500	0.02	0.98	0.00	0.04	0.04
Haas-DWeibull	0.01	1000	0.04	0.96	0.00	0.03	0.02
Haas-DWeibull	0.05	250	0.70	0.30	0.02	0.02	0.02
Haas-DWeibull	0.05	500	0.84	0.16	0.01	0.01	0.01
Haas-DWeibull	0.05	1000	0.92	0.08	0.01	0.00	0.01

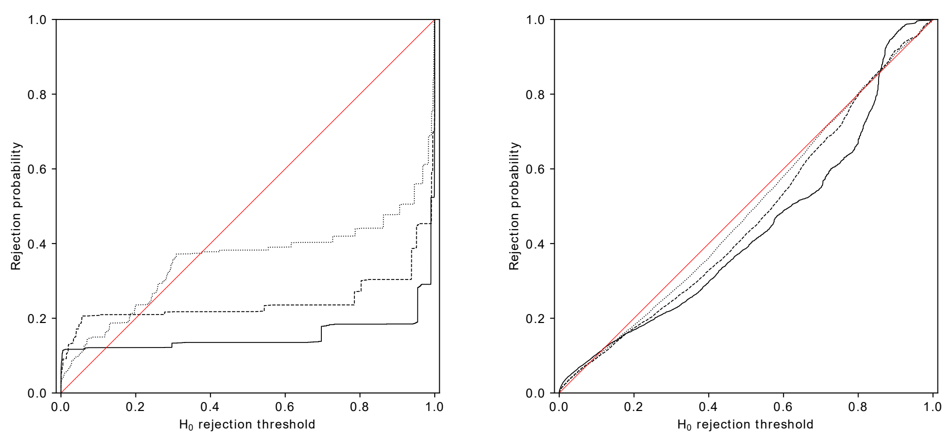
Source: authors' calculation.

Due to the approach applied in the test, it is usually compared with its continuous version, i.e. the Christoffersen Continuous Weibull. Regarding those two specifications, the discrete version preserves better size properties taking into account the size evaluation statistics.

4.8. Engle and Manganelli 2004 – DQ

The Engle and Manganelli backtest verifies whether VaR violations can be explained by a linear regression of previous violations (in fact, this test can also take into account other exogenous variables).

Figure 11. Size analysis for the Engle DQ test for $\boldsymbol{\alpha = 0.01}$ (left plot) and $\boldsymbol{\alpha = 0.05}$ (right plot). Key: black $\boldsymbol{n = 250}$, black dashed $\boldsymbol{n = 500}$, grey dashed $\boldsymbol{n = 1000}$. The red line represents a correct-size test



Source: authors' calculation.

Table 9. Size evaluation statistics – the Engle DQ test

Test name	α	n	T_O	T_U	A_O	A_U	A
Engle-DQ	0.01	250	0.12	0.88	0.06	0.40	0.36
Engle-DQ	0.01	500	0.21	0.79	0.08	0.35	0.29
Engle-DQ	0.01	1000	0.38	0.62	0.04	0.25	0.17
Engle-DQ	0.05	250	0.26	0.74	0.03	0.09	0.08
Engle-DQ	0.05	500	0.25	0.75	0.01	0.05	0.04
Engle-DQ	0.05	1000	0.27	0.73	0.00	0.02	0.02

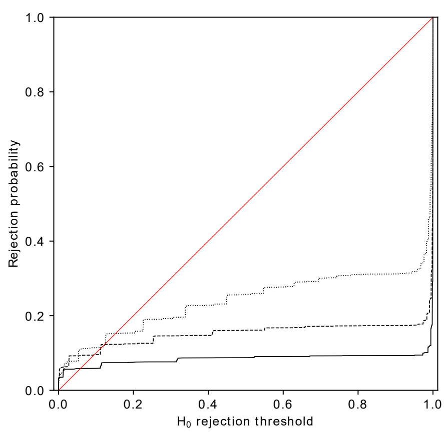
Source: authors' calculation.

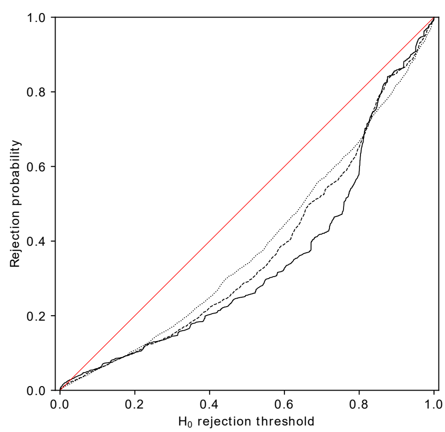
The p-value of the test relating to low VaR violation cases is highly deviated from the uniform distribution. As far as the high VaR violation cases are concerned, the size of the tests significantly improves.

4.9. Berkowitz 2005 – Ljung-Box

Berkowitz's Ljung-Box backtest verifies whether VaR violations are autocorrelated with the degree of k (in this experiment, a $k = 5$ set is implemented).

Figure 12. Size analysis for Berkowitz's Ljung-Box test for $\boldsymbol{\alpha} = \mathbf{0.01}$ (left plot) and $\boldsymbol{\alpha} = \mathbf{0.05}$ (right plot). Key: black $\boldsymbol{n} = \mathbf{250}$, black dashed $\boldsymbol{n} = \mathbf{500}$, grey dashed $\boldsymbol{n} = \mathbf{1000}$. The red line represents a correct-size test





Source: authors' calculation.

Table 10. Size evaluation statistics – Berkowitz's Ljung-Box test

Test name	α	n	T_O	T_U	A_O	A_U	A
Berkowitz-BoxLjung	0.01	250	0.06	0.94	0.02	0.44	0.42
Berkowitz-BoxLjung	0.01	500	0.11	0.89	0.03	0.39	0.36
Berkowitz-BoxLjung	0.01	1000	0.14	0.86	0.03	0.31	0.27

Table 11. Size evaluation statistics – Berkowitz's Ljung-Box test (cont.)

Test name	α	n	T_O	T_U	A_O	A_U	A
Berkowitz-BoxLjung	0.05	250	0.03	0.97	0.01	0.16	0.15
Berkowitz-BoxLjung	0.05	500	0.02	0.98	0.00	0.13	0.13
Berkowitz-BoxLjung	0.05	1000	0.01	0.99	0.00	0.11	0.11

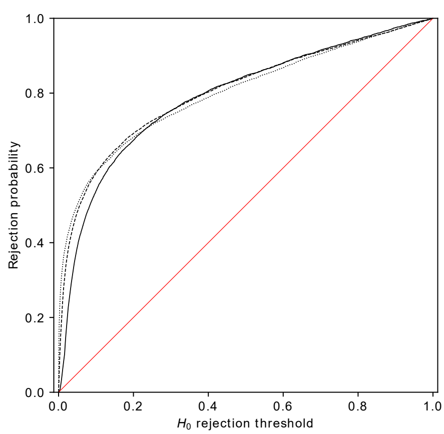
Source: authors' calculation.

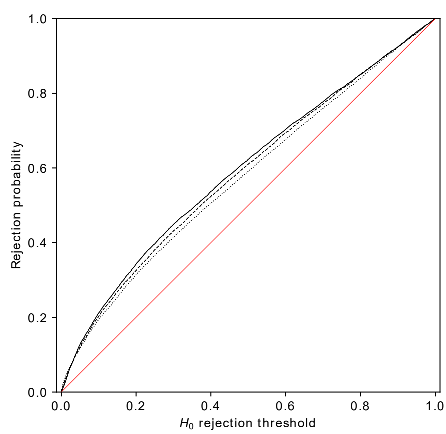
The p-value of the test for the low VaR violation cases is highly deviated from the uniform distribution. Concerning the high VaR violation cases, the size of the tests improves, but, nevertheless, remains below the correct value.

4.10. Krämer and Wied 2015 – Gini coefficient

The Krämer and Wied backtest is a duration-type test, but contrary to the previous ones, it is based on the Gini coefficient.

Figure 13. Size analysis of Krämer's Gini coefficient test for $\boldsymbol{\alpha} = \mathbf{0.01}$ (left plot) and $\boldsymbol{\alpha} = \mathbf{0.05}$ (right plot). Key: black $\boldsymbol{n} = \mathbf{250}$, black dashed $\boldsymbol{n} = \mathbf{500}$, grey dashed $\boldsymbol{n} = \mathbf{1000}$. The red line represents a correct-size test





Source: authors' calculation.

Table 12. Size evaluation statistics – Krämer's Gini coefficient test

Test name	α	n	T_O	T_U	A_O	A_U	A
Kramer-GINI	0.01	250	1.00	0.00	0.29	0.00	0.29
Kramer-GINI	0.01	500	1.00	0.00	0.30	0.00	0.30
Kramer-GINI	0.01	1000	1.00	0.00	0.30	0.00	0.30
Kramer-GINI	0.05	250	1.00	0.00	0.09	0.00	0.09
Kramer-GINI	0.05	500	1.00	0.00	0.09	0.00	0.09
Kramer-GINI	0.05	1000	1.00	0.00	0.07	0.00	0.07

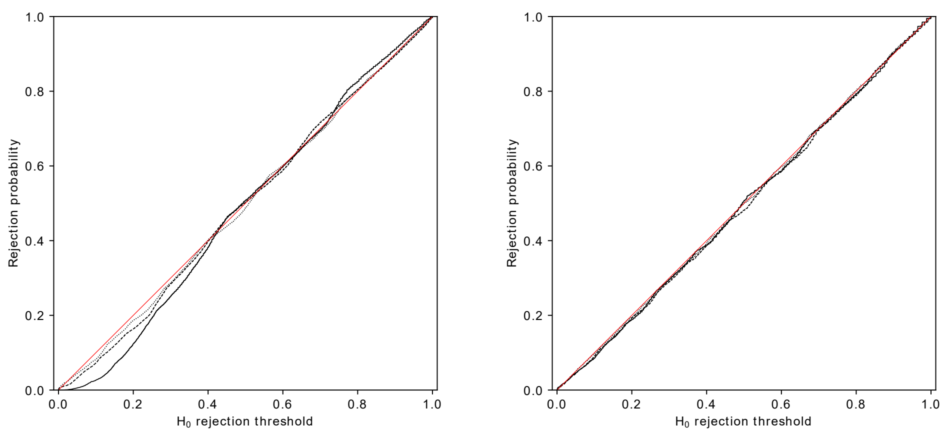
Source: authors' calculation.

As the authors emphasise in the article (Krämer and Wied 2015), simulation is the preferable approach to size evaluation. Based on our calculation (assuming asymptotic distribution of a test's statistics), the test for low VaR violation instances proves strongly oversized. This problem is much smaller in the case of the high-volume VaR violations scenarios.

4.11. Candelon 2011 – GMM test

The Candelon backtest is a duration-type test based on the GMM approach, which assumes that the distribution of failures is geometric. The size-assessment results of the unconditional coverage variant of the GMM test is presented below. As Fig. 14 and the results from Table 12 indicate, the test shows a low level of size-related problems in comparison to other approaches.

Figure 14. Size analysis of the Candelon GMM Unconditional Coverage test for $\boldsymbol{\alpha} = \mathbf{0.01}$ (left plot) and $\boldsymbol{\alpha} = \mathbf{0.05}$ (right plot). Key: black $\boldsymbol{n} = \mathbf{250}$, black dashed $\boldsymbol{n} = \mathbf{500}$, grey dashed $\boldsymbol{n} = \mathbf{1000}$. The red line represents a correct-size test



Source: authors' calculation.

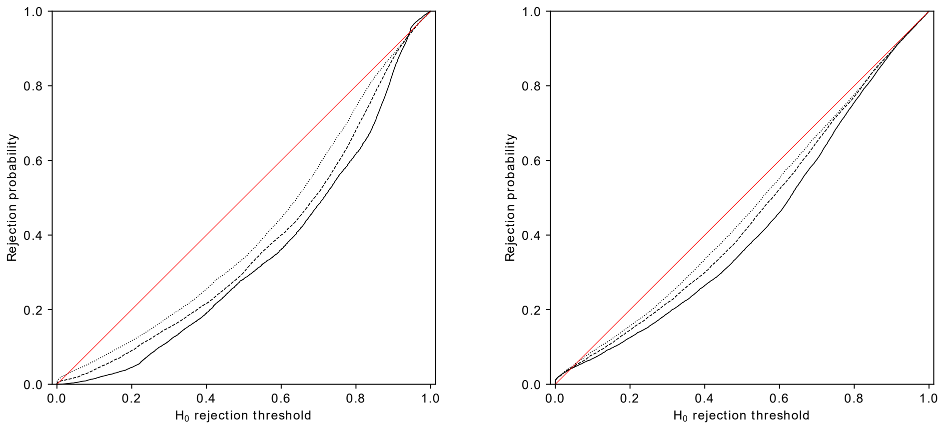
Table 13. Size evaluation statistics – the Candelon GMM Unconditional Coverage test

Test name	α	n	T_O	T_U	A_O	A_U	A
Candelon-GMM-UC	0.01	250	0.42	0.58	0.01	0.04	0.03
Candelon-GMM-UC	0.01	500	0.31	0.69	0.01	0.02	0.01
Candelon-GMM-UC	0.01	1000	0.24	0.76	0.00	0.01	0.01
Candelon-GMM-UC	0.05	250	0.20	0.79	0.00	0.01	0.01
Candelon-GMM-UC	0.05	500	0.09	0.91	0.00	0.01	0.01
Candelon-GMM-UC	0.05	1000	0.28	0.72	0.00	0.00	0.00

Source: authors' calculation.

In terms of the Conditional Coverage variant of that test, the simulation results are shown in the figure / table below.

Figure 15. Size analysis of the Candelon GMM Conditional Coverage test for $\boldsymbol{\alpha} = \mathbf{0.01}$ (left plot) and $\boldsymbol{\alpha} = \mathbf{0.05}$ (right plot). Key: black $\boldsymbol{n} = \mathbf{250}$, black dashed $\boldsymbol{n} = \mathbf{500}$, grey dashed $\boldsymbol{n} = \mathbf{1000}$. The red line represents a correct-size test



Source: authors' calculation.

Table 14. Size evaluation statistics – the Candelon GMM Conditional Coverage test

Test name	α	n	T_O	T_U	A_O	A_U	A
Candelon-GMM-CC	0.01	250	0.06	0.94	0.01	0.16	0.16
Candelon-GMM-CC	0.01	500	0.02	0.98	0.00	0.13	0.12
Candelon-GMM-CC	0.01	1000	0.05	0.95	0.00	0.09	0.09
Candelon-GMM-CC	0.05	250	0.04	0.96	0.01	0.08	0.08
Candelon-GMM-CC	0.05	500	0.05	0.95	0.01	0.05	0.05
Candelon-GMM-CC	0.05	1000	0.08	0.92	0.00	0.04	0.04

Source: authors' calculation.

5. Conclusions

The presented methodology and size plots indicate the discrete nature of backtests for small samples. One of the findings demonstrates that even though backtests may have unbiased sizes, the comparison of the size of VaR backtesting procedures should be based on the distribution of empirical p-values due to the fact that tests' statistics can take discrete values. The authors' intention was to strongly emphasise the relatively significant discretisation of POF tests, which is less severe in the case of duration-based tests. This effect results from the number of possible (and probable) values of the tests' inputs. As regards frequency-based tests, for small samples the test statistic is usually limited to only a few values, and in effect a few test outcomes –

the p-values. As far as duration-based tests are concerned, the numbers of possible test outcomes are much broader, which results in a less discrete p-value cumulative distribution.

Considering exclusively average-size deviation from the correct size in the case of small samples, duration-based tests appear to be superior, especially the Kupiec TUFF and the Haas DWeibull. On the other hand, the Christoffersen's Conditional Coverage test demonstrates a significant deviation from the correct size – especially when considering a low, $\alpha = 0.01$ level. The Christoffersen's Continuous Weibull is another example of a backtest which shows a significant deviation from the correct test size, in particular for the $\alpha = 0.01$ level.

In order to facilitate the comparison of all the analysed tests, a summary of the backtests' size assessment is presented in Appendix A. In addition to the measures proposed in Section 2.4, i.e. measures for the assessment of the size of backtests, a comparative measure of discretisation levels of individual tests – a D measure – is also included. The applied D measure is the number of the unique p-values in the range of 0.01–0.1, i.e. in the range of H_0 rejection threshold which is typically encountered in practice. The results indicate that the tests with the highest levels of discretisation ($D \geq 50$), along with the smallest deviation from the correct size ($A \leq 0.05$) for small samples, i.e. $n = 250$ and $\alpha = 0.01$, are the Candelon GMM (Unconditional Coverage variant), the Haas Discrete Weibull, and the Haas TBF tests. In addition, the results confirm the intuitive observation that the level of discretisation (i.e. the number of unique p-values) decreases along with the increase of n, i.e. the length of the time window at which VaR models are validated. The authors would also like to point out that each of the backtests is designed to measure a particular type of a deviation/problem. Bearing that in mind, it is recommended that the results presented in this paper be used to compare backtests with their benchmarks. For instance, in terms of duration-based test, for small samples (i.e., $n = 250$ and $\alpha = 0.01$) the best backtest is Candelon-GM, even though the Kupiec TUFF tests have a lower A, they also have a small number of unique p-values denoted by D. The summary table in Appendix A is sorted by the average deviation A.

We are aware that when selecting a test for VaR backtesting it is essential for it to be of a large power. However, the usage of an ill-sized test leads to unreliable results. As a consequence, a proper size of the test should be a *screening criterion* applied prior to using the test in practice. This issue is illustrated by, e.g., the fact that the Christoffersen Independence test remains a popular and widely-used test in VaR diagnostics, even though it significantly deviates from the correct size (as the results of our analysis show). In practice, the analysis of the power of the considered tests should be performed along with the consideration of the proper size of the test. However, regarding VaR backtesting, it is challenging to provide a similar analysis to the one we presented for test sizes, as there are no equally-powerful VaR backtests (different tests are sensitive to different violations of the assumptions). Therefore, the choice of an appropriate backtest should depend on the kind of deviation the analyst strives most to detect (alternatively, using several tests in combination may be considered, provided that all of them are of an acceptable quality in terms of their size).

The practical suggestion resulting from this study is that instead of using theoretical formulas for p-values of the discussed tests (that are only asymptotic), which is common practice, it is advisable to produce a simulated distribution of the statistics for a given test (knowing α and n), and compute the p-values against such a distribution. This procedure makes it possible, at least to some extent, to mitigate the risk of applying over- or undersized tests in the case of the limited sample size n and small α level. Unfortunately, such a simulation does not remove the discretisation effect in tests which display such features.

References

Altman, D. G. (1991). *Practical Statistics for Medical Research*. London: Chapman & Hall/CRC. https://www.scribd.com/doc/273959883/Douglas-G-Altman-Practical-Statistics-for-Medical-Research-Chapman-Hall-CRC-1991

BCBS. (1996). *Supervisory framework for the use of "backtesting" in conjunction with the internal models approach to market risk capital requirements.* Basel: Basle Committee on Banking Supervision. https://www.bis.org/publ/bcbs22.pdf

BCBS. (2009). *Revisions to the Basel II market risk framework*. Basel: Bank for International Settlements. https://www.bis.org/publ/bcbs158.pdf

Berkowitz, J. (2001). Testing Density Forecasts, with Applications to Risk Management. *Journal of Business & Economic Statistics*, *19*(4), 465–474. https://doi.org/10.1198/07350010152596718

Berkowitz, J., Christoffersen, P., Pelletier, D. (2011). Evaluating Value-at-Risk Models with Desk--Level Data. *Management Science*, *57*(12), 2213–2227. https://doi.org/10.1287/mnsc.1080.0964

Bontemps, C. (2014). *Moment-based tests for discrete distributions*. (IDEI Working Paper, n. 772). http://idei.fr/sites/default/files/medias/doc/by/bontemps/discrete-15oct2014.pdf

Campbell, S. D. (2006). A review of backtesting and backtesting procedures. *Journal of Risk*, *9*(2), 1–17. http://dx.doi.org/10.21314/JOR.2007.146

Candelon, B., Colletaz, G., Hurlin, C., Tokpavi, S. (2011). Backtesting Value-at-Risk: a GMM Duration-Based Test. *Journal of Financial Econometrics*, *9*(2), 314–343. https://doi.org/10.1093/jjfinec/nbq025

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International economic review*, *39*(4), 841–862. https://doi.org/10.2307/2527341

Christoffersen, P., Pelletier, D. (2004). Backtesting Value-at-Risk: A Duration-Based Approach. *Journal of Financial Econometrics*, *2*(1), 84–108. https://doi.org/10.1093/jjfinec/nbh004

Clements, M. P., Taylor, N. (2003). Evaluating interval forecasts of high-frequency financial data. *Journal of Applied Econometrics*, *18*(4), 445–456. https://doi.org/10.1002/jae.703

Dowd, K. (1998). *Beyond Value at Risk: The New Science of Risk Management.* Chichester: John Wiley & Sons.

Dumitrescu, E. I., Hurlin, C., Pham, V. (2012). Backtesting Value-at-Risk: From Dynamic Quantile to Dynamic Binary Tests. *Finance*, *33*(1), 79–112. https://www.cairn-int.info/journal-finance-2012-1-page-79.htm?WT.tsrc=cairnPdf#

Engle, R. F., Manganelli, S. (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics*, *22*(4), 367–381. https://doi.org/10.1198/073500104000000370

Escanciano, J. C., Olmo, J. (2011). Robust Backtesting Tests for Value-at-risk Models. *Journal of Financial Econometrics*, *9*(1), 132–161. https://doi.org/10.1093/jjfinec/nbq021

Everitt, B. S. (Ed.). (2006). *The Cambridge Dictionary of Statistics* (3rd edition). Cambridge: Cambridge University Press.

Evers, C., Rohde, J. (2014). *Model Risk in Backtesting Risk Measures* (HEP Discussion Paper No. 529). http://diskussionspapiere.wiwi.uni-hannover.de/pdf_bib/dp-529.pdf

Haas, M. (2001). *New Methods in Backtesting*. https://www.ime.usp.br/~rvicente/risco/haas.pdf

Haas, M. (2005). Improved duration-based backtesting of value-at-risk. *Journal of Risk*, *8*(2), 17–38. http://dx.doi.org/10.21314/JOR.2006.128

Hurlin, C. (29.04.2013). *Backtesting Value-at-Risk Models*. Séminaire Validation des Modèles Financiers, University of Orléans. https://www.univ-orleans.fr/deg/masters/ESA/CH/Slides_Seminaire_ Validation.pdf

Hurlin, C., Tokpavi, S. (2006). Backtesting Value-at-Risk Accuracy: A Simple New Test. *Journal of Risk*, *9*(2), 19–37. http://dx.doi.org/10.21314/JOR.2007.148

Jorion, P. (2007). *Value at Risk: The New Benchmark for Managing Financial Risk* (3rd edition). New York: The McGraw-Hill Companies. https://www.academia.edu/8519246/Philippe_Jorion_Value_at_Risk_The_New_Benchmark_for_Managing_Financial_Risk_3rd_Ed_2007

Jorion, P. (2010). *Financial Risk Manager Handbook: FRM Part I/Part II*. Hoboken: John Wiley & Sons.

Krämer, W., Wied, D. (2015). A simple and focused backtest of value at risk. *Economics Letters*, *137*, 29–31. https://doi.org/10.1016/j.econlet.2015.10.028

Kupiec, P. H. (1995). Techniques For Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives*, *3*(2), 73–84. https://doi.org/10.3905/jod.1995.407942

Lopez, J. A. (1998). Methods for Evaluating Value-at-Risk Estimates. *Economic Policy Review*, *4*(3), 119–124. https://www.newyorkfed.org/medialibrary/media/research/epr/1998/EPRvol4no3.pdf

Małecka, M. (2014). Duration-Based Approach to VaR Independence Backtesting. *Statistics in Transition new series*, *15*(4), 627–636. http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.ekon-element-000171338797

Murdoch, D. J., Tsai, Y. L., Adcock, J. (2008). P-Values are Random Variables. *The American Statistician*, *62*(3), 242–245. https://doi.org/10.1198/000313008X332421

Nieto, M. R., Ruiz, E. (2016). Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting*, *32*(2), 475–501. https://doi.org/10.1016/j.ijforecast.2015.08.003

Pajhede, T. (2015). *Backtesting Value-at-Risk: A Generalized Markov Framework* (Univ. of Copenhagen Dept. of Economics Discussion Paper No. 15–18). http://dx.doi.org/10.2139/ssrn.2693504

Pelletier, D., Wei, W. (2016). The Geometric-VaR Backtesting Method. *Journal of financial econometrics*, *14*(4), 725–745. https://pdfs.semanticscholar.org/644b/ced159cafb17e48a24ffa36bbaf2ac776f18.pdf?_ga=2.260873924.817706028.1606139341-1947910505.1605732686

Zhang, Y., Nadarajah, S. (2017). A review of backtesting for value at risk. *Communications in Statistics – Theory and Methods*, *47*(15), 3616–3639. https://doi.org/10.1080/03610926.2017.1361984

Ziggel, D., Berens, T., Weiß, G. N. F., Wied, D. (2014). A new set of improved Value-at-Risk backtests. *Journal of Banking & Finance*, *48*, 29–41. https://doi.org/10.1016/j.jbankfin.2014.07.005

Appendix A

Table 15. Summary table – an assessment of size of VaR backtests ($\boldsymbol{p}$-values ranging from 0 to 1)

Test	α	n	T_O	T_U	A_O	A_U	A	D
Kupiec-TUFF	0.01	250	0.78	0.22	0.02	0.02	0.02	10
Candelon-GMM-UC	0.01	250	0.42	0.58	0.01	0.04	0.03	187
Haas-DWeibull	0.01	250	0.17	0.83	0.00	0.04	0.03	69
Haas-TBF	0.01	250	0.86	0.14	0.05	0.01	0.05	834
Kupiec-POF	0.01	250	0.64	0.36	0.08	0.08	0.08	3
Binomial-POF	0.01	250	0.55	0.45	0.09	0.08	0.09	1
Christoffersen-CCoverage	0.01	250	0.29	0.71	0.04	0.13	0.10	11
Candelon-GMM-CC	0.01	250	0.06	0.94	0.01	0.16	0.16	135
Christoffersen-CWeibull	0.01	250	0.00	1.00	0.00	0.28	0.28	357
Christoffersen-Ind.	0.01	250	0.07	0.93	0.04	0.31	0.29	9
Kramer-GINI	0.01	250	1.00	0.00	0.29	0.00	0.29	2,938
Engle-DQ	0.01	250	0.12	0.88	0.06	0.40	0.36	27
Berkowitz-BoxLjung	0.01	250	0.06	0.94	0.02	0.44	0.42	77
Candelon-GMM-UC	0.01	500	0.31	0.69	0.01	0.02	0.01	428
Kupiec-TUFF	0.01	500	0.98	0.02	0.02	0.00	0.02	104
Haas-DWeibull	0.01	500	0.02	0.98	0.00	0.04	0.04	123
Haas-TBF	0.01	500	1.00	0.00	0.05	0.00	0.05	1,328
Kupiec-POF	0.01	500	0.52	0.48	0.05	0.06	0.05	3
Binomial-POF	0.01	500	0.45	0.55	0.06	0.06	0.06	4
Christoffersen-CCoverage	0.01	500	0.50	0.50	0.04	0.09	0.07	20
Candelon-GMM-CC	0.01	500	0.02	0.98	0.00	0.13	0.12	285
Christoffersen-CWeibull	0.01	500	0.00	1.00	0.00	0.18	0.18	637
Christoffersen-Ind.	0.01	500	0.14	0.86	0.03	0.26	0.23	11
Engle-DQ	0.01	500	0.21	0.79	0.08	0.35	0.29	490
Kramer-GINI	0.01	500	1.00	0.00	0.30	0.00	0.30	3,355
Berkowitz-BoxLjung	0.01	500	0.11	0.89	0.03	0.39	0.36	148
Candelon-GMM-UC	0.01	1000	0.24	0.76	0.00	0.01	0.01	546
Kupiec-TUFF	0.01	1000	0.99	0.01	0.02	0.00	0.02	149
Haas-DWeibull	0.01	1000	0.04	0.96	0.00	0.03	0.02	263
Kupiec-POF	0.01	1000	0.53	0.47	0.04	0.04	0.04	6
Binomial-POF	0.01	1000	0.46	0.54	0.05	0.04	0.04	6
Christoffersen-CCoverage	0.01	1000	0.67	0.33	0.05	0.05	0.05	29
Haas-TBF	0.01	1000	1.00	0.00	0.08	0.00	0.08	1,498
Candelon-GMM-CC	0.01	1000	0.05	0.95	0.00	0.09	0.09	440
Christoffersen-CWeibull	0.01	1000	0.00	1.00	0.00	0.11	0.11	769
Christoffersen-Ind.	0.01	1000	0.28	0.72	0.09	0.19	0.16	20
Engle-DQ	0.01	1000	0.38	0.62	0.04	0.25	0.17	719
Berkowitz-BoxLjung	0.01	1000	0.14	0.86	0.03	0.31	0.27	335
Kramer-GINI	0.01	1000	1.00	0.00	0.30	0.00	0.30	2518

Table 16. Summary table – an assessment of size of VaR backtests (p-values ranging from 0 to 1) (cont.)

Test	α	n	T_O	T_U	A_O	A_U	A	D
Candelon-GMM-UC	0.05	250	0.20	0.79	0.00	0.01	0.01	362
Haas-DWeibull	0.05	250	0.70	0.30	0.02	0.02	0.02	638
Kupiec-TUFF	0.05	250	0.92	0.08	0.02	0.01	0.02	49
Kupiec-POF	0.05	250	0.55	0.45	0.03	0.03	0.03	7
Binomial-POF	0.05	250	0.51	0.49	0.04	0.04	0.04	6
Candelon-GMM-CC	0.05	250	0.04	0.96	0.01	0.08	0.08	474
Christoffersen-CWeibull	0.05	250	0.26	0.74	0.00	0.09	0.07	932
Engle-DQ	0.05	250	0.26	0.74	0.03	0.09	0.08	752
Haas-TBF	0.05	250	1.00	0.00	0.09	0.00	0.09	1,643
Kramer-GINI	0.05	250	1.00	0.00	0.09	0.00	0.09	1,828
Christoffersen-Ind.	0.05	250	0.77	0.23	0.12	0.03	0.10	49
Christoffersen-CCoverage	0.05	250	1.00	0.00	0.14	0.00	0.14	69
Berkowitz-BoxLjung	0.05	250	0.03	0.97	0.01	0.16	0.15	397
Candelon-GMM-UC	0.05	500	0.09	0.91	0.00	0.01	0.01	502
Haas-DWeibull	0.05	500	0.84	0.16	0.01	0.01	0.01	913
Kupiec-POF	0.05	500	0.51	0.49	0.02	0.03	0.02	9
Kupiec-TUFF	0.05	500	0.93	0.07	0.03	0.01	0.03	46
Binomial-POF	0.05	500	0.47	0.53	0.03	0.03	0.03	8
Engle-DQ	0.05	500	0.25	0.75	0.01	0.05	0.04	745
Christoffersen-CWeibull	0.05	500	0.60	0.40	0.03	0.06	0.04	1,161
Christoffersen-Ind.	0.05	500	0.81	0.19	0.06	0.01	0.05	89
Candelon-GMM-CC	0.05	500	0.05	0.95	0.01	0.05	0.05	600
Kramer-GINI	0.05	500	1.00	0.00	0.09	0.00	0.09	1,702
Christoffersen-CCoverage	0.05	500	0.99	0.01	0.12	0.00	0.12	110
Berkowitz-BoxLjung	0.05	500	0.02	0.98	0.00	0.13	0.13	456
Haas-TBF	0.05	500	1.00	0.00	0.14	0.00	0.14	1,869
Candelon-GMM-UC	0.05	1000	0.28	0.72	0.00	0.00	0.00	588
Haas-DWeibull	0.05	1000	0.92	0.08	0.01	0.00	0.01	955
Engle-DQ	0.05	1000	0.27	0.73	0.00	0.02	0.02	788
Kupiec-POF	0.05	1000	0.54	0.46	0.02	0.02	0.02	13
Binomial-POF	0.05	1000	0.50	0.50	0.02	0.02	0.02	12
Christoffersen-Ind.	0.05	1000	0.91	0.09	0.02	0.01	0.02	162
Kupiec-TUFF	0.05	1000	0.94	0.06	0.03	0.01	0.03	44
Christoffersen-CWeibull	0.05	1000	0.78	0.22	0.06	0.04	0.05	1,375
Candelon-GMM-CC	0.05	1000	0.08	0.92	0.00	0.04	0.04	669
Kramer-GINI	0.05	1000	1.00	0.00	0.07	0.00	0.07	1,564
Christoffersen-CCoverage	0.05	1000	1.00	0.00	0.11	0.00	0.11	208
Berkowitz-BoxLjung	0.05	1000	0.01	0.99	0.00	0.11	0.11	478
Haas-TBF	0.05	1000	1.00	0.00	0.20	0.00	0.20	2,389

Source: authors' calculation.

Appendix B

Table 17. Summary table – an assessment of size of VaR backtests (p-values ranging from 0 to 0.1)

Test	α	n	T_O	T_U	A_O	A_U	A	D
Kupiec-TUFF	0.01	250	0.95	0.05	0.01	0.00	0.01	10
Christoffersen-CCoverage	0.01	250	0.46	0.54	0.01	0.01	0.01	11
Haas-DWeibull	0.01	250	0.00	1.00	0.00	0.01	0.01	82
Binomial-POF	0.01	250	0.27	0.73	0.01	0.03	0.02	1
Berkowitz-BoxLjung	0.01	250	0.62	0.38	0.02	0.02	0.02	90
Christoffersen-CWeibull	0.01	250	0.00	1.00	0.00	0.02	0.02	412
Candelon-GMM-UC	0.01	250	0.00	1.00	0.00	0.04	0.04	169
Kupiec-POF	0.01	250	0.75	0.25	0.05	0.01	0.04	3
Christoffersen-ICoverage	0.01	250	0.00	1.00	0.00	0.04	0.04	8
Candelon-GMM-CC	0.01	250	0.00	1.00	0.00	0.04	0.04	137
Haas-TBF	0.01	250	1.00	0.00	0.05	0.00	0.05	791
Engle-DQ	0.01	250	1.00	0.00	0.06	0.00	0.06	20
Kramer-GINI	0.01	250	0.97	0.03	0.30	0.00	0.29	3,059
Kupiec-POF	0.01	500	0.52	0.48	0.01	0.01	0.01	3
Binomial-POF	0.01	500	0.56	0.44	0.01	0.01	0.01	4
Christoffersen-CWeibull	0.01	500	0.00	1.00	0.00	0.01	0.01	624
Kupiec-TUFF	0.01	500	0.96	0.04	0.02	0.00	0.02	108
Haas-DWeibull	0.01	500	0.00	1.00	0.00	0.02	0.02	151
Candelon-GMM-UC	0.01	500	0.12	0.88	0.00	0.02	0.02	394
Candelon-GMM-CC	0.01	500	0.07	0.93	0.00	0.03	0.03	321
Christoffersen-ICoverage	0.01	500	0.00	1.00	0.00	0.03	0.03	13
Berkowitz-BoxLjung	0.01	500	0.96	0.04	0.04	0.00	0.03	160
Christoffersen-CCoverage	0.01	500	1.00	0.00	0.04	0.00	0.04	21
Haas-TBF	0.01	500	1.00	0.00	0.04	0.00	0.04	1,253
Engle-DQ	0.01	500	1.00	0.00	0.11	0.00	0.11	485
Kramer-GINI	0.01	500	1.00	0.00	0.38	0.00	0.38	3,394
Binomial-POF	0.01	1000	0.47	0.53	0.01	0.01	0.01	6
Christoffersen-CWeibull	0.01	1000	0.00	1.00	0.00	0.01	0.01	736
Candelon-GMM-UC	0.01	1000	0.17	0.83	0.00	0.01	0.01	532
Kupiec-POF	0.01	1000	0.64	0.36	0.01	0.01	0.01	6
Kupiec-TUFF	0.01	1000	0.98	0.02	0.02	0.00	0.01	146
Candelon-GMM-CC	0.01	1000	0.31	0.69	0.01	0.02	0.01	447
Haas-DWeibull	0.01	1000	0.00	1.00	0.00	0.03	0.03	253
Christoffersen-ICoverage	0.01	1000	0.00	1.00	0.00	0.03	0.03	17
Berkowitz-BoxLjung	0.01	1000	1.00	0.00	0.04	0.00	0.04	337
Christoffersen-CCoverage	0.01	1000	1.00	0.00	0.05	0.00	0.05	24
Haas-TBF	0.01	1000	1.00	0.00	0.06	0.00	0.06	1,503
Engle-DQ	0.01	1000	1.00	0.00	0.06	0.00	0.06	710
Kramer-GINI	0.01	1000	1.00	0.00	0.41	0.00	0.41	2,576

Table 18. Summary table – an assessment of size of VaR backtests ($\boldsymbol{p}$-values ranging from 0 to 0.1) (cont.)

Test	α	n	T_O	T_U	A_O	A_U	A	D
Candelon-GMM-UC	0.05	250	0.36	0.64	0.00	0.00	0.00	365
Christoffersen-CWeibull	0.05	250	0.99	0.01	0.00	0.00	0.00	920
Binomial-POF	0.05	250	0.45	0.55	0.01	0.01	0.01	6
Kupiec-POF	0.05	250	0.60	0.40	0.01	0.01	0.01	7
Engle-DQ	0.05	250	1.00	0.00	0.01	0.00	0.01	768
Candelon-GMM-CC	0.05	250	0.34	0.66	0.01	0.02	0.01	469
Berkowitz-BoxLjung	0.05	250	0.32	0.68	0.01	0.02	0.01	399
Kupiec-TUFF	0.05	250	0.80	0.20	0.02	0.00	0.02	46
Haas-DWeibull	0.05	250	0.00	1.00	0.00	0.03	0.03	685
Christoffersen-ICoverage	0.05	250	0.00	1.00	0.00	0.03	0.03	50
Haas-TBF	0.05	250	1.00	0.00	0.05	0.00	0.05	1,636
Kramer-GINI	0.05	250	0.99	0.01	0.07	0.00	0.06	1,744
Christoffersen-CCoverage	0.05	250	1.00	0.00	0.07	0.00	0.07	69
Candelon-GMM-UC	0.05	500	0.20	0.80	0.00	0.00	0.00	478
Binomial-POF	0.05	500	0.56	0.44	0.01	0.01	0.01	8
Engle-DQ	0.05	500	0.88	0.12	0.01	0.00	0.01	799
Kupiec-POF	0.05	500	0.76	0.24	0.01	0.00	0.01	9
Candelon-GMM-CC	0.05	500	0.50	0.50	0.01	0.01	0.01	554
Haas-DWeibull	0.05	500	0.00	1.00	0.00	0.01	0.01	889
Christoffersen-ICoverage	0.05	500	0.21	0.79	0.01	0.01	0.01	86
Berkowitz-BoxLjung	0.05	500	0.14	0.86	0.00	0.02	0.02	444
Kupiec-TUFF	0.05	500	0.78	0.22	0.02	0.00	0.02	43
Christoffersen-CWeibull	0.05	500	1.00	0.00	0.02	0.00	0.02	1,179
Kramer-GINI	0.05	500	1.00	0.00	0.07	0.00	0.07	1,603
Haas-TBF	0.05	500	1.00	0.00	0.08	0.00	0.08	1,881
Christoffersen-CCoverage	0.05	500	1.00	0.00	0.08	0.00	0.08	106
Haas-DWeibull	0.05	1000	0.56	0.44	0.00	0.00	0.00	948
Engle-DQ	0.05	1000	0.83	0.17	0.00	0.00	0.00	806
Candelon-GMM-UC	0.05	1000	0.23	0.77	0.00	0.00	0.00	590
Kupiec-POF	0.05	1000	0.73	0.27	0.00	0.00	0.00	13
Binomial-POF	0.05	1000	0.65	0.35	0.00	0.00	0.00	12
Candelon-GMM-CC	0.05	1000	0.58	0.42	0.01	0.01	0.01	641
Berkowitz-BoxLjung	0.05	1000	0.06	0.94	0.00	0.02	0.02	452
Kupiec-TUFF	0.05	1000	0.82	0.18	0.02	0.00	0.02	48
Christoffersen-ICoverage	0.05	1000	0.86	0.14	0.03	0.00	0.03	145
Christoffersen-CWeibull	0.05	1000	1.00	0.00	0.04	0.00	0.04	1,370
Kramer-GINI	0.05	1000	1.00	0.00	0.06	0.00	0.06	1,490
Christoffersen-CCoverage	0.05	1000	1.00	0.00	0.10	0.00	0.10	186
Haas-TBF	0.05	1000	1.00	0.00	0.12	0.00	0.12	2,404

Source: authors' calculation.

Table 19. Definitions of the utilised measures

Measure	Description
α	VaR significance level
n	length of the backtesting time-window
T_O	oversize frequency
T_U	undersize frequency
A_0	average oversize value
A_U	average undersize value
A	Ill-size measure; average deviation from the correct size
D	discretization measure; number of unique p-values in 0.01 – 0.1

Source: authors' work.

Przegląd Statystyczny. Statistical Review, 2020, vol. 67, 2, 152–173 DOI: 10.5604/01.3001.0014.5727

The application of selected methods of multivariate statistical analysis to study objective quality of life in Polish and Belarusian regions

Valiantsina Lialikava,[a] Iwona Skrodzka,[b] Alena Kalinina[c]

Abstract. The concept of life quality has been studied by specialists from a variety of scientific fields: economics, social geography, sociology, psychology, medicine, political sciences, and others. This contributes to the complementariness of the notion and broadens its interdisciplinary perspective, but on the other hand, it leads to a lack of unanimity in terms of the definition and measurement of the quality of life. Meanwhile, all developed countries in the world regard enhancing life quality as a priority of state policy. With the further advancement of our civilisation, quality of life will become a major issue in economic development. Therefore, monitoring this aspect of economic life, at both country and regional level, seems to be of particular significance. The paper aims to assess the suitability of selected methods of multivariate statistical analysis for the construction of a synthetic measure of objective quality of life. The study employs two methods of constructing synthetic measures of objective life quality: the linear ordering method – TOPSIS, and factor analysis. The results obtained by means of multivariate statistical analysis methods made it possible to create ratings of Polish and Belarusian regions in terms of objective quality of life and to further divide the regions into typological groups.

Keywords: objective quality of life, TOPSIS, factor analysis
JEL: C38; I31; R13

1. Introduction

As advanced integration processes are being implemented, globalisation is becoming a significant factor of economic and social change in modern society. Not only are socio-economic inequalities between the populations of different countries persisting, they are actually increasing. These inequalities, accompanied by a declining standard and quality of life (QoL), undermine economic growth, have an adverse effect on social stability, and exclude large groups of individuals from participating in the political, economic and social life of their countries. Therefore, in all developed countries in the world, improving life quality is considered a priority of state policy. It is becoming increasingly evident that with the further advancement of civilisation, quality of life is bound to be a major factor of economic development. The standard and quality of life will fully reflect the efficiency of state structures and the social policy of governments.

[a] Yanka Kupala State University of Grodno, Faculty of Economics and Management, e-mail: vlialikova@tut.by.
[b] University of Bialystok, Faculty of Economics and Finance, e-mail: i.skrodzka@uwb.edu.pl, ORCID: https://orcid.org/0000-0002-3261-8687.
[c] Yanka Kupala State University of Grodno, Faculty of Economics and Management, e-mail: elena-niko@inbox.ru.

Internal disparities exist among regions both in Poland and Belarus (voivodships and oblasts, respectively) as far as QoL is concerned. This is confirmed not only by statistical analyses of indicators relating to QoL (National Statistical Committee of the Republic of Belarus [Belstat], 2018; Central Statistical Office [GUS], 2017), but also by research conducted by numerous scientists (Bąk & Szczecińska, 2016; Lialikava et al., 2017; Lialikava & Kalinina, 2016; Nowak, 2018; Winiarczyk-Raźniak & Raźniak, 2011).

Polish strategic documents, directly or indirectly, refer to the category 'quality of life'. The improvement of Polish citizens' life quality is the main strategic goal of the long-term national development strategy (Ministry of Administration and Digitization, 2013, p. 42). The strategy provides for increased expenditure in the following areas: education, health, infrastructure, research and development, and culture. Also the medium-term national development strategy (Ministry of Regional Development [MRD], 2012, p. 20) is concerned with improving QoL. Its main aim is to strengthen and exploit economic, social and institutional potentials to ensure faster and sustainable growth of the economy, and the improvement of the quality of life of the population.

As regards Belarus, the national strategy for sustainable socio-economic development (Ministry of Economy of the Republic of Belarus [MINEC], 2015) confirms the urgency of the task of enhancing people's life quality. The strategy highlights the following aspects of quality of life: accessibility of high-quality education and health services, ensuring high-quality housing, a wide access to cultural goods and high standards of personal and environmental security. Monitoring the QoL in regions, analysing interregional variations, and seeking factors that could contribute to reducing socio-economic inequalities thus seem crucial here. This is acknowledged by state authorities, institutions which gather and analyse data, as well as economic researchers.

The purpose of this paper is to assess the suitability of selected methods of multivariate statistical analysis (MSA) for the construction of a synthetic measure of objective life quality, in particular by means of the TOPSIS method and factor analysis. Sixteen Polish voivodships and seven Belarusian regions are considered in the study. Due to the variety of differences between the countries, including the level of economic development, political systems, cultural backgrounds, and considering the fact that the present paper is a pilot study, the two countries were analysed separately. Year 2016 was selected as the period of interest because of the availability of statistical data.

Section 2 discusses the theoretical aspects of the 'quality of life' category. Subsequently, a description of the multivariate statistical analysis methods used in the research is provided. Section 4 contains a presentation of the diagnostic variables used to construct synthetic measures of objective QoL. Sections 5 and 6 are devoted to a discussion of the results. Section 7 offers a comparative analysis of the outcomes of the study. The paper closes with conclusions on the findings.

2. The concept of quality of life

The term 'quality of life' was first used in 1958 by a British economist A. C. Pigou (see Pigou, 1920, p. 32), but it did not gain much popularity at the time. The first scientific approach to the problem appeared in the theories proposed by Bell (1976), Galbraith (1958), and Toffler (1980). In the 1960s, a quantitative approach to the concept in question prevailed. With time, however, a qualitative perspective became more prominent. In the 1970s, the so-called 'binary' concept of life quality was developed. The level of life quality, understood as physical, emotional, material and social well-being, began to be considered not only with regard to objective facts, but also individual, subjective notions and perceptions. The 1980s saw an increased interest in quality of life as referred to an individual. Apart from investigating the socio-economic aspects, researchers also began to analyse non-material factors, like the welfare of a person or life satisfaction (Lialikava et al., 2017).

Despite the fact that the phrase 'quality of life' has functioned in the theory of economics and in economic practice for many years, the debate on its precise definition is still going on. In the debate, one can distinguish two opposing positions (Borys, 2015, pp. 2–3):

- the belief that quality of life cannot be universally defined, because there are too many ways of interpreting the notion and too many dimensions which would have to be taken into account should a uniform definition be adopted;
- attempts at creating a universal definition of life quality in spite of the numerous difficulties resulting from a number of factors, including the complexity of the problem, its interdisciplinary character or an overlapping of the scientific and colloquial understanding of the phrase 'quality of life'.

The term 'quality of life' is used interchangeably with the following phrases: well-being, living conditions, level of living, living standards, way of life or lifestyle. The differences or similarities between these expressions have not been clearly identified, which often leads to theoretical and practical contradictions (Borys, 2015, p. 2).

Table 1 presents selected approaches to defining QoL.

Table 1. Selected approaches to defining quality of life

Author	Definition
T. Słaby	All aspects of human life associated with the existence of a person, being someone, and experiencing various emotional states caused by, e.g. having a family, colleagues, friends (Słaby, 1990, p. 8).
T. Borys	The image of life perceived on the basis of a specific system of values (axiological system). This image (as a collective attribute of an individual or a group) can be described in a subjective or objective manner, from a one-dimensional or a multidimensional perspective etc., depending on the tools used. The tools applied to describing quality of life create its different typologies (Borys, 2015, p. 4).
E. Skrzypek	A combination of objective conditions: economic circumstances, leisure time, housing conditions, natural environment, health, social environment, and subjective conditions, which are perceived in a unique way by every individual and are reflected in their well-being (Skrzypek, 2001, p. 8).
R. Kolman	The degree to which the spiritual and material needs of individuals and society as a whole are satisfied, the degree to which the expectations of contractual normality in everyday activities of individuals and the society are met (Kolman, 2000, p. 2).
WHO	An individual's perception of their position in life in the context of culture and systems of values in which they live and in relation to their goals, expectations, standards and concerns.[1]

Source: authors' work.

The empirical research presented in the paper concerned objective QoL and included such areas as the quality of the population, material living conditions, social sphere, environment, and cultural sphere.

3. Research methodology

TOPSIS is a linear ordering method. It involves calculating the distance of each multi-attribute object from the pattern and anti-pattern, followed by a linear ordering of the objects. In taxonomic studies, the first linear ordering method using a pattern was presented by a Polish statistician Zdzisław Hellwig in 1968 (Hellwig, 1968). Hellwig's article initiated intensive research in this field, carried on by other Polish scientists, including Bartosiewicz (1976), Borys (1978), Cieślak (1974), Pluta (1976), Strahl (1978) and Walesiak (1993). In terms of the decision theory, the first linear ordering method with a pattern and anti-pattern was proposed by C.L. Hwang and K. Yoon in 1981, and was named TOPSIS (Hwang & Yoon, 1981).

The study's objective was achieved in the following 6 stages:

Stage 1. Selection of diagnostic variables on the basis of substantive and statistical factors. The diagnostic variables which were initially chosen for analysis (see

[1] See: WHOQOL: Measuring Quality of Life, https://www.who.int/healthinfo/survey/whoqol-qualityoflife/en/ (access: 12.08.2020)

Table 2.) were universally acknowledged, substantively valuable, measurable and confirmed by accessible statistics. In statistical terms, the level of variation was examined (a 10% value of the classical coefficient of variation was assumed as critical), as was the level of correlation (in order to eliminate excessively correlated variables, the inverse correlation matrix by Malina and Zeliaś (1997) was applied).

Stage 2. Division of diagnostic variables into stimulants (a higher value of such a variable means a higher level of the studied phenomenon) and destimulants (a higher value of such a variable means a lower level of the studied phenomenon).[2]

Stage 3. Normalisation of the values of diagnostic variables. The zero unitarisation procedure was adopted,[3] as represented by the equations below (Kukuła, 2000):

- for stimulants

$$z_{ik} = \frac{x_{ik} - \min_{\mathrm{i}}\{x_{ik}\}}{\max_{\mathrm{i}}\{x_{ik}\} - \min_{\mathrm{i}}\{x_{ik}\}}, \tag{1}$$

- for destimulants

$$z_{ik} = \frac{\max_{\mathrm{i}}\{x_{ik}\} - x_{ik}}{\max_{\mathrm{i}}\{x_{ik}\} - \min_{\mathrm{i}}\{x_{ik}\}}, \tag{2}$$

where
i – number of region ($i = 1, 2, \ldots, n$),
k – number of diagnostic variable ($k = 1, 2, \ldots, m$).

Stage 4. Calculation of the Euclidean distance of each region from the pattern $z_k^+ = [1,\ 1, \ldots,\ 1]$ and from the anti-pattern $z_k^- = [0,\ 0, \ldots,\ 0]$, according to the following equations:

- distance from the pattern

$$d_i^+ = \sqrt{\sum_{k=1}^{m} (z_{ik} - z_k^+)^2}\,, \tag{3}$$

- distance from the anti-pattern

[2] The concepts of stimulants and destimulants were introduced into the literature by Hellwig (1968).
[3] In Hwang and Yoon's original work, the quotient transformation was used to normalise the variables (Hwang & Yoon, 1981, pp. 131–132).

$$d_i^+ = \sqrt{\sum_{k=1}^{m}(z_{ik} - z_k^-)^2}\,, \tag{4}$$

where $(i = 1, 2, \dots, n)$.

Stage 5. Calculation of the value of the synthetic measure for each region, by means of the following formula (see Hwang & Yoon, 1981, p. 132):

$$q_i = \frac{d_i^-}{d_i^- + d_i^+}. \tag{5}$$

The values of the synthetic measure fall within the range [0,1]. The measure takes the value of 1 for the pattern and 0 for the anti-pattern. The closer the value of the measure to 1, the less the given region diverges from the pattern.

Stage 6. Ordering of the studied regions and their division into typological groups.

The process of ordering was conducted on the basis of the value of the synthetic measure calculated in the previous stage. The boundaries of the intervals were established through arithmetic means and standard deviation of the synthetic measure:

- group I (very high and high objective QoL): $q_i \geq \bar{q} + s_q$,
- group II (medium-higher objective QoL): $\bar{q} \leq q_i < \bar{q} + s_q$,
- group III (medium-lower objective QoL): $\bar{q} - s_q \leq q_i < \bar{q}$,
- group IV (low and very low objective QoL): $q_i < \bar{q} - s_q$.

Factor analysis is a collection of techniques and procedures used to reduce a large number of studied variables to a far smaller group of mutually independent factors or principal components. It consists of the classical factor analysis and the principal component analysis. The former, whose main ideas were developed by Spearman (1904) and Thurstone (1931), is primarily used to investigate the internal relationships between variables. The latter, whose theoretical foundations were devised by Hotelling (1933) and Pearson (1901), is applied for analysing interdependencies within sets of variables or for studying the structures of sets of observations.

In the case of factor analysis, the following algorithm was used:

Stage 1. Selection of diagnostic variables on the basis of substantive and statistical reasons. Here, the applied substantive criteria were analogous to those used in the TOPSIS method and therefore, the set of diagnostic variables was also analogous (see Table 2.). The level of correlation of diagnostic variables was studied in statistical terms – the use of factor analysis is only justified if at least some of the variables are correlated.

Stage 2. Division of diagnostic variables into stimulants and destimulants.

Stage 3. Normalisation of the values of diagnostic variables – the zero unitarisation procedure was applied, according to Equations (1) and (2).

Stage 4. Estimation of the factor analysis model using the principal components method (Härdle & Simar, 2015, pp. 367–375; Timm, 2002, pp. 502–506).

Stage 5. Establishing the number of principal factors.

In the paper, the Kaiser criterion was adopted, which involves the elimination of those principal factors whose singular values are lower than 1 (Jolliffe, 2002, pp. 114–115).

Stage 6. Calculation of the value of the synthetic measure for each region, by means of the equation below:

$$R = \left(\sum_{i=1}^{m} \lambda_i \right)^{-1} (\lambda_1 F_1 + \cdots + \lambda_m F_m) \times 100, \tag{6}$$

where

F_i – values of the first m factors

λ_i – singular values of the covariance matrix.

Stage 7. Ordering of the studied regions and their division into typological groups on the basis of the values of synthetic measure. The division into groups was conducted through k-means clustering (Härdle & Simar, 2015, pp. 385–406; Timm, 2002, pp. 522–523).

4. Diagnostic variables

The lack of a single, widely accepted definition of QoL results in the fact that there is no unambiguous method of measuring this category. International organisations, e.g. the European Union, the United Nations, the World Bank, the OECD, as well as individual countries, including Poland[4] and Belarus,[5] have been involved in developing criteria for assessing the standard and quality of people's lives. Scientists and practitioners make attempts at constructing synthetic measures of quality of life. The

---

[4] Research into QoL in Poland is conducted by Statistics Poland, whose published reports (every two years) contain indicators on the following areas of QoL: material situation, work, health, education, free time and social relations, personal safety, state quality and basic rights, quality of the natural environment in the place of residence as well as the subjective well-being. In addition, the Social Monitoring Board has been conducting research within the framework of the project 'Social Diagnosis' since 2000. Data relating to households and attitudes, state of mind and behaviours of their members are obtained.

[5] Research into QoL in Belarus is conducted by the National Statistical Committee of the Republic of Belarus. The published reports contain socio-economic indicators measuring the quality and standard of living of the inhabitants of Belarusian cities and regions.

most widely known ones include: the Human Development Index, the Physical Quality of Life Index, Gross National Happiness, the Happy Planet Index, The Economist Intelligence Unit's Quality-of-Life Index, the Gallup-Healthways Global Well-being Index, and the Legatum Prosperity Index. Also this paper undertakes to develop synthetic measures of quality of life on a regional level.

Table 2 presents a set of diagnostic values used in the study. The set is the result of a compromise between knowledge and experience in measuring objective QoL and the accessibility of comparable data for the two groups of regions investigated by the authors. The variables were classified according to five categories: quality of the population, material living conditions, social sphere, environment and cultural sphere. The statistical data were obtained from the official databases of statistical offices in Poland and Belarus. The study covers the year 2016.

Table 2. Diagnostic variables of objective QoL

Symbol	Diagnostic variable	Type of variable[a]
	Quality of population	
*X*1	Net migration rate	S
*X*2	Birth rate per 1,000 population	S
*X*3	Life expectancy	S
*X*4	Infant mortality rate	D
*X*5	Pre-working age population per 1,000 persons of working age	S
*X*6	Post-working age population per 1,000 persons of working age	D
*X*7	Percentage of population with tertiary education employed in the economy (Polish regions) Percentage of population with tertiary education employed in organisations (Belarusian regions)	S
*X*8	Number of deaths per 1,000 population	D
*X*9	Number of marriages per 1,000 population	S
*X*10	Number of divorces per 1,000 population	S
	Material living conditions	
*X*11	Average monthly salary	S
*X*12	Average usable floor area of residential premises per person	S
*X*13	Number of passenger cars per 1,000 population	S
*X*14	GDP per capita	S
*X*15	Retail sale of goods per person	S
*X*16	Percentage of households below poverty line	D
	Social sphere	
*X*17	Unemployment rate	D
*X*18	Employment rate	S
*X*19	Number of doctors per 10,000 population	S
*X*20	Number of nurses and midwives per 10,000 population	S
*X*21	Number of injured in accidents at work per 1,000 employed	D
*X*22	Number of offences per 1,000 population	D

a S – stimulant, D – destimulant.

Table 2. Diagnostic variables of objective QoL (cont.)

Symbol	Diagnostic variable	Type of variable[a]
	Environment	
*X*23	Emission of particulate pollutants by plants of significant impact on air quality	D
*X*24	Industrial and municipal waste water requiring treatment discharged into waters or into the ground	D
	Cultural sphere	
*X*25	Audience in theatres and music institutions per 1,000 population	S
*X*26	Number of museum admissions per 10,000 population	S

a S – stimulant, D – destimulant.
Source: authors' work.

5. Results obtained by means of TOPSIS

The diagnostic variables from Table 2 were verified statistically in order to eliminate data which were insufficiently varied or excessively correlated. Table 3 contains diagnostic variables used to construct the synthetic measures of objective QoL in the Polish and Belarusian regions. Each of the specified areas of objective QoL was represented by at least one variable.

Table 3. Diagnostic variables used for constructing synthetic measures of objective QoL

Polish regions	Belarusian regions
Quality of population	
*X*1, *X*4, *X*10	*X*1, *X*4, *X*7
Material living conditions	
*X*14, *X*16	*X*14
Social sphere	
*X*17, *X*21, *X*22	*X*17, *X*19, *X*21
Environment	
*X*23	*X*23
Cultural sphere	
*X*25	*X*25, *X*26

Source: authors' work.

The values of the variables were normalised according to Equations (1) and (2). Next, the values of the synthetic measures of objective QoL were calculated and, on this basis, linear ordering of the regions was performed, followed by their division into typological groups. The division into groups was conducted with the help of the mean and standard deviation of the synthetic measures. The results are presented in Tables 4 and 5.

The variation of the value of the synthetic measure was approximately 36%. In the case of seven voivodships, the value of the synthetic measure was higher than the average (i.e. 0.49).

It was found that in 2016, Mazowieckie offered the best objective QoL in Poland, whereas in Warmińsko-Mazurskie, the QoL was the poorest of all the voivodships. In Mazowieckie Voivodship, five out of ten diagnostic variables reached the best values (these were $X1$, $X4$, $X14$, $X16$, $X21$), while the majority of diagnostic variables in Warmińsko-Mazurskie assumed the lowest values. The only exception was variable $X23$, in which Warmińsko-Mazurskie ranked the highest of all the voivodships. Mazowieckie, Małopolskie and Pomorskie were those three voivodships in which QoL was very high or high. The objective QoL of inhabitants of Dolnośląskie, Łódzkie, Podlaskie, and Wielkopolskie was medium-high. Eight voivodships: Podkarpackie, Lubelskie, Opolskie, Świętokrzyskie, Śląskie, Lubuskie, and Kujawsko-Pomorskie offered medium-low objective QoL. Only one voivodship, Warmińsko-Mazurskie, offered low or very low QoL.

Table 4. Ordering and classification of Polish voivodships in terms of objective QoL in 2016 – the TOPSIS method

Voivodship	Value of synthetic measure	Rating position	Group
Mazowieckie	0.904	1	1
Małopolskie	0.751	2	1
Pomorskie	0.689	3	1
Dolnośląskie	0.606	4	2
Łódzkie	0.584	5	2
Podlaskie	0.509	6	2
Wielkopolskie	0.491	7	2
Podkarpackie	0.474	8	3
Lubelskie	0.429	9	3
Opolskie	0.416	10	3
Świętokrzyskie	0.411	11	3
Śląskie	0.359	12	3
Lubuskie	0.358	13	3
Kujawsko-Pomorskie	0.349	14	3
Zachodniopomorskie	0.347	15	3
Warmińsko-Mazurskie	0.178	16	4

Source: authors' calculation.

The City of Minsk clearly stood out from the rest of the Belarusian regions. It was at the top of the rating and was the only region with very high or high objective QoL. For the City of Minsk, eight out of ten diagnostic variables took the highest values (except for $X4$ and $X23$). In the other six regions, the objective QoL was medium-low, which shows the vastness of the gap between the capital city and the rest of the country. The variation in the value of the synthetic measure, assessed by means of the classical coefficient of variation, reached almost 56%, whereas after the exclusion of the City of Minsk region, it was only 20%.

Table 5. Ordering and classification of Belarusian regions in terms of objective QoL in 2016 – the TOPSIS method

Region	Value of synthetic measure	Rating position	Group
City of Minsk	0.8862	1	1
Minsk Region	0.3687	2	3
Grodno Region	0.3603	3	3
Gomel Region	0.3214	4	3
Vitebsk Region	0.2961	5	3
Brest Region	0.2406	6	3
Mogilev Region	0.1947	7	3

Source: authors' calculation.

6. Results obtained by means of factor analysis

The factor analysis procedure began with the analysis of the correlation matrix of the diagnostic variables in order to find whether at least some of the variables are correlated. After assessing that the condition was fulfilled, the variables were normalised according to Equations (1) and (2). Next, estimation of the model was performed and the principal factors were identified. Table 6 presents the obtained results.

In the case of Polish voivodships, the first principal factor explains 34% of the total variation, the second – approximately 20%, and the third – 15.5%. Therefore, the first three principal factors explain together 70.3% of the total variation. As regards the Belarusian regions, the first principal factor explains about 60% of the total variation, whereas the second one – approximately 20%, which accounts for nearly 80% of the total variation.

Table 6. Results of the factor analysis

Factor	Polish regions		Belarusian regions	
	percentage of total variation	cumulative percentage of total variation	percentage of total variation	cumulative percentage of total variation
F_1	34.21	34.21	59.39	59.39
F_2	20.56	54.77	19.69	79.08
F_3	15.54	70.32	–	–

Source: authors' calculation.

The synthetic measure for assessing objective QoL in the Polish and Belarusian regions was constructed on the basis of the following formulae:

$$R_{PL} = 34.21F_1 + 20.56F_2 + 15.54F_3, \tag{7}$$

$$R_{BLR} = 59.39F_1 + 19.96F_2, \tag{8}$$

where F_1, F_2, F_3 are the estimated values of the first three principal factors, while the accompanying coefficients represent the percentages of the total variation given in Table 6.

Table 7 provides the values of the synthetic measure and the rating and division of Polish voivodships into typological groups. The division into typological groups was conducted according to the k-means clustering method.

Table 7. Ordering and classification of Polish voivodships in terms of objective QoL in 2016 – results of factor and cluster analysis

Voivodship	Rating position	Value of synthetic measure	Group
Mazowieckie	1	98.15	1
Małopolskie	2	72.01	1
Pomorskie	3	52.58	1
Wielkopolskie	4	33.00	1
Dolnośląskie	5	7.33	2
Podkarpackie	6	3.22	2
Podlaskie	7	−2.93	2
Lubelskie	8	−15.98	2
Śląskie	9	−18.39	2
Lubuskie	10	−19.61	2
Kujawsko-Pomorskie	11	−22.37	2
Zachodniopomorskie	12	−29.66	3
Opolskie	13	−32.24	3
Łódzkie	14	−34.54	3
Świętokrzyskie	15	−38.77	3
Warmińsko-Mazurskie	16	−51.80	3

Source: authors' calculation.

Mazowieckie, Małopolskie, Pomorskie and Wielkopolskie were the leaders of the rating. They formed the first group of voivodships, characterised by very high or high objective QoL. The second group consisted of seven voivodships: Dolnośląskie, Podkarpackie, Podlaskie, Lubelskie, Śląskie, Lubuskie and Kujawsko-Pomorskie. The remaining five regions were classified in the third, and last, typological group.

The results of the ordering and classification of the Belarusian regions are shown in Table 8.

Table 8. Ordering and classification of Belarusian regions in terms of objective QoL in 2016 – results of factor and cluster analysis

Region	Rating position	Value of synthetic measure	Group
City of Minsk	1	131.07	1
Minsk Region	2	25.95	2
Grodno Region	3	−11.56	3
Brest Region	4	−24.05	3
Mogilev Region	5	−38.81	3
Gomel Region	6	−39.22	3
Vitebsk Region	7	−43.37	3

Source: authors' calculation.

The City of Minsk was the leader of the Belarusian regions and alone formed the first typological group. The other regions were classified in the second (Minsk Region) or the third group (Grodno Region, Brest Region, Mogilev Region, Gomel Region and Vitebsk Region).

7. Comparison of research results

Tables 9 and 10 provide the results of the ordering of the Polish and Belarusian regions in terms of objective QoL, done by means of the multivariate statistical analysis methods.

The ratings of Polish regions are characterised by high correlation, as reflected by the values of Spearman's rank correlation coefficients, at 0.77. In both ratings, the first, second, third and sixteenth places are occupied by the same voivodships (Mazowieckie, Małopolskie, Pomorskie, and Warmińsko-Mazurskie, respectively). The largest difference between the two ratings was obeserved for Łódzkie Voivodship. The low rank of the region in the rating constructed through factor analysis may result from the fact that Łódzkie was rated near the bottom in terms of four diagnostic variables which were strongly correlated with the estimated values of the first principal factor (high values of factor loadings).

Table 9. Ordering of Polish voivodships in terms of objective QoL – comparison of results obtained by means of applied MSA methods

Voivodship	TOPSIS	Factor analysis	Difference in rating
Dolnośląskie	4	5	1
Kujawsko-Pomorskie	14	11	3
Lubelskie	9	8	1
Lubuskie	13	10	3
Łódzkie	5	14	9
Małopolskie	2	2	0
Mazowieckie	1	1	0
Opolskie	10	13	3
Podkarpackie	8	6	2
Podlaskie	6	7	1
Pomorskie	3	3	0
Śląskie	12	9	3
Świętokrzyskie	11	15	4
Warmińsko-Mazurskie	16	16	0
Wielkopolskie	7	4	3
Zachodniopomorskie	15	12	3

Source: authors' calculation.

Table 10. Ordering of Belarusian regions in terms of objective QoL – comparison of results obtained by means of the applied MSA methods

Region	TOPSIS	Factor analysis	Difference in rating
City of Minsk	1	1	0
Brest Region	6	4	2
Grodno Region	3	3	0
Gomel Region	4	6	2
Minsk Region	2	2	0
Mogilev Region	7	5	2
Vitebsk Region	5	7	2

Source: authors' calculation.

The level of correlation among the ratings of the Belarusian regions is similar to that among the Polish voivodships. The value of the Spearman's rank amounts to 0.71. In both ratings of the Belarusian regions, the first three places are the same: the City of Minsk, Minsk Region and Grodno Region. The rankings of the other regions vary.

Tables 11 and 12 represent the division of the Polish and Belarusian regions into typological groups created through the use of k-means clustering and the method based on the mean and standard deviation of the synthetic measure determined by TOPSIS.

Table 11. Division of Polish voivodships into typological groups in terms of objective QoL in 2016 – comparison of results obtained by means of the applied MSA methods

Cluster analysis \ TOPSIS	Group I	Group II	Group III	Group IV
Group I	Mazowieckie, Małopolskie, Pomorskie	Wielkopolskie		
Group II		Dolnośląskie, Podlaskie	Podkarpackie, Lubelskie, Śląskie, Lubuskie, Kujawsko-Pomorskie	
Group III		Łódzkie	Opolskie, Świętokrzyskie, Zachodnio-pomorskie	Warmińsko-Mazurskie

Source: authors' calculation.

The cluster analysis revealed three clusters of Polish voivodships and three clusters of Belarusian regions. The classification based on the results of the TOPSIS procedure led to the division of Polish voivodships into four typological groups, whereas in the case of the Belarusian regions, it was two groups. While the results concerning the Belarusian regions are similar, the groupings of the Polish voivodships manifest considerable differences.

Table 12. Division of Belarusian regions into typological groups in terms of objective QoL in 2016 – comparison of results obtained by means of the applied MSA methods

Cluster analysis \ TOPSIS	Group I	Group II	Group III	Group IV
Group I	City of Minsk			
Group II			Minsk Region	
Group III			Grodno Region, Gomel Region, Vitebsk Region, Brest Region, Mogilev Region	

Source: authors' calculation.

Research into quality of life at regional level has been conducted by various specialists (e.g. Bąk & Szczecińska, 2016; Lialikava et al., 2017; Lialikava & Kalinina, 2016; Nowak, 2018; Winiarczyk-Raźniak & Raźniak, 2011). However, it is difficult, and in many cases even impossible, to compare the results of these studies, since they pertain to different periods or examine different research objects.

8. Conclusions

The paper presents the results of research into objective QoL in Polish and Belarusian regions, obtained through the application of selected methods of multivariate statistical analysis. Two very different methods – TOPSIS and factor analysis – were used in order to construct synthetic measures of objective QoL and to create ratings and typological groups of the studied regions. The methods differed already at the stage of selecting diagnostic variables. In TOPSIS, the set of diagnostic variables had to be narrowed down by eliminating too strongly correlated variables, whereas in the factor analysis procedure, quite the opposite – the variables had to be correlated. The TOPSIS method involved studying the distance of the regions from the pattern and anti-pattern, and the obtained values of the synthetic measure fell within the range [0, 1]. Factor analysis involved searching for hidden factors in a set of diagnostic variables, and the obtained values of the synthetic measure were infinite. Due to the above-mentioned differences between the applied methods, discrepancies occurred in the ratings and typological groups. The choice of the method affects the outcome of the study, so it should be determined by the purpose of the research and by what is expected of the selected method, as well as on the statistical properties of the analysed set of diagnostic variables.

Evaluating QoL at the regional level is a particularly significant task in the context of socio-economic analyses. The results of the present study can serve as a tool for planning or monitoring the utilisation of financial resources granted to local government units. They can also be used to assess the efficiency of the already implemented socio-economic policies.

Appendix

Table 1A. Basic descriptive statistics of diagnostic variables in the 'quality of population' area – Polish regions in 2016

Symbol	Min	Max	Mean	Coefficient of variation
*X*1 (S)	–2.07 Warmińsko-Mazurskie	2.41 Mazowieckie	–0.38	326.78
*X*2 (S)	–2.98 Łódzkie	2.04 Pomorskie	–0.40	355.61
*X*3 (S)	76.58 Łódzkie	79.26 Podkarpackie	77.95	0.91
*X*4 (D)	3.27 Mazowieckie	5.86 Lubuskie	4.18	16.60
*X*5 (S)	25.21 Opolskie	31.59 Pomorskie	28.66	5.82
*X*6 (D)	29.20 Warmińsko-Mazurskie	37.00 Łódzkie	32.48	6.36
*X*7 (S)	26.70 Łódzkie	42.91 Mazowieckie	32.15	11.48
*X*8 (D)	8.98 Podkarpackie	12.14 Łódzkie	10.09	7.85
*X*9 (S)	4.65 Warmińsko-Mazurskie	5.36 Pomorskie	5.01	4.57
*X*10 (S)	1.20 Podkarpackie	1.90 Warmińsko-Mazurskie, Lubuskie	1.62	12.20

Source: authors' calculation.

Table 2A. Basic descriptive statistics of diagnostic variables in the 'material living conditions' area – Polish regions in 2016

Symbol	Min	Max	Mean	Coefficient of variation
*X*11 (S)	3619.16 Warmińsko-Mazurskie	5240.86 Mazowieckie	3993.79	9.94
*X*12 (S)	24.2 Warmińsko-Mazurskie	29.9 Mazowieckie	27.11	5.58
*X*13 (S)	485.19 Podlaskie	626.59 Wielkopolskie	555.82	7.30
*X*14 (S)	33371 Lubelskie	77359 Mazowieckie	43765.31	24.81
*X*15 (S)	8597 Opolskie	40383 Mazowieckie	16295.69	56.77
*X*16 (D)	8.5 Mazowieckie	21.3 Podkarpackie	13.32	31.34

Source: authors' calculation.

Table 3A. Basic descriptive statistics of diagnostic variables in the 'social sphere' area – Polish regions in 2016

Symbol	Min	Max	Mean	Coefficient of variation
*X*17 (D)	4.90 Wielkopolskie	14.20 Warmińsko-Mazurskie	9.09	26.36
*X*18 (S)	49.00 Warmińsko-Mazurskie	56.60 Mazowieckie	52.19	3.98
*X*19 (S)	35.97 Wielkopolskie	71.32 Mazowieckie	52.59	18.87
*X*20 (S)	50.32 Wielkopolskie	76.45 Śląskie	66.95	11.71
*X*21 (D)	4.84 Mazowieckie	9.17 Warmińsko-Mazurskie	7.38	16.50
*X*22 (D)	11.02 Podkarpackie	25.73 Dolnośląskie	18.87	19.21

Source: authors' calculation.

Table 4A. Basic descriptive statistics of diagnostic variables in the 'environment' and 'cultural sphere' areas – Polish regions in 2016

Symbol	Min	Max	Mean	Coefficient of variation
*X*23 (D)	0.03 Warmińsko-Mazurskie	0.74 Śląskie	0.14	112.47
*X*24 (D)	2.00 Podlaskie	30.10 Śląskie	7.60	89.22
*X*25 (S)	100 Podkarpackie	716 Dolnośląskie	332.50	54.76
*X*26 (S)	2412 Opolskie	29363 Małopolskie	7674.49	90.86

Source: authors' calculation.

Table 5A. Basic descriptive statistics of diagnostic variables in the 'quality of population' area – Belarusian regions in 2016

Symbol	Min	Max	Mean	Coefficient of variation
*X*1 (S)	−1.75 Grodno Region	4.96 City of Minsk	0.37	793.15
*X*2 (S)	−3.52 Vitebsk Region	2.65 City of Minsk	−0.47	−381.11
*X*3 (S)	73.1 Minsk Region	76.5 City of Minsk	74.00	1.50
*X*4 (D)	2.8 Vitebsk Region, Gomel Region	3.8 Minsk Region	3.14	10.46
*X*5 (S)	267 City of Minsk	344 Brest Region	310.43	8.15
*X*6 (D)	369 City of Minsk	486 Vitebsk Region	450.29	7.90
*X*7 (S)	21.9 Minsk Region	40.7 City of Minsk	26.09	23.07
*X*8 (D)	8.7 CIty of Minsk	14.6 Vitebsk Region	12.93	14.15
*X*9 (S)	6.3 Vitebsk Region, Gomel Region	7.7 City of Minsk	6.69	6.61
*X*10 (S)	3.0 Grodno Region, Brest Region	3.8 City of Minsk	3.40	8.02

Source: authors' calculation.

Table 6A. Basic descriptive statistics of diagnostic variables in the 'material living conditions' area – Belarusian regions in 2016

Symbol	Min	Max	Mean	Coefficient of variation
*X*11 (S)	241.2 Gomel Region	445.0 City of Minsk	285.51	23.49
*X*12 (S)	22.5 City of Minsk	29.7 Minsk Region	27.14	8.24
*X*13 (S)	267 Gomel Region	352 Grodno Region	310.14	8.74
*X*14 (S)	6295.1 Viciebsk Region	12960.0 City of Minsk	8001.43	28.18
*X*15 (S)	3251.7 Mogilev Region	11285.4 City of Minsk	5274.77	48.13
*X*16 (D)	1.4 City of Minsk	8.1 Brest Region	5.81	39.07

Source: authors' calculation.

Table 7A. Basic descriptive statistics of diagnostic variables in the 'social sphere' area – Belarusian regions in 2016

Symbol	Min	Max	Mean	Coefficient of variation
*X*17 (D)	0.5 City of Minsk	1.0 Vitebsk Region, Gomel Region	0.84	19.91
*X*18 (S)	63.1 Gomel Region	71.5 City of Minsk	66.31	4.28
*X*19 (S)	32.8 Minsk Region	58.7 City of Minsk	42.76	20.07
*X*20 (S)	120.6 Minsk Region	137.2 Grodno Region	132.69	3.97
*X*21 (D)	0.30 City of Minsk	0.52 Grodno Region	0.44	16.49
*X*22 (D)	826 Brest Region	1203 Minsk region	974.57	11.68

Source: authors' calculation.

Table 8A. Basic descriptive statistics of diagnostic variables in the 'environment' and 'cultural sphere' areas – Belarusian regions in 2016

Symbol	Min	Max	Mean	Coefficient of variation
*X*23 (D)	1.45 Mogilev Region	60.33 City of Minsk	10.38	196.52
*X*24 (D)	0.09 City of Minsk	0.22 Minsk Region	0.14	29.26
*X*25 (S)	413.8 Minsk Region	3976.7 City of Minsk	1135.89	106.48
*X*26 (S)	434 Mogilev Region	827 City of Minsk	655.14	20.12

Source: authors' calculation.

References

Bartosiewicz, S. (1976). Propozycja metody tworzenia zmiennych syntetycznych. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, *84*, 5–7.

Bąk, I., Szczecińska, B. (2016). The Use of Multi-Criteria Taxonomy in the Study of Objective Quality of Life in Polish Voivodeships. *Folia Oeconomica Stetinensia*, *16*(1), 7–20. https://doi.org/10.1515/foli-2016-0001

Bell, D. (1976). *The Coming Of Post-Industrial Society: A vanture in Social Forecasting*. New York: Basic Books.

Belstat. (2018). *Regions of the Republic of Belarus. Socio-economic indicators*. Minsk: National Statistical Committee of the Republic of Belarus.

Borys, T. (1978). Propozycja agregatowej miary rozwoju obiektów. *Przegląd Statystyczny*, *25*(3), 371–381.

Borys, T. (2015). Typologia jakości życia i pomiar statystyczny. *Wiadomości Statystyczne*, *60*(7), 1–18. http://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.ekon-element-000171388701

Cieślak, M. (1974). Taksonomiczna procedura prognozowania rozwoju gospodarczego i określania potrzeb na kadry kwalifikowane. *Przegląd Statystyczny*, *21*(1), 29–39.

Galbraith, J. K. (1958). *The Affluent Society*. London: Hamish Hamilton. http://pinguet.free.fr/affluent58.pdf

GUS. (2017). *Quality of life in Poland. 2017 edition*. Warsaw: Central Statistical Office. https://stat.gov.pl/en/topics/living-conditions/living-conditions/quality-of-life-in-poland-2017-edition,5,4.html#

Härdle, W. K., Simar, L. (2015). *Applied Multivariate Statistical Analysis* (4th edition). Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-662-45171-7

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

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*(6), 417–441. https://doi.org/10.1037/h0071325

Hwang, C. L., Yoon, K. (1981). *Multiple Attribute Decision Making: Methods and Applications*. Berlin, Heidelberg: Springer-Verlag. http://dx.doi.org/10.1007/978-3-642-48318-9

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd edition). New York: Springer-Verlag. https://doi.org/10.1007/b98835

Kolman, R. (2000). Zespoły badawcze jakości życia. *Problemy Jakości*, *32*(2), 2–5.

Kukuła, K. (2000). *Metoda unitaryzacji zerowanej*. Warszawa: Wydawnictwo Naukowe PWN.

Lialikava, V., Kalinina, A. (2016). Factors improving quality of life in the Polish regions. *Vesnik of Yanka Kupala State University of Grodno. Series 5. Economics. Sociology. Biology*, *6*(3), 61–72.

Lialikava, V., Kalinina, A., Ziczkene, S. (2017). Essential Factors Determining the Quality of Population Life. *Socialiniai Tyrimai / Social Research*, *40*(2), 55–65. http://doi.org/10.21277/st.v40i2.197

Malina, A., Zeliaś, A. (1997). Taksonomiczna analiza przestrzennego zróżnicowania jakości życia ludności w Polsce w 1994 r. *Przegląd Statystyczny*, *44*(1), 11–27.

MINEC. (2017). *National Strategy of the Republic of Belarus for Sustainable Socio-Economic Development for the period until 2030*. Minsk: Ministry of Economy of the Republic of Belarus. http://www.economy.gov.by/uploads/files/NSUR2030/Natsionalnaja-strategija-ustojchivogo-sotsialno-ekonomicheskogo-razvitija-Respubliki-Belarus-na-period-do-2030-goda.pdf

Ministry of Administration and Digitization. (2013). *Poland 2030. The third wave of modernity. Long-term National Development Strategy*. Warsaw: Ministry of Administration and Digitization.

MRD. (2012). *National Development Strategy 2020*. Warsaw: Ministry of Regional Development. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiQyNzzhILtAhXotYsKHWrMA4IQFjAAegQIBBAC&url=https%3A%2F%2Fwww.trade.gov.tw%2FApp_Ashx%2FFile.ashx%3FFilePath%3D..%2FFiles%2FDoc%2F%25E6%25B3%25A2%25E8%2598%25AD2020%25E5%259C%258B%25E5%25AE%25B6%25E7%2599%25BC%25E5%25B1%259

5%25E7%25B6%25B1%25E9%25A0%2598(%25E8%258B%25B1%25E6%2596%2587).pdf&usg=AOvVaw1ta9rIayjtd9QX3rIqjeCd

Nowak, P. (2018). Regional variety in quality of life in Poland. *Oeconomia Copernicana*, *9*(3), 381–401. https://doi.org/10.24136/oc.2018.019

Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559–572. https://doi.org/10.1080/14786440109462720

Pigou, A. C. (1920). *The Economics of Welfare*. London: Macmillan and Co. http://pombo.free.fr/pigou1920.pdf

Pluta, W. (1976). Taksonomiczna procedura prowadzenia syntetycznych badań porównawczych za pomocą zmodyfikowanej miary rozwoju gospodarczego. *Przegląd Statystyczny*, *23*(4), 511–517.

Skrzypek, E. (2001). Ekonomiczne aspekty jakości życia. *Problemy Jakości*, *33*(1), 8–14.

Słaby, T. (1990). Poziom życia, jakość życia. *Wiadomości Statystyczne*, *35*(6), 8–10.

Spearman, C. (1904). 'General intelligence', objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–292. https://doi.org/10.2307/1412107

Strahl, D. (1978). Propozycja konstrukcji miary syntetycznej. *Przegląd Statystyczny*, *25*(2), 205–215.

Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*, *38*(5), 406–427. https://doi.org/10.1037/h0069792

Timm, N. H. (2002). *Applied Multivariate Analysis*. New York: Springer-Verlag.

Toffler, A. (1980). *The Third Wave*. New York: William Morrow and Company.

Walesiak, M. (1993). *Statystyczna analiza wielowymiarowa w badaniach marketingowych*. Wrocław: Wydawnictwo Akademii Ekonomicznej we Wrocławiu. http://keii.ue.wroc.pl/pracownicy/mw/1993_Walesiak_SAW_w_badaniach_marketingowych_OCR.pdf

Winiarczyk-Raźniak, A., Raźniak, P. (2011). Regional differences in the standard of living in Poland (based on selected indices). *Procedia – Social and Behavioral Sciences*, *19*, 31–36. https://doi.org/10.1016/j.sbspro.2011.05.103

Przegląd Statystyczny. Statistical Review, 2020, vol. 67, 2, 174–189

Professor Stanisława Bartosiewicz celebrates her 100th birthday

Józef Dziechciarz[a]

On 8 May 2020, Professor Stanisława Bartosiewicz celebrated her 100th birthday. Her friends and students gathered to celebrate this event and to honour the Professor's outstanding achievements in the field of research and didactics, and the academic guidance provided to her former students, who are now recognised scientists.

1. An outline of the biography

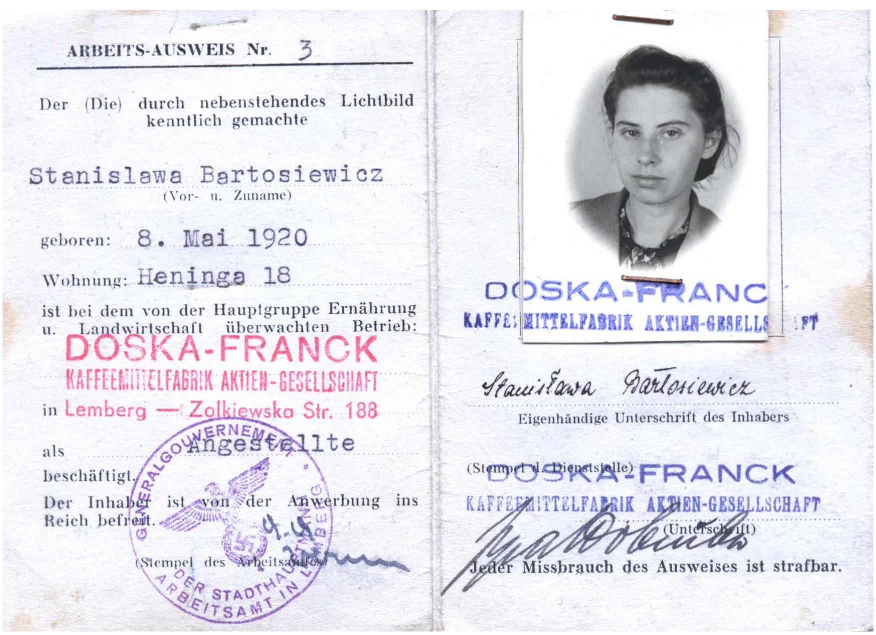
Professor Stanisława Bartosiewicz was born on 8 May 1920 in Brzeżany, Podolia, now Ukraine, in a family with academic background. In 1938, she graduated from the neoclassical gymnasium and began studying at the Academy of Foreign Trade in Lviv, which was interrupted by the outbreak of the Second World War.

[a] Wrocław University of Economics and Business, Department of Econometrics and Operations Research, e-mail: Józef.Dziechciarz@ue.wroc.pl, ORCID: https://orcid.org/0000-0002-5261-0578.


Stanisława Bartosiewicz
Własnoręczny podpis właściciela(-ki) książki

AKADEMIA HANDLU ZAGRANICZNEGO WE LWOWIE
L. 18
My Rektor
Akademii Handlu Zagranicznego we Lwowie
stwierdzamy niniejszym, że
P. Bartosiewicz
Stanisława Maria Helena
urodzony (a)
w Brzezanach rok 1920
został (a) wpisany (a) w album studentów
Akademii Handlu Zagranicznego we Lwowie
Lwów, dnia -3 XII 1938 193
T. cz. Rektor

The need for gainful employment primarily marked the war period.

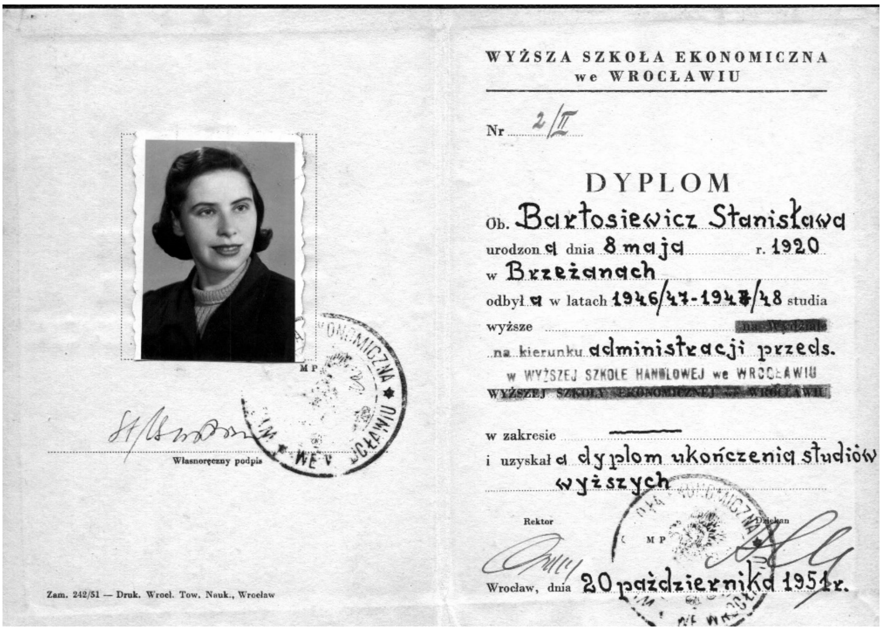

ARBEITS-AUSWEIS Nr. 3
Der (Die) durch nebenstehendes Lichtbild kenntlich gemachte
Stanislawa Bartosiewicz
(Vor- u. Zuname)
geboren: 8. Mai 1920
Wohnung: Heninga 18
ist bei dem von der Hauptgruppe Ernährung u. Landwirtschaft überwachten Betrieb:
DOSKA-FRANCK
KAFFEEMITTELFABRIK AKTIEN-GESELLSCHAFT
in Lemberg — Zolkiewska Str. 188
als Angestellte
beschäftigt.
Der Inhaber ist von der Anwerbung ins Reich befreit.
(Stempel des Arbeitsamtes)

Stanisława Bartosiewicz
Eigenhändige Unterschrift des Inhabers
(Stempel d. Dienststelle) DOSKA-FRANCK
KAFFEEMITTELFABRIK AKTIEN-GESELLSCHAFT
(Unterschrift)
Jeder Missbrauch des Ausweises ist strafbar.

In 1946, Professor Stanisława Bartosiewicz settled with her parents in Lower Silesia. She lived in the town of Żarów for the first six months and then moved to Wrocław.

While working, Professor Stanisława Bartosiewicz resumed her previously interrupted studies at the beginning of 1947, at a newly-established tertiary-level institution – the private Higher School of Commerce.

She completed her first-cycle studies in economics in 1949 as one of the first graduates of this University – her diploma was issued as a second such document awarded by this university.


WYŻSZA SZKOŁA EKONOMICZNA
we WROCŁAWIU

Nr 2/II

DYPLOM

Ob. Bartosiewicz Stanisława
urodzona dnia 8 maja r. 1920
w Brzeżanach
odbyła w latach 1946/47-1947/48 studia
wyższe [redacted]
na kierunku administracji przeds.
w WYŻSZEJ SZKOLE HANDLOWEJ we WROCŁAWIU
[redacted]
w zakresie ——
i uzyskała dyplom ukończenia studiów wyższych

Rektor Dziekan

M.P.

Wrocław, dnia 20 października 1951 r.

Własnoręczny podpis

Zam. 242/51 — Druk. Wrocł. Tow. Nauk., Wrocław

During her studies, as early as in 1947, at the invitation of her future mentor, Professor Jan Falewicz, she started combining her professional work outside the University with the function of a volunteer assistant at the Department of Business Economics, later Department of Statistics, at the Wrocław Higher School of Commerce. Her scientific interests, influenced by Professor Falewicz, focused on the application of quantitative methods in enterprise management. Since then, her entire life has become associated with the Wrocław University of Economics and Business, where she has obtained all possible levels of academic qualification and within the structures of which she performed many responsible functions

In 1952, she married Tadeusz Bartosiewicz.

In the difficult post-war years, Professor Stanisława Bartosiewicz combined her professional work with an assistantship at the University and her studies. She gained a master's degree in 1953, completing her second-cycle education at her Alma Mater, which in 1950, after undergoing nationalisation, changed its name to the Higher

School of Economics in Wrocław. Her master's thesis was entitled *Regression analysis as a tool for the assessment of the economic efficiency in an enterprise*. In the years 1953–1957, she worked as a senior assistant at the Central Institute of Scientific and Technical Documentation in Warsaw. During that period, Professor Bartosiewicz's two daughters, Anna and Ewa were born.

In 1953, after returning to Wrocław, she started working as an assistant professor at the Wrocław University of Economics and Business (Pol. Akademia Ekonomiczna) in the Department of Statistics, headed by Professor Jan Falewicz. She obtained a doctorate in economics in 1962 at her Alma Mater, with a dissertation entitled *Adequacy of indicators characterising the activity of enterprises*.

In 1966, she was appointed associate professor at the Institute of Economic Accounting Methods. At the same time, she assumed the position of the head of the Econometrics Unit (later Department of Econometrics). She remained the head of the Department of Econometrics until her retirement in 1990.

On 2 February 1984, Professor Bartosiewicz received a postdoctoral degree in economic sciences (Pol. habilitacja). In 1988, she was awarded the title of Professor of Economic Sciences.

Below is an outline of Professor Bartosiewicz's employment record:

25 March 1947 – 31 Dec. 1947 – assistant volunteer;

1 Jan. 1948 – 31 Dec. 1949 – assistant;

1 Jan. 1950 – 30 Aug. 1953 – senior assistant;

1 Nov. 1953 – 30 Nov. 1957 – senior assistant (Warsaw);

1 Dec. 1957 – 31 Dec. 1957 – senior assistant;

1 Jan. 1958 – 31 Oct. 1968 – assistant professor;

1 Nov. 1968 – 31 May 1988 – associate professor;

1 June 1988 – 30 Nov. 1990 – professor;

1 Dec. 1990 – 30 Sept. 2006 – part-time professor.

2. The person

Professor Stanisława Bartosiewicz has been a mentor to many students. She pays attention to the individuality and development of a scientific personality. She is always ready to help and offer sound advice, especially being able to interpret complicated quantitative methods quickly and identify an adequate practical approach, which are her particularly appreciated gifts.

She is valued and held in high esteem by Polish econometricians and statisticians. She was elected member of the Statistics and Econometrics Committee of the Polish Academy of Sciences several times.

Her numerous distinguishing abilities include her unique talent for leadership, which is reflected in her own demeanour, not only limited to giving orders. Her colleagues and students know what kind of behaviour and actions are expected and appreciated and which are unwelcomed. The professor needs not articulate her expectations, as she herself sets an excellent example to follow.

An equally distinctive feature of her character is the rare ability to focus solely on a problem, not letting any personal sympathies or animosities towards the person presenting the problem to interfere with the process.

The great mind of Professor Stanisława Bartosiewicz is widely recognised. Professor Juliusz Siedlecki, in his address to the scholar expressed the wish for her wise advice and assistance to continue benefitting him throughout his further career. In this context, he referred to Seneca's thought: *supporting and helping is a testimony of a noble and wonderful mind. Whoever is such a benefactor, imitates God.*

Professor Janusz Łyko considers Stanisława Bartosiewicz a role model of a dedicated professor, an example of high-standard decency, a recognised scientist and teacher, a person of an exceptionally vivid mind.

Professor Bartosiewicz's friends and colleagues underline her life wisdom and understanding. As a scientist, she lets the principle of self-verification of previously created theories, concepts or definitions guide her actions, which is what makes a scientist a researcher.

Her life philosophy which stresses the necessity of dialogue and search for compromise-based solutions, is what guides the activities of her numerous followers, including the academic community of the Wrocław University of Economics and Business. Thanks to this inspiration, they developed a habit to look for solutions to any arising issues together, in the course of group discussion. Professor Bartosiewicz's unique skills, such as her widely-recognised ability to analyse and accurately generalise notions and situations, her perceptiveness, talent to motivate others to engage in creative activities and to conduct productive discussions, her problem-solving abilities as well as scientific intuition, are all greatly appreciated and admired. The Professor's credo: *don't talk about forms, talk about ideas*, accurately illustrates her general approach to both science and life.

Professor Stanisława Bartosiewicz has a great sense of humour, which enables her to see people, phenomena and actions in a proper perspective. She tends to apply this sense of humour to her professional life as well, which is reflected, for example, in the title of her book: *Econometrics with a pinch of salt*, a book on complicated issues relating to econometrics.

She has always sought opportunities for self-development and broadening her knowledge. With her distinct self-distance and in a playful tone, she spoke of her latest achievements, which included learning to use a computer at the age of 84.

Professor Bartosiewicz displays her great sense of humour also in private life. She is known for being an expert in eastern borderland jokes, which she tells with great eloquence and in original accent.

Professor Bartosiewicz has a rich life history. She shared some of her life experiences in her autobiography entitled *Crumbs of life.*

In the book, she tells the story of how Professor Jan Falewicz offered her the position of an assistant: she attended his lectures on *Business Economics,* which were essentially about the econometric analysis of costs. After reviewing her exceptionally diligent notes from the course, Professor Falewicz invited Professor Bartosiewicz to become his assistant.

Her ability to write structured poems based on Japanese patterns is also widely admired.

When analysing the sources of satisfaction and success in life, Professor Bartosiewicz mentions such factors as scientific intuition, experience and teaching skills. She says that her ongoing participation in academic life gives her much strength and fosters new interests. Meeting friends inspires her to engage in new activities, such as writing books.

3. The scholar

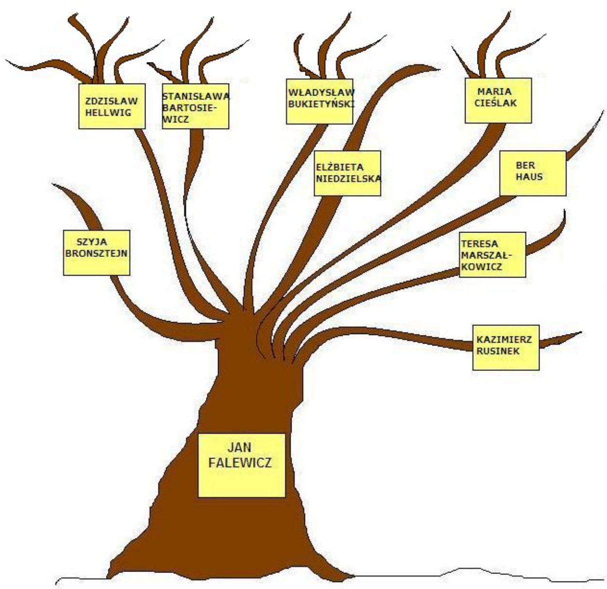
Professor Stanisława Bartosiewicz is one of the most outstanding scientists in the field of econometrics in Poland. Her research interests were shaped during her studies, under the influence of Professor Jan Falewicz, who, along with Paweł Ciompa, is a pioneer in micro-econometric research. She is considered the founder of the Scientific School of Econometric Modelling. Her theoretical and methodological research included econometric modelling, multivariate statistical analysis and methods of decision-making under uncertainty.

In an attempt to characterise her extraordinary scientific achievements, two general trends may be distinguished: theoretical and methodological, and application. As mentioned before, the practical part of the research interests of Professor Bartosiewicz has been influenced by her mentor, Professor Jan Falewicz, who, as she often recalls, conducted the econometric analysis of costs. He formulated pioneer proposals for the application of mathematical tools in enterprise management. The subject of cost analysis has remained of particular interest to Professor Bartosiewicz throughout the entire period of her scientific and research activity. She is the author of a chapter on econometric analysis of costs in the textbook *Econometrics. Methods and analysis of economic problems*. She always stresses the fact that costs is an area of a company's operations, which involves numerous management elements, requiring the application of a variety of quantitative methods.

During her studies under the supervision of Professor Falewicz, Stanisława Bartosiewicz worked on the construction of Clark's cost budget at the Pafawag State Wagon Factory in Wrocław. It was then that her scientific interests began to develop to their fullest. Her scientific interests focused on the search for optimal statistical (one could even say, econometric) methods. Those methods were considered tools serving to control and program the economic activity of enterprises. Her bachelor's thesis entitled *Criticism of Clark's system* and master's thesis entitled *Regression analysis as a tool for examining the economic efficiency of enterprises* reflected these interests.

Further research relating to the indicated field was presented in the published doctoral dissertation entitled *Adequacy of indicators characterising the activity of an enterprise* (1962). Her doctoral dissertation, and especially its first part, entitled *Theoretical issues,* constitutes an original systemic view of an enterprise's activity as a relatively isolated element of a more extensive system of the national economy, which places the study within the scope of cybernetics. The whole dissertation (the above-mentioned first part and the second part, entitled *Statistical issues*), was published as a monograph entitled *On the correctness of the construction of indicators characterising the activity of an enterprise.*

The most important areas of Professor Bartosiewicz's theoretical and methodological research include: econometric methods, the decision-making theory, statistical multidimensional comparative analysis, input-output methods as well as the mathematical theory of organisation. In the field of econometric modelling, Professor Bartosiewicz is the author of influential and pioneering publications, such as the monograph entitled *Econometrics. The technology of econometric information processing* (1976, 1989). In this book, she outlines the entire procedure of processing economic information by means of an econometric model. Along with an analysis of standard econometric methods, she presents techniques which constitute an original contribution to the theory of econometrics.



One of the most useful solutions the author introduces is a graph-based method of selecting explanatory variables for the econometric model, which is supplemented with a technique of choosing variables in nonlinear models. The most significant concept here concerns constructing replacement variables (proxies), i.e. an algorithm meant to substitute variables that cannot be measured directly with proxy variables in the form of composite indicators obtained by methods of multivariate comparative analysis. Professor Bartosiewicz's other valuable input includes devising a procedure for constructing a composite indicator which maintains the original value of the variance. The most utilitarian approach is the technique for selecting a model function class based on the visual assessment of empirical regression with projections of points on the coordinate plane. Among her other significant achievements is the introduction of a modification which simplifies the testing of autocorrelation with the Student's test.

The essential part of Professor Bartosiewicz's scientific activity is combining the econometric modelling theory with practical economic applications, especially in micro-econometrics, which is applied e.g. in a monograph entitled *On the correctness of the construction of indicators characterising the company's activity* (1965). In this work, she introduces a cybernetic, relatively isolated system of an enterprise.

Professor Bartosiewicz's particularly notable achievement is ordering the procedure of the verification of econometric models by arranging the criteria of goodness of the model. She developed a method of selecting the analytical form of a model with several explanatory variables and formulated procedures for testing the residuals' symmetry and autocorrelation.

The second important monograph relating to this field is entitled *Specificity of econometric models and their application in the analysis of socio-economic phenomena* (1987), which contains a summary and a review of the professor's original scientific achievements. Additionally, it includes interesting reflections on the application of econometric models to economic practice.

Professor Bartosiewicz has also done research on multidimensional statistical analysis and methods of decision making under uncertainty. In the work *Elements of economic calculus* (1978, co-authorship), she presented an original approach to multi-criteria mathematical programming based on the concept of the game theory.

Another significant area of her research is the application of mathematical tools in business management, mainly in terms of cost analysis. A reflection on this part of her research is included in the book *Econometrics. Methods and Analysis of Economic Problems* (1998), which Professor Bartosiewicz co-authored.

As far as econometrics is concerned, Professor Bartosiewicz introduced decision-making, balance sheet accounts (input-output analysis) and organisation-related (network analysis) issues into the Polish theory and practice. Particularly noteworthy are three papers related to the PERT method, in which she provides original algorithms for solving network problems. Her research results regarding the PERT method were published in a series of articles in the Statistical Review and Scientific Journals of the Wrocław University of Economics and Business. These include: *On the technique of applying the PERT method* (1966), *On ordering nodes in the PERT network* (1966), *Contribution to the technique of applying the PERT method* (1967).

The original scientific achievements of Professor Bartosiewicz include repealing the assumption of simple proportionality of the dependence of input and output tables and the transfer of the technique of flow analysis onto a microscale, i.e. the enterprise.

An undoubtedly significant publication of Professor Bartosiewicz is her book *Wrocław Econometrics*, which includes a complete list of scholars working in the widely-defined field of econometrics and linked to tertiary institutions in Wrocław. Moreover, the book contains a comprehensive overview of the achievements of the scientists involved in developing econometric methods within the Wrocław academic centre.

A distinctive feature of her scientific work is keeping theoretical considerations simple, communicative and practically useful. She withstands the current tendency of presenting simple notions in a complicated manner, i.e. she makes sure that complex scientific problems are presented in a simple and approachable form. It is manifested, for example, in the way she presents computational algorithms in the form of procedures, providing a comparatively easy and convenient means to computerising econometric research.

Professor Stanisława Bartosiewicz has always sought to utilise quantitative methods in specific practical applications. The most notable achievements in this area include methods of settling spare parts in repaired vehicles, planning the optimal size of a warehouse and a system of technical and economic indicators for the management of the auxiliary economy.

Professor Bartosiewicz presents the connection of theoretical tools with economic applications in her acknowledged book entitled *Specification of econometric models and their use in the analysis of socio-economic phenomena* (1987).

Professor Bartosiewicz's second most important area of research activity is multidimensional statistical analysis. She concentrates especially on one of its sections, namely on multidimensional comparative analysis. Methodological achievements of Professor Bartosiewicz include devising a method of determining the path of proportional development, the technique of composite indicators construction, the algorithm for determining subsets of similar objects, whose similarity criterion is a parallel line of regression, as well as a cybernetic system of an isolated enterprise.

Professor Bartosiewicz also researched mathematical methods of decision-making, nowadays referred to as description operations research. In particular, she devised a macro-scale decision-making tool using elements of the game theory. These propositions were published in *Elements of economic calculus*, a book co-authored by the Professor and published in three editions.

It is worth mentioning that Professor Bartosiewicz is always eager to participate in scientific conferences, during which she provides young scientists reporting on their research with perceptive and kind comments, as she assumes a practical approach and is able to interpret complicated quantitative methods quickly and efficiently. Her favourite conferences are those held in Zakopane, Toruń and Szczecin. In 2019, she prepared and presented several lectures for some of the planned seminars.

4. The teacher

Professor Stanisława Bartosiewicz can boast outstanding achievements in teaching and education-related activities. She is well known to every employee of the Wrocław University of Economics and Business. During their studies, some of them had the privilege of attending her classes and lectures.

At the turn of the 1950s and 1960s, the subject of econometrics appeared in the curricula of economic schools for the first time. Professor Bartosiewicz initiated lectures on this subject at the University. Jointly with Professor Zdzisław Hellwig, she wrote a textbook entitled *Representative method*, which at that time was considered a very modern publication. She was the author of a section entitled *Selected issues in econometrics*. Professor Bartosiewicz is also the co-author of a textbook published by PWE (three editions) entitled *Outline of econometrics*.

Stanisława Bartosiewicz has an outstanding record of teaching achievements. She taught several different subjects, among which was the theory of statistics, industrial statistics, representative methods, econometrics with elements of input-output analysis techniques, mathematical programming, as well as multidimensional comparative analysis.

Her didactic work involved not only conducting classes, but also preparing teaching aids – she is the author and co-author of almost twenty manuals and textbooks.

Professor Bartosiewicz was the supervisor and reviewer of several hundred master's and bachelor's theses.

5. A Mentor

Just as Professor Jan Falewicz was a source of inspiration throughout her scientific career, Professor Bartosiewicz went on to inspire her own students (and she still does so), stimulating their individual development. Professor Bartosiewicz is the founder of the scientific school of econometric modelling developed at the Wrocław University of Economics and Business. The research trends she initiated sparked her students' scientific interests.

The Professor's outstanding teaching accomplishments are reflected in the academic achievements of her own students: thirteen obtained a doctorate under her supervision, five doctoral students received the title of professor and two of habilitated doctor. She reviewed thirty-six doctoral dissertations and nine habilitation theses.

The professor always focuses on individuality and the development of a scientific personality. She never attempts to impose research topics, but rather offers her assistance and valuable scientific advice.

Wiesław Pluta has been developing research on the use of quantitative methods in corporate finance, which he discusses in his articles, including *Multidimensional comparative analysis in econometric modelling* (1986), or *Financial planning in an enterprise* (1999). Edward Nowak focuses on the use of quantitative methods in accounting and is the author of several related publications, e.g. *Taxonomic methods in the classification of socio-economic objects* (1990), *Theory of costs in enterprise management* (1996), or *Accounting in enterprise controlling* (1996, co-authorship). Józef Dziechciarz and his team continue to examine the issues of econometric modelling, especially modelling in conditions of heterogeneity of data sets, modelling based on qualitative data and the development of the robust regression techniques, and their findings are described in publications including *Econometric modelling of economic processes: models with variable and random parameters* (1993), or *The decision-making process support in the economy: econometric models with variable and random parameters as a simulation and analysis tool* (1995, co-authorship). The research interests of Krzysztof Jajuga focus on the broadly-understood problems of the financial market. His scientific work is described in articles entitled *Capital Management* (1993), *Investments: financial instruments, financial risk, financial engineering* (1998, co-authorship). Ludmiła Waszkiewicz continued to develop her scientific career in the field of health care management and biostatistics, while Jerzy Jakubczyc works in the finance and banking sector. All the above-mentioned scientists are professor Bartosiewicz's former students.

It is worth noting that Professor Bartosiewicz's work also inspires the youngest generation of researchers. A group of over fifty authors of doctoral theses, whose supervisors were once her students, undertook the issues of econometric modelling in their scientific work.

While being the dean, Professor Bartosiewicz informally oversaw numerous post-doctoral and doctoral students in the final stage of their habilitation or doctoral dissertation processes. Prior to the habilitation colloquium or the defence of the doctorate she offered her advice which proved invaluable, very insightful and kind, and has helped many doctoral and postdoctoral students to present their achievements to the best of their ability.

6. The organiser

Professor Stanisława Bartosiewicz has been inextricably associated with the Wrocław University of Economics and Business. During the establishment of the Wrocław Higher School of Commerce (today's Wrocław University of Economics and Business) in 1947, she was among its first students and employees.

Professor Bogusław Fiedor, former rector of the Wrocław University of Economics and Business, said that Professor Bartosiewicz, a student of the Lviv Academy of Foreign Trade, played a unique role in the history of their University. She is a person who unifies the Wrocław and the Lviv community. The Lviv academic community is a source of tradition for the city and its scientific development. The rector had the opportunity to learn about the professor's teaching skills in practice, as he attended her statistics classes during his studies. Thanking for all the years Professor Bartosiewicz worked for the University, he wished that she would be able to continue to add her tremendous intellectual and physical potential to the further development of the school, and that for many years to come her students and colleagues would be able to continue to experience her extraordinary humour, kindness and positive energy.

Professor Stanisława Bartosiewicz is one of the greatest contributors to the development of the Wrocław University of Economics and Business, starting in 1947 and continuing to this day. Throughout this period she has carried out several responsibilities.

Professor Bartosiewicz was actively involved in the creation and development of the Faculty of Management and Informatics, which evolved into the Faculty of Management, Informatics and Finance, at the Wrocław University of Economics and Business. It was in the early 1970s that Professor Bartosiewicz participated in the development of a new programme and organisational concept of the University, which led to the establishment of a new unit – the Faculty of Management and Informatics. In the years 1976-1990, she was the deputy dean and dean of this faculty. She was the deputy dean since the establishment of the faculty until 1984, and then she held the position of dean until 30 November 1990.

From the 1960s until her retirement in 1990, Professor Bartosiewicz was a member of the Faculty Council and the Senate at the University's Parliament.

Her research work, as mentioned before, inspired scientific interests of her students. As a result, she created the Department of Econometrics and headed it until her retirement.

Professor Stanisława Bartosiewicz managed research programs for many years. Moreover, she was the head of many scientific and research studies, often carried out for business practice. The vast majority of the findings could be directly applied in real life, including the already-mentioned system of settling spare parts in repaired motor vehicles, the planning of the optimal size of a warehouse, and the system of technical and economic indicators for the management of the auxiliary economy.

During her work at the University, Professor Bartosiewicz held many positions, which included serving as head of the consultation point in Jelenia Góra.

Professor Andrzej Gospodarowicz, dean of the Faculty of Management and Informatics in the years 1990–1996, recalled numerous occasions when Professor Bartosiewicz offered her valuable advice and enormous organisational assistance upon his assuming the function of the dean. At that time, a new law on higher education came into force and so the acquaintance of the previous dean, Professor Bartosiewicz, proved invaluable in the process of organising the scientific and didactic life in that challenging period.

7. Expressions of respect and recognition of success

Professor Stanisława Bartosiewicz received several decorations and distinctions awarded by the state, the scientific community and other institutions, for her outstanding achievements.

The highest academic title that an academic teacher may earn is an honorary doctorate, *doctor honoris causa*. The Senate of the Wrocław University of Economics and Business decided to award this honourable title to Professor Stanisława Bartosiewicz. The ceremony was held in the Senate Hall of Wrocław University of Economics and Business in Wrocław, on 12 March 2020, at 10 a.m.

The procedure of awarding the *doctor honoris causa* title to the professor consisted of the following stages: on 4 July 2019, the Faculty Council of the Faculty of Management, Informatics and Finance of the Wrocław University of Economics and Business appointed Professor Józef Dziechciarz (chairman), Ewa Stańczyk-Hugiet and Edward Nowak the Commission members to process the application to award Professor Stanisława Bartosiewicz with the title of *doctor honoris causa* of Wrocław University of Economics and Business.

On 12 September 2019, the Faculty Council of the Faculty of Management, Informatics and Finance of the Wrocław University of Economics and Business resolved to initiate the procedure to award Professor Stanisława Bartosiewicz with the honorary doctorate from the Wrocław University of Economics and Business. The reviewers of Professor Bartosiewicz's scientific achievements were: Teodor Kulawczuk, PhD, DSc, ProfTit, professor emeritus of the University of Gdańsk and Józef Pociecha, PhD, DSc, ProfTit, from the University of Economics in Krakow. Krzysztof Jajuga, PhD, DSc, ProfTit, Doctor Honoris Causa was appointed the promoter in the proceedings for the granting of the honorary title.

On 19 December 2019, the Senate of the Wrocław University of Economics and Business, after reviewing the Faculty Council's request, passed a resolution to confer the title of an honorary doctor of the Wrocław University of Economics and Business on Professor Stanisława Bartosiewicz.

Professor Bartosiewicz also holds the honour of the Lady Knight of the Order Golden Cross of Merit, the Knight's Cross and the Officer's Cross of the Order of Polonia Restituta. She was also awarded the National Education Commission Medal and the Teacher Badge of Merit.

On 21 May 2019, Dominik Rozkrut, PhD, President of Statistics Poland, awarded her with an honorary badge of merit for her contribution in the development of statistics of the Republic of Poland.

Professor Bartosiewicz also holds regional awards of the Builder Badge of Wrocław and Badge of Merit of Lower Silesia.

The Wrocław University of Economics and Business honoured Professor Stanisława Bartosiewicz with the title of Honorary Professor of the Wrocław University of Economics and Business.

She also holds the Crystal Alumnus title. It is awarded on behalf of the Wrocław University of Economics and Business, as a form of recognition for the Professor's numerous and longstanding achievements in research and teaching, as well as for her committed work for the academic community.

Her Alma Mater conferred on her the Medal of Merit of the Wrocław University of Economics and Business.

Acknowledgements

In the preparation of this text I used words of several persons. The list includes (featured in alphabetical order): Bogusław Fiedor, Andrzej Gospodarowicz, Krzysztof Jajuga, Stanisław Krawczyk, Teodor Kulawczuk, Janusz Łyko, Walenty Ostasiewicz, Józef Pociecha, Juliusz Siedlecki, Lucyna Wasylina. I am grateful to them for their support in providing any data, information, assessments and evaluative formulations. Many thanks. A part of the information provided in this paper was drawn from the numerous conversations and discussions I was privileged to have with the distinguished honouree, Professor Stanisława Bartosiewicz.

References

Bartosiewicz, S. (2006). *Okruchy życia*. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego.

Bartosiewicz, S. (2007). *Ekonometria wrocławska*. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego.

Bartosiewicz, S., Ostasiewicz, W. (1997). School of quantitative methods. *Argumenta Oeconomica*, (1), 73–81. https://www.dbc.wroc.pl/Content/14800/download/

Dudycz, H. (Ed.). (2001). *Księga pamiątkowa Wydziału Zarządzania i Informatyki*. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego.

Ostasiewicz, W. (Ed.). (2000). *Katedra Statystyki Akademii Ekonomicznej we Wrocławiu 1950–2000*. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego.

UE we Wrocławiu. (2020a). *Doktorat honoris causa UEW dla prof. Stanisławy Bartosiewicz.* https://youtu.be/UKrT7HsArzk

UE we Wrocławiu. (2020b). *Profesor Stanisława Bartosiewicz doktorem honoris causa naszej uczelni.* http://www.portal.ue.wroc.pl/nauka/21402/profesor_stanislawa_bartosiewicz_doktorem_honoris_causa_naszej_uczelni.html#.X2HDVIvgqUk

UE we Wrocławiu. (2020c). *Profesor Stanisława Bartosiewicz ma 100 lat.* http://www.ue.wroc.pl/aktualnosci/21693/profesor_stanislawa_bartosiewicz_ma_100_lat.html#.X2HFrIvgq60