Acta Universitatis Lodziensis

FOLIA OECONOMICA

4(349)
2020

WYDAWNICTWO UNIWERSYTETU ŁÓDZKIEGO

COPE
Member since 2018
JM13703

Acta Universitatis Lodziensis

FOLIA OECONOMICA

4(349)
2020

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

Acta Universitatis Lodziensis

4(349)
2020

FOLIA OECONOMICA

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

Łódź 2020

COPE
Member since 2018
JM13703

REDAKTOR NACZELNY

Mariusz Plich

ZASTĘPCA REDAKTORA NACZELNEGO

Agnieszka Rossa, Ewa Kusideł

REDAKTORZY TEMATYCZNI

Jerzy Czarnecki, Czesław Domański, Jacek Gad, Jerzy Gajdka, Beata Guziejewska Magdalena Kalińska-Kula, Piotr Krajewski, Jakub Marszałek, Janusz Świerkocki

REDAKTOR STATYSTYCZNY

Alina Jędrzejczak

RADA PROGRAMOWA

Barbara Błaszczyk (Instytut Nauk Ekonomicznych PAN), *Krystyna Gawlikowska-Hueckel* (Uniwersytet Gdański), *Grażyna Gierszewska* (Politechnika Warszawska), *Marian Gorynia* (Uniwersytet Ekonomiczny w Poznaniu), *Marc Lautier* (Uniwersytet w Paryżu), *Cedric Lesage* (HEC Paryż), *Elżbieta Mączyńska* (SGH), *Bogdan Nogalski* (Uniwersytet Gdański), *Witold Orłowski* (Szkoła Biznesu Politechniki Warszawskiej), *Eulalia Skawińska* (Uniwersytet Zielonogórski) *Roman Sobiecki* (SGH), *Sławomir Sojak* (Uniwersytet Mikołaja Kopernika w Toruniu) *Matti Viren* (University of Turku)

SEKRETARZE NAUKOWI

Jurand Skrzypek, Katarzyna Walerysiak-Grzechowska

SEKRETARZ WYDAWNICZY

Tamara Franiak-Krupińska, Aleksandra Przybył

REDAKTORZY JĘZYKOWI

Emilia Poisson (język angielski), *Grzegorz Rudziński* (język polski)

RECENZJA

double-blind review

© Copyright by Authors, Łódź 2020
© Copyright for this edition by Uniwersytet Łódzki, Łódź 2020

ISSN 0208-6018
e-ISSN 2353-7663

Czasopismo afiliowane przy Wydziale Ekonomiczno-Socjologicznym Uniwersytetu Łódzkiego

Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663
www.czasopisma.uni.lodz.pl/foe/
4(349) 2020

Spis treści

Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663
www.czasopisma.uni.lodz.pl/foe/
4(349) 2020

DOI: http://dx.doi.org/10.18778/0208-6018.349.01

Przemysław Pomykalski
Łódź University of Technology, Faculty of Management and Production Engineering
Department of Management Systems and Innovation, przemyslaw.pomykalski@p.lodz.pl

Piotr Filipiak
Koźmiński University, piotr.filipiak@me.com

Factors Influencing IPO Underpricing in Poland

Abstract: We review the theory and evidence on IPO activity and underpricing focusing on the Warsaw Stock Exchange. Although the topic has been under investigation in the past, we believe that the recent decade of low interest rates deserves inquiry. We research the extent of underpricing during this period and further conclude that three factors had a statistically significant influence on initial public offering underpricing during this period: the year of IPO, risk-free rate and WIG close value.

Keywords: IPO, underpricing, Poland, behavioural finance

JEL: G11, G15, G24

1. Introduction

Companies that sell shares on the stock exchange for the first time (Initial Public Offering) advertise and offer incentives to new shareholders. To attract investors, shares are usually offered at a price which is lower than that resulting from valuations. The procedure is called "underpricing". The size of the incentives and pre-IPO valuations are confidential and as such difficult to research. To assess underpricing, an indicator of Initial Return (IR) is calculated based on the price at the end of the first day of trading and the price at which shares were sold to new investors.

Data gathered by Jay Ritter (2018) indicate that IPO underpricing in the United States fluctuates substantially, averaging 21.2 in the 1960s, 7.1% in 1970s, 6.9% in the 1980s, 21% in the 1990s, and 21.1% in the years 2001–2017[1]. Ritter further indicates that underpricing depends upon the size of the company issuing shares and prior financing sources. Initial return decreases with the size of the company (measured in revenues). Underpricing is also higher for venture capital backed companies than it is for growth and buy-out funds financed companies.

Underpricing on the Warsaw Stock Exchange also varied. Earlier research was conducted by Aussenegg (2000: 69–99), Lyn and Zychowicz (2003: 181–195), Jewartowski and Lizińska (2012: 59–75), Sieradzki (2013), Lizińska and Czapiewski (2015: 112–125). Higher underpricing in the 1990s is attributed to privatisation processes in the Polish economy. After the year 2000, underpricing decreased. Czapiewski et al. (2014) indicate underpricing of 34.1% in the period 1991–2000 and 13.5% in the period 2001–2011. Pomykalski and Domagalski (2015) reported 11.89% in the period 2005–2013. Going a step further, Wołoszyn and Zarzecki (2013: 121–135) researched the impact of "the January effect" on IPO underpricing in Poland.

Similar research has been conducted in other countries by Chowdhry and Sherman (1996: 359–381), Habib and Ljungqvist (2001: 433–458), Banerjee, Dai and Shrestha (2011: 1289–1305), Chan, Wang and Wei (2004: 409–430) in China, Cassia et al. (2004: 179–194) in Italy, Chambers and Dimson in the UK (2009: 1407–1443), Mezhoud and Boubaker in France (2011: 166–180), Ganesamoorthy and Shankar in India (2013: 36–38).

In this study, we investigate underpricing in IPOs on the Warsaw Stock Exchange (WSE) between 2005 and 2016. Since there is sufficient scientific evidence to assume that IPO underpricing existed in this period and that it is likely to exist in the future, these results are used as a background for further analysis of factors

1 Based on data available on Jay Ritter's website: https://site.warrington.ufl.edu/ritter/ipo-data/ [accessed: 1.04.2018].

influencing the extent of underpricing in this period. We analysed three groups of potential factors and focused on the impact of three: the year of IPO, risk-free rate and WIG close value.

We admit that findings of this study are limited to the WSE and one economic cycle.

We believe that the results of this study may be interesting for analysts, investors, consultants and managers involved in IPOs. The researched period was characterised by low interest rates and we believe that this factor makes our research and its results worth considering.

2. Dataset and methodology

In this paper, we investigate IPOs of 349 companies quoted on the Main Market of the Warsaw Stock Exchange that took place between 2005 and 2016. Only companies which offered their shares to the public for the first time were taken into consideration. Firms previously quoted on the WSE or another market were excluded from the sample.

Table 1. Number of IPOs in the years 2005–2016 (adjusted for companies previously quoted)

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
No of IPOs included	35	37	81	33	13	26	29	15	16	25	25	14

Source: own elaboration

Data for the WIG – Warszawski Indeks Giełdowy (Warsaw Stock Exchange Index) – are published by the Warsaw Stock Exchange. During the researched period, the number of IPOs and WIG closing values were loosely correlated. The number of public offerings dropped after 2008 and remained lower than in prior years even though the stock market recovered and reached higher valuations. The reasons for that may vary but the primary reason may be that better access to debt financing combined with low interest rates seem to decrease the attractiveness of public equity financing.

Our research results indicate that average underpricing on the WSE in the years 2005–2016 was 12.35%. Further analysis indicates that underpricing changes over time, with the size of the offer, and due to other factors.

To assess underpricing, we used two methods. Using the first method, we compared the first day closing price to the offering price. Using the second method, we subtracted broad market index change.

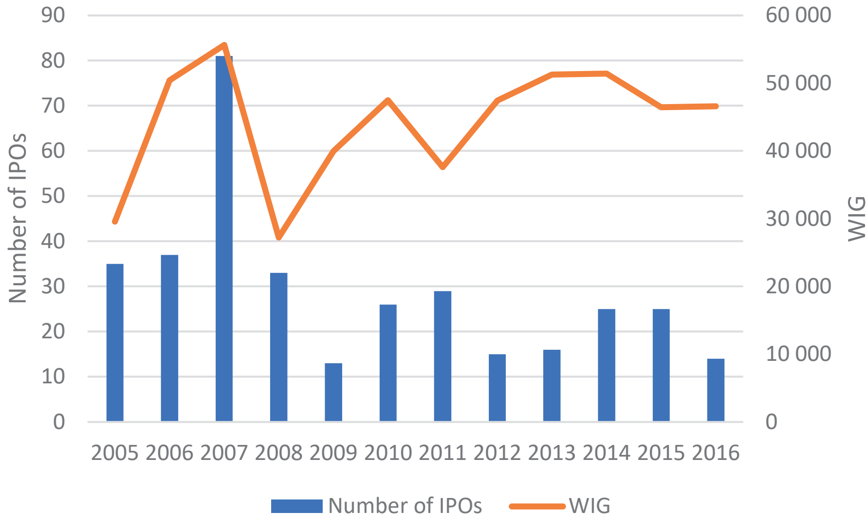


Figure 1. WIG and number of IPOs

Source: own elaboration

To assess underpricing, a ratio of initial return (IR) is calculated (Saunders, 1990: 3–12; Fijałkowska, Muszyński, Pauka, 2013: 415–426).

$$IR_{i,t} = \frac{PC_{i,t} - PO_i}{PC_{i,t}} \cdot 100\%, \quad (1)$$

where:

$IR_{i,t}$ – initial return (return on the first day (t) the company (i) was traded),
$PC_{i,t}$ – first day (t) closing price of company (i) shares,
PO_i – offering price of the company (i).

In the second step, the broad market index change is subtracted from the initial return. In this paper, we used Warszawski Indeks Giełdowy (WIG), the main index of the Warsaw Stock Exchange. This method is less popular than the first one but, according to Hunger (2003), Sieradzki (2013: 1–37) and Czapiewski et al. (2014: 571–590), it provides more reliable results, as stock market (index) changes may impact first day closing prices.

$$IAR_{i,t} = IR_{i,t} - IR_t^{WIG}, \quad (2)$$

where:

IAR – initial adjusted return on the day (t) that the company (i) debuted,
$IR^{WIG}_{i,t}$ – return of broad market index on the day (t) that the company (i) debuted.

Any list of factors influencing underpricing is subjective. The list of factors used in this research was created from most commonly used statistics in publications of the WSE.

Table 2. Factors affecting underpricing

Factor	Description
Year	The year of IPO.
Offering Price	Initial price of one share.
New issue [%]	Part of the company that was offered to the public in percentage.
IR [%]	Initial rate of return (underpricing).
WIG change [%]	Percentage change of WIG on the day that the company was offered.
WIG close	Value of broad market index in base points.
Offering value	Total offering value of the company in PLN.
Risk-free rate	Risk free rate on the day that the company was offered.
Number of IPOs	Number of IPOs in the year that the company was offered.

Source: own elaboration

As a risk-free rate, we have taken the reference rate of the National Bank of Poland.

Table 3. Reference rate (on 31 December)

Year	2005	2006	2007	2008	2009	2010
Reference rate	6.00%	4.00%	4.00%	5.70%	3.81%	3.50%
Year	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Reference rate	4.15%	4.55%	2.73%	2.34%	1.54%	1.50%

Source: National Bank of Poland data

We have chosen ordinary least squares methodology to look for correlations between underpricing and factors. This methodology was previously used in underpricing examination of IPOs on New Connect by Fijałkowska, Muszyński and Pauka (2013: 415–426).

We used Gretl open source statistical package to perform the analysis. IR was the dependent variable. Factors listed in Table 2 were used as regressors.

3. Results

3.1. Underpricing

Average underpricing (initial return) of the 349 companies which debuted on the Warsaw Stock Exchange from 2005 to 2016 was 12.35%. In comparison to 2005–2013 (Pomykalski, Domagalski, 2015: 117–131), the average rate of return on the first day increased by 0.46%.

The average adjusted IR (IAR) underpricing of IPOs is 11.84%, which is smaller than IR. Decreasing change of the adjusted IR is expected. It means that IPOs were increasing their value more when there was a positive change of WIG. According to the calculation performed, the average influence of WIG change on the closing price is 0.51%.

Table 4. Analysis of Initial Return (IR) and corrected Initial Return (IAR)

	IR	IAR
Average	12.35%	11.84%
Median	4.71%	4.15%
Minimum	–74.07%	–72.20%
Maximum	481.33%	480.07%
Standard deviation	36.71%	36.32%
Skewness	7.34	7.42
Curtosis	81.52	82.83
IPOs with negative initial returns	94.00	100.00
IPOs with positive initial returns	238.00	249.00
IPOs with initial returns equal to 0	17.00	0.00

Source: own elaboration

The median for IR is 4.71% and it is a lot lower than the average which is 12.35%. The same situation is for the corrected IAR – 4.15% compared to 11.84%.

The distribution is far from normal (curtosis of 81.52 for IR and 82.23 for IAR). Skewness is positive (7.34 for IR and 7.42% for IAR), which means there are more results above the average than below the average values.

Almost 27% of the offers were overpriced (29% if IAR is used). 238 companies brought a positive initial return while considering IR, which is 68% of all IPOs (249 companies or 71% if IAR is used). The majority of IPOs are underpriced and results indicate the existence of underpricing on the Warsaw Stock Exchange.

The level of underpricing varies. In 27% of the cases, initial return was negative and in one case it amounted to –74.07%. This means that investing in IPOs on the WSE in hope of exploiting underpricing is associated with risk and may bring disappointing results.

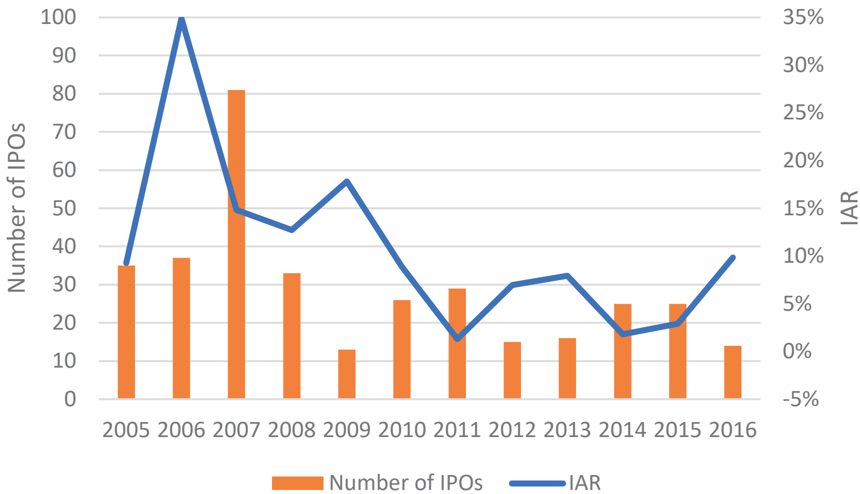


Figure 2. Number of IPOs and IAR in the years: 2005 to 2016

Source: own elaboration

During the researched period, IAR was inversely proportional to the number of IPOs in individual years (Figure 2). This confirms similar conclusions by Henricson (2012: 1–45) on the Swedish market and Chi and Padgett (2005: 71–93) on the Chinese market.

3.2. Analysis of independent variables

Table 5. Summary statistics

	Offering price	New issue	WIG change	WIG close	Offering value	Risk free rate	Number of IPOs
Average	28.759	0.750	3.378e-05	47351,000	2.539e+08	0.039	40.728
Median	13.330	1.000	0.001	47872,000	4.240e+07	0.040	33.000
Minimum	0.510	0.000	–0.043	24333,000	0.000	0.015	13.000
Maximum	539.500	5.500	0.043	67289,000	1.072e+10	0.060	81.000
Standard deviation	57.629	0.457	0.010	10209,000	9.760e+08	0.013	23.174
Skewness	5.436	3.247	–0.451	–0.353	7.310	–0.159	0.958
Curtosis	35.119	37.975	2.375	–0.322	60.633	–0.462	–0.608

Source: own elaboration

Table 6. Correlations between independent variables

		Year	Offering Price	New issue [%]	IR [%]	WIG change [%]	WIG close	Offering value	Risk free rate	Number of IPOs	WIG log
Year	Pearson correlation	1.000	0.002	–0.059	–0.157**	0.072	0.179**	–0.001	–0.766**	–0.498**	0.243**
	significance		0.970	0.310	0.007	0.216	0.002	0.991	0.000	0.000	0.000
Offering Price	Pearson correlation	0.002	1.000	–0.128*	0.039	0.017	0.128*	0.260**	–0.065	0.053	0.120*
	significance	0.970		0.028	0.510	0.777	0.028	0.000	0.264	0.364	0.040
New issue [%]	Pearson correlation	–0.059	–0.128*	1.000	0.014	–0.076	0.016	–0.157**	0.045	0.055	0.017
	significance	0.310	0.028		0.813	0.193	0.788	0.007	0.439	0.346	0.772
IR [%]	Pearson correlation	–0.157**	0.039	0.014	1.000	0.048	0.073	–0.021	–0.014	0.101	0.073
	significance	0.007	0.510	0.813		0.413	0.210	0.723	0.809	0.085	0.213
WIG change [%]	Pearson correlation	0.072	0.017	–0.076	0.048	1.000	0.038	–0.009	–0.066	0.007	0.043
	significance	0.216	0.777	0.193	0.413		0.510	0.873	0.262	0.908	0.458
WIG close	Pearson correlation	0.179**	0.128*	0.016	0.073	0.038	1.000	0.008	–0.470**	0.624**	0.990**
	significance	0.002	0.028	0.788	0.210	0.510		0.897	0.000	0.000	0.000
Offering value	Pearson correlation	–0.001	0.260**	0.157**	–0.021	–0.009	0.008	1.000	0.022	–0.024	0.009
	significance	0.991	0.000	0.007	0.723	0.873	0.897		0.712	0.686	0.873
Risk free rate	Pearson correlation	–0.766**	–0.065	0.045	–0.014	–0.066	–0.470**	0.022	1.000	0.136*	–0.514**
	significance	0.000	0.264	0.439	0.809	0.262	0.000	0.712		0.020	0.000
Number of IPOs	Pearson correlation	–0.498**	0.053	0.055	0.101	0.007	0.624**	–0.024	0.136*	1.000	0.546**
	significance	0.000	0.364	0.346	0.085	0.908	0.000	0.686	0.020		0.000
WIG log	Pearson correlation	0.243**	0.120*	0.017	0.073	0.043	0.990**	0.009	–0.514**	0.546**	1.000
	significance	0.000	0.040	0.772	0.213	0.458	0.000	0.873	0.000	0.000	

** Correlation is significant at the 0.01 level. * Correlation is significant at the 0.05 level.

Source: own elaboration

There are statistically significant correlations between independent variables at a level of 0.01 and 0.05. In order to meet ordinary least squares requirements, we have performed Variance Inflation Factors statistics to examine a collinearity problem. Since all of the values are below 10, collinearity problem does not exist.

Table 7. Variance Inflation Statistics

Year	4.423
OfferingPrice	1.114
NewIssue	1.238
WIGchange	1.017
l_WIGclose	3.280
l_OfferingValue	1.338
RiskFreeRate	3.361
NumberofIPOs	3.634

Source: own elaboration

Two of the independent variables have much higher values than the rest: WIGclose and OfferingValue. In order to have a similar scale, natural logarithms of these variables were used in the model.

Four OLS models were built. The first model consists of all variables listed in Table 7. P-value was the highest for the variable NewIssue which was rejected from the second model. The second model R-squared was higher than the first model R-squared. In the second model, p-value was the highest for the variable OfferingPrice which was rejected from the third model. The third model R-squared was higher than the second model R-squared. Since rejections from the third model would lower the R-squared, the third model was the final one.

Two of the variables have a statistically significant impact on IR at a level of 0.01: the year and risk-free rate. One variable has a statistically significant impact on IR at a level of 0.05: the number of IPOs. Considering significance at the 0.1 level, one variable has an impact on IR: WIGclose.

Table 8. Ordinary least squares analysis

Dependent variable = IR				
	coefficient	**std. error**	**t-ratio**	**p-value**
Const	126.2360	25.9428	4.8660	1.8800e-06***
Year	−0.06375	0.0131	−4.8710	1.8400e-06***
WIGchange	2.0631	1.9385	1.0640	0.2881
l_WIGclose	0.2658	0.1504	1.7670	0.0783*
l_OfferingValue	−0.0202	0.0126	−1.6070	0.1092
RiskFreeRate	−9.4979	2.9669	−3.2010	0.0015***
NumberofIPOs	−0.0037	0.0017	−2.2300	0.0265**

<table>
<tr><th colspan="5">Dependent variable = IR</th></tr>
<tr><th></th><th>coefficient</th><th>std. error</th><th>t-ratio</th><th>p-value</th></tr>
<tr><td>Mean dependent var</td><td>0.1175</td><td>S. D. dependent var</td><td colspan="2">0.3581</td></tr>
<tr><td>Sum squared resid</td><td>34.1815</td><td>S.E. of regression</td><td colspan="2">0.3445</td></tr>
<tr><td>R-squared</td><td>0.0936</td><td>Adjusted R-squared</td><td colspan="2">0.0747</td></tr>
<tr><td>F(8, 286)</td><td>4.9570</td><td>P-value (F)</td><td colspan="2">0.0001</td></tr>
<tr><td>Log-likelihood</td><td>−100.6813</td><td>Akaike criterion</td><td colspan="2">215.3625</td></tr>
<tr><td>Schwarz criterion</td><td>241.1714</td><td>Hannan-Quinn</td><td colspan="2">225.6970</td></tr>
</table>

Source: own elaboration

The factor with the most significance is the year of the IPO (Figure 3). The coefficient of the year variable was –0.06, which means that the higher the year value, the lower the underpricing level. In developing markets, this can be explained by efficient market hypothesis (EMH). Fama (1965: 34–105) described an efficient market as “a market where prices at every point in time represent best estimates of intrinsic value. This implies in turn that, when an intrinsic value changes, the actual price will adjust “instantaneously”, where instantaneously means, among other things, that the actual price will initially overshoot the new intrinsic values as often as it will undershoot it”. Significance of *year* variable means that the WSE is more efficient now than it was in the past. This explanation is controversial. Adams, Thornton and Hall (2008: 67–74) in their study of IPO pricing argue against associating EMH with IPOs.

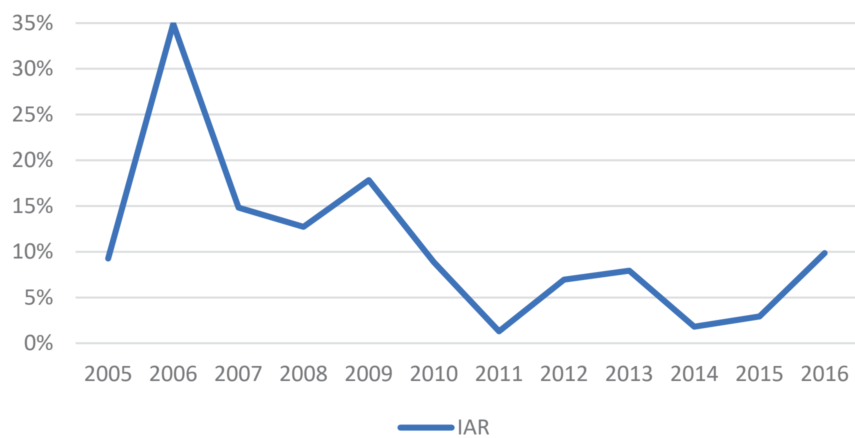


Figure 3. IAR
Source: own elaboration

Results of this and other studies conducted in Poland using data starting after the year 2000 (Czapiewski et al., 2014: 571–590; Pomykalski, Domagalski, 2015: 117–132) indicate that the level of underpricing was below 21.1% observed by Ritter in the US markets. Assuming that US markets are more developed, the conclusion that underpricing is higher in less developed countries does not hold.

The second most significant factor is the risk-free rate (Figure 4). The coefficient is –9.49. It means that the higher is the risk-free rate, the lower is the IPO initial rate of return. The sign of the coefficient is also expected. When interest rates are high, investors usually save their money because profits from bank deposits are satisfactory. Conversely, if interest rates are low, investors are looking for different investment opportunities because they cannot earn as much as they wish on a bank deposit. For that reason, more people are likely to buy shares when interest rates are low. High interest rates increase demand for shares and consequently first day closing prices.

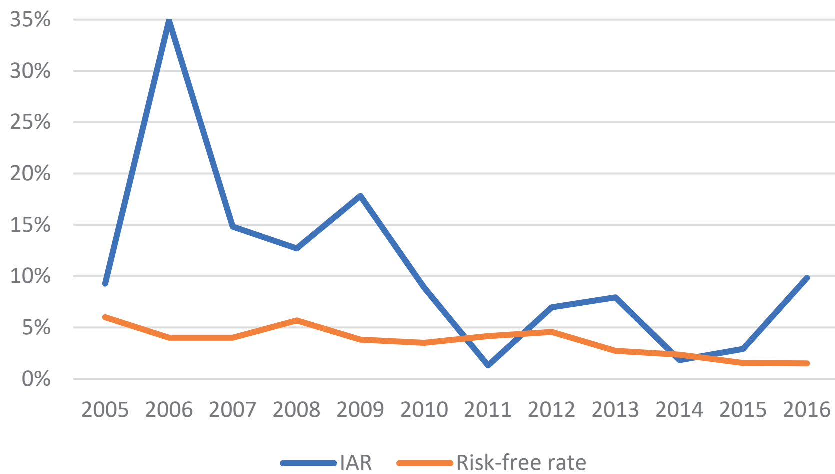


Figure 4. Risk-free rates and IAR

Source: own elaboration

The third most significant factor influencing IPO underpricing is the number of IPOs that took place in a given year (Figure 2). The coefficient is –0.0037 and it is negative, which means that when more IPOs took place in a given year, the underpricing level was lower. One possible explanation is the law of supply and demand. When there are more IPO offers on the market, people are not willing to pay as much as if there were fewer offers.

The least important but also significant factor is the WIG closing level. The WIG (Warszawski Indeks Giełdowy) is a cumulated value of all securities quoted on the Warsaw Stock Exchange. In this case, the WIG close is the level measured at the end of IPO day. The WIG coefficient is positive, which means that the higher the broad market index, the higher the IPO initial return. There is no clear explanation of this phenomenon in the literature. One possible reason of this dependency is stock exchange attractiveness. If the value of broad market index is at a high level, the stock exchange seems to be more attractive to investors who are more likely to invest in IPOs. That causes higher initial returns of IPOs. Adjusted R-squared of the final model is 7.47%. Since the goal of this paper was not to develop a model describing underpricing but to identify factors influencing initial return, the value is rewarding.

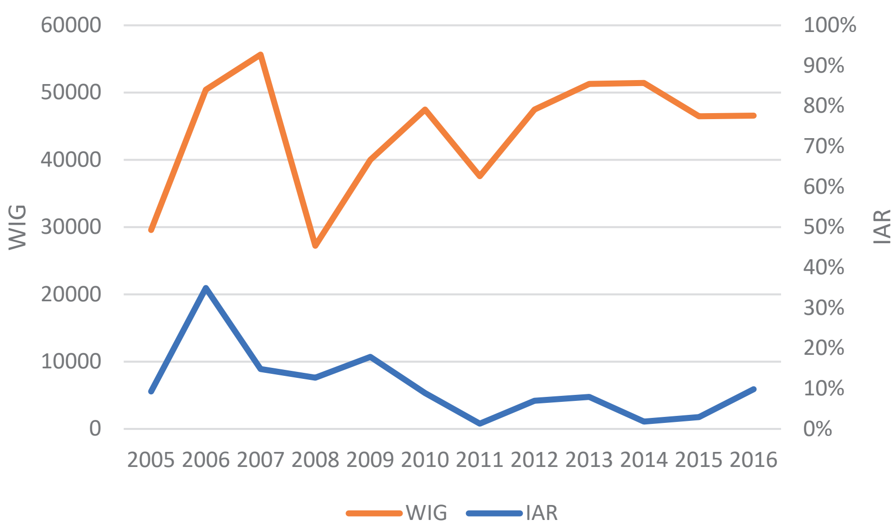


Figure 5. Comparison of WIG and IAR

Source: own elaboration

The impact of WIG on IAR was stronger in the period 2005–2008. This indicates that lower interest rates and the number of IPOs in the later years impacted underpricing to a larger extent than the broad stock market index. This conclusion is based on a short period of analysis and requires further research.

3.3. Limitations of this study

This study is limited to the WSE and one economic cycle. Results may be different in both more mature and less mature markets. We researched the impact of factors from a list of statistics used by the WSE. There may be other factors influencing underpricing as shown by Wołoszyn and Zarzecki (2013: 121–135).

Further research can concentrate on factors such as underwriters' reputation, free float of shares, market segment affiliation and oversubscription.

4. Conclusions

We have examined 349 companies which went public in 2005–2016. We have calculated the initial return and adjusted initial return and obtained results of 12.35% and 11.84% respectively. We have confirmed that underpricing existed during the researched period.

We have also examined the influence of selected factors on underpricing and can conclude that during the researched period three of the examined factors had a statistically significant influence on initial public offering underpricing. The year of IPO (negatively), risk-free rate (negatively) and WIG close value (positively) influenced underpricing during the researched period.

Due to a limited scope of our research (a short period and one market), our results should be treated with caution and used in further inquiries into a possible impact of interest rates and stock market indexes on underpricing.

References

Adams M., Thornton B., Hall G. (2008), *IPO pricing phenomena: empirical evidence of behavioral biases*, "Journal of Business & Economics Research", vol. 6, no. 4, pp. 67–74, https://doi.org/10.19030/jber.v6i4.2412

Aussenegg W. (2000), *Privatization versus Private Sector Initial Public Offerings in Poland*, "Multinational Finance Journal", vol. 4, no. 1/2, pp. 69–99

Banerjee S., Dai L., Shrestha K. (2011), *Cross-country IPOs: What explains differences in underpricing?*, "Journal of Corporate Finance", vol. 17, no. 5, pp. 1289–1305, https://doi.org/10.1016/j.jcorpfin.2011.06.004

Cassia L., Giudici G., Paleari S., Redondi R. (2004), *IPO underpricing in Italy*, "Applied Financial Economics", vol. 14, no. 3, pp. 179–194, https://doi.org/10.1080/0960310042000187333

Chambers D., Dimson E. (2009), *IPO underpricing over the very long run*, "The Journal of Finance", vol. 64, no. 3, pp. 1407–1443, https://doi.org/10.1111/j.1540-6261.2009.01468.x

Chan K., Wang J., Wei K.J. (2004), *Underpricing and long-term performance of IPOs in China*, "Journal of Corporate Finance", vol. 10, no. 3, pp. 409–430, https://doi.org/10.1016/S0929-1199(03)00023-3

Chi J., Padgett C. (2005), *Short-run underpricing and its characteristics in Chinese initial public offering (IPO) markets*, "Research in International Business and Finance", vol. 19, no. 1, pp. 71–93, https://doi.org/10.1016/j.ribaf.2004.10.004

Chowdhry B., Sherman A. (1996), *International differences in oversubscription and underpricing of IPOs*, "Journal of Corporate Finance", vol. 2, no. 4, pp. 359–381, https://doi.org/10.1016/0929-1199(96)00002-8

Czapiewski L., Jewartowski T., Kałdoński M., Mizerka J. (2014), *Determinanty natychmiastowych stóp zwrotu z akcji spółek skarbu państwa dokonujących pierwotnych ofert publicznych*, "Ekonomista" vol. 4, pp. 571–590.

Fama E.F. (1965), *The behavior of stock-market prices*, "The Journal of Business", vol. 38, no. 1, pp. 34–105.

Fijałkowska D., Muszyński M., Pauka M. (2013), *Zjawisko underpricingu a ład korporacyjny na rynku NewConnect*, "Zarządzanie i Finanse", vol. 11, no. 2/6, pp. 415–426.

Ganesamoorthy L., Shankar H. (2013), *The Performance of Initial Public Offerings Based on Their Size: An Empirical Analysis of the Indian Scenario*, "IUP Journal of Applied Finance", vol. 19, no. 4, pp. 84–99.

Habib M.A., Ljungqvist A.P. (2001), *Underpricing and entrepreneurial wealth losses in IPOs: Theory and evidence*, "The Review of Financial Studies", vol. 14, no. 2, pp. 433–458, https://doi.org/10.1093/rfs/14.2.433

Henricson T. (2012), *Underpricing in the Swedish IPO market: Can investors earn abnormal returns by investing in IPOs?*, Unpublished thesis, Jönköping University, Jönköping.

Hunger A. (2003), *Market Segmentation and Ipo-Underpricing: The German Experience*, "Munich Business Working Paper", no. 2003–14, http://dx.doi.org/10.2139/ssrn.496584

Jewartowski T., Lizińska J. (2012), *Short-and Long-Term Performance of Polish IPOs*, "Emerging Markets Finance and Trade", vol. 48, no. 2, pp. 59–75, https://doi.org/10.2753/REE1540-496X480204

Lizińska J., Czapiewski L. (2015), *Determinanty underpricingu w Polsce i na innych rynkach wschodzących*, "Research Papers of The Wroclaw University of Economics/Prace Naukowe Uniwersytetu Ekonomicznego We Wroclawiu", no. 386, pp. 112–125.

Lyn E.O., Zychowicz E.J. (2003), *The Performance of New Equity Offerings in Hungary and Poland*, "Global Finance Journal", vol. 14, no. 2, pp. 181–195, https://doi.org/10.1016/S1044-0283(03)00011-5

Mezhoud M., Boubaker A. (2011), *Determinants of the Components of IPO Initial Returns: Paris Stock Exchange*, "International Journal of Accounting and Financial Reporting", vol. 1, no. 1, pp. 166–180, https://doi.org/10.5296/ijafr.v1i1.986

Pomykalski P., Domagalski M. (2015), *IPO Underpricing and Financial Analysts' Forecast Accuracy in Poland*, "Acta Universitatis Lodziensis. Folia Oeconomica", vol. 1, no. 310, pp. 117–132, http://dx.doi.org/10.18778/0208-6021.310.09

Ritter J. (2018), *IPO data website*, https://site.warrington.ufl.edu/ritter/ipo-data/ [accessed: 1.04.2018].

Saunders A. (1990), *Why are so many new stock issues underpriced*, "Business Review", March, pp. 3–12

Sieradzki R. (2013), *Does it pay to invest in IPOs? Evidence from the Warsaw Stock Exchange*, "National Bank of Poland Working Paper", no. 139, pp. 1–37, http://dx.doi.org/10.2139/ssrn.2213770

Wołoszyn A., Zarzecki D. (2013), *The impact of the January effect on the IPO underpricing in Poland*, "Folia Oeconomica Stetinensia", vol. 13, no. 1, pp. 121–135.

Czynniki wpływające na niedoszacowanie cen emisyjnych pierwszych ofert publicznych akcji w Polsce

Streszczenie: W artykule analizujemy teorię oraz dowody występowania zjawiska niedoszacowania cen emisyjnych pierwszych ofert publicznych akcji na Giełdzie Papierów Wartościowych w Warszawie. Mimo że temat był już w przeszłości badany, uważamy, że ostatnia dekada niskich stóp procentowych zasługuje na szczególną uwagę. Zbadaliśmy zjawisko niedoszacowania cen emisyjnych w tym okresie, a także wykazaliśmy, że trzy czynniki miały istotnie statystyczny wpływ na wielkość zjawiska: rok emisji, stopa procentowa wolna od ryzyka oraz poziom zamknięcia indeksu szerokiego rynku w dniu emisji.

Słowa kluczowe: IPO, niedoszacowanie cen emisyjnych, Polska, finanse behawioralne

JEL: G11, G15, G24

© by the author, licensee Łódź University – Łódź University Press, Łódź, Poland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license CC-BY (https://creativecommons.org/licenses/by/4.0/)

Received: 2019-10-05; verified: 2020-03-12. Accepted: 2020-08-19

C O P E
Member since 2018
JM13703

This journal adheres to the COPE's Core Practices https://publicationethics.org/core-practices

Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663
www.czasopisma.uni.lodz.pl/foe/
4(349) 2020
DOI: http://dx.doi.org/10.18778/0208-6018.349.02

Artur Włodzimierz Parys
Uniwersytet Łódzki, Wydział Ekonomiczno-Socjologiczny, parysartur@wp.pl

Ocena lokalnej polityki mieszkaniowej Łodzi i Krakowa w świetle wybranych aspektów

Streszczenie: W artykule przedstawiono badanie dotyczące lokalnej polityki mieszkaniowej Łodzi i Krakowa w latach 2006 i 2017. Dokonano analizy porównawczej w zakresie ludności, zasobów mieszkaniowych (w tym lokali socjalnych i komunalnych), stopy bezrobocia, dochodów i wydatków budżetów badanych miast, sytuacji ekonomicznej wyrażonej dochodami budżetowymi na jednego mieszkańca. Wyniki przeprowadzonych analiz ilościowych pokazują trend zmniejszania się liczby zasobów mieszkaniowych w badanych miastach. W ostatnich latach intensywniej zmniejszają się zasoby mieszkaniowe w Łodzi niż w Krakowie, a struktura własności budynków mieszkalnych Łodzi różni się od struktury budynków mieszkalnych w Krakowie.

Słowa kluczowe: mieszkaniowy zasób gminy, zapotrzebowanie na mieszkania, polityka mieszkaniowa

JEL: R21

1. Wprowadzenie

Mieszkalnictwo rozumiane jest jako zespół decyzji i poczynań prowadzących do osiągnięcia założonego modelu warunków i stosunków mieszkaniowych oraz zasad funkcjonowania gospodarki mieszkaniowej (Kulesza, Nieciuński, 1989: 46). Jest jednym z głównych obszarów gospodarki i w znaczący sposób wpływa na procesy społeczno-gospodarcze oraz skuteczność działań rozwojowych państwa. Oddziałuje również istotnie na poziom zaspokojenia potrzeb społecznych. Widoczne powiązanie systemu mieszkaniowego z gospodarką wskazuje, że powinien on odgrywać ważną rolę w podnoszeniu jakości społecznej, gospodarczej i przestrzennej kraju. Mieszkalnictwo zawiera zarówno elementy z obszaru polityki gospodarczej, jak i polityki społecznej. Wymiar ingerencji państwa jest wynikiem przyjętej ideologii i wyraża się w deklarowanej i realizowanej polityce mieszkaniowej.

Celem artykułu jest ocena lokalnej polityki mieszkaniowej na przykładzie Łodzi i Krakowa oraz odpowiedź na następujące pytanie badawcze: „Czy poprawa warunków mieszkaniowych Łodzi i Krakowa będzie miała odzwierciedlenie w zwiększeniu atrakcyjności oraz dochodów badanych miast?".

Analizie retrospektywnej poddano dane mieszkaniowe, demograficzne oraz gospodarcze Łodzi i Krakowa w latach 2006 oraz 2017. Zastosowano badania i analizę dokumentów oraz metody statystyczne.

2. Definicje polityki mieszkaniowej

W literaturze nie istnieje jednolita definicja polityki mieszkaniowej. Adam Andrzejewski definiuję politykę mieszkaniową jako ogólny kierunek i metody działania stosowane przez państwo i gminę do osiągnięcia określonych celów w dziedzinie mieszkalnictwa i zaspokojenia potrzeb mieszkaniowych (Andrzejewski, 1979: 25). Problemem pozostaje zakres ingerencji państwa w mechanizmy rynku nieruchomości. Wskazano, że interwencja państwa w obszarze mieszkalnictwa jest uzasadniona w przypadkach, gdy ma ona skorygować pojawiające się nieefektywności i niesprawności rynków mieszkaniowych, zapewnić sprawiedliwość społeczną oraz stabilność społeczną i finansową (Lis, 2017: 54).

Szersze ujęcie definiujące politykę mieszkaniową przedstawia ją jako działania podejmowane przez państwo czy też poszczególne władze lokalne, dlatego według Andrzejewskiego stosowane instrumenty polityki mieszkaniowej muszą być nakierowane na likwidację przyczyn niekorzystnych zjawisk, a nie jedynie na redukowanie niekorzystnych skutków (Andrzejewski, 1979: 25). Koordynacja poszczególnych narzędzi polityki gospodarczej, pozwalających na ochronę gospodarki przed ewentualnym kryzysem gospodarczym, wymaga przyjęcia przez państwo odpowiednich kierunków i metod działania. Metody te przyjmowane są również na polu polityki mieszkaniowej.

Analizując różne definicje i ujęcia polityki mieszkaniowej, trzeba zaznaczyć, że odnosi się ona zarówno do kwestii ilościowych, jak i jakościowych. Polityka mieszkaniowa w ujęciu ilościowym polega na wprowadzeniu innych instrumentów w zakresie rozwiązań instytucjonalno-prawnych. Natomiast w ujęciu jakościowym zawiera projekty dążące do zmian w warstwie instytucjonalno-prawnej. Według Radosława Cyrana efektem tych działań powinno być zmniejszenie kosztów utrzymania zasobu mieszkaniowego, usprawnienie procesu zarządzania zasobem oraz przewidywalność decyzji w odniesieniu do kosztów utrzymania zasobu (Cyran, 2013: 218). W innych interpretacjach polityka mieszkaniowa ma na celu również wprowadzenie nowych instrumentów lub usunięcie złych, dotychczas wykorzystywanych, a także wprowadzenie znaczących zmian systemu mieszkaniowego oraz norm i reguł kierujących jego działaniem. Aby osiągnąć założony cel – polepszenie warunków mieszkaniowych w gminie − należy łącznie stosować wszelkie podejmowane przez gminy działania pomocowe z zakresu polityki mieszkaniowej (Cyran, 2013).

3. Prawo do mieszkania jako fundament polityki mieszkaniowej

Mieszkanie traktuje się jako dobro niezbędne każdej rodzinie, które zaspokaja zarówno potrzeby podstawowe, jak i potrzeby wyższego rzędu (Cyran, 2013). Prawo popytu dotyczy tej części potrzeb, która ma swoje odzwierciedlenie w środkach finansowych pozwalających na zaspokojenie tej potrzeby. Prawo podaży wskazuje zależności między ilością dóbr i usług dostarczanych na rynek a ich cenami. Część osób ma możliwość zaspokojenia swoich potrzeb mieszkaniowych samodzielnie, przekształcając je w popyt na mieszkania. W tej sytuacji najlepszym rozwiązaniem będzie wynajem mieszkania o odpowiednim standardzie oraz odpowiedniej do niego wysokości czynszu. Można wyróżnić jeszcze grupy gospodarstw domowych nieposiadających żadnych możliwości finansowych, które nie mogą pozwolić sobie na przekształcenie potrzeb mieszkaniowych w realny popyt. Do tych grup gospodarstw domowych zaliczają się najczęściej osoby bezrobotne lub niesamodzielne ekonomicznie z różnych powodów. Wymagają one pomocy ze strony państwa w postaci interwencji socjalnej, mającej na celu zaspokojenie ich minimalnego poziomu potrzeb mieszkaniowych.

Dobro mieszkaniowe to jeden z najistotniejszych czynników w życiu człowieka. Prawo do mieszkania zostało unormowane w regulacjach światowych, europejskich oraz krajowych. Po raz pierwszy prawo do zamieszkania na poziomie międzynarodowym zostało wymienione pośród praw podstawowych w Powszechnej Deklaracji Praw Człowieka ONZ w 1948 roku. Zgodnie z artykułem 25.1 tego dokumentu każdy człowiek ma prawo do życia na poziomie zapewniającym zdrowie i pomyślność jemu i jego rodziny, włączając w to wyżywienie, odzież, mieszkanie, opiekę lekarską oraz niezbędne świadczenia socjalne. Kolejnym międzynarodowym aktem prawnym mówiącym o prawie do zamieszkania jest Międzynarodowy Pakt Praw Gospodarczych, Społecznych i Kulturalnych z 1966 roku. W artykule 11 punkt 1 potwierdzono prawo każdego człowieka i jego rodziny do odpowiedniego poziomu życia, włączając w to wyżywienie, odzież i mieszkanie, jak również prawo do stałej poprawy warunków egzystencji. W 1969 roku w Deklaracji Postępu Społecznego i Rozwoju w artykule 10 zapisano, że zapewnienie odpowiedniego mieszkania i usług społecznych wszystkim obywatelom, a w szczególności najbiedniejszym i wielodzietnym rodzinom, warunkuje podstawowe wolności człowieka. Natomiast w Konwencji o Prawach Dziecka z 1989 roku w artykule 27 punkt 3 zawarte zostały wspólne zobowiązania państw ratyfikujących do podejmowania odpowiednich działań, uzależnionych od warunków i środków, jakimi dysponuje dane państwo, aby pomóc rodzinom i osobom mającym na utrzymaniu dziecko, poprzez wsparcie materialne i programy pomocowe, a w szczególności w zakresie wyżywienia, odzieży i mieszkania. W Deklaracji Światowej Konferencji ONZ Habitat II w Stambule z 1996 roku zapisano, że poszczególne kraje zobowiązują się do pełnej i postępującej realizacji prawa do odpowiedniego mieszkania, zgodnie z postanowieniami dokumentów międzynarodowych oraz do zapewnienia równego dostępu do tanich i odpowiednich mieszkań wszystkim osobom i ich rodzinom (Deklaracja Światowej Konferencji ONZ Habitat II).

Obszar mieszkalnictwa realizowany jest również przez agendę lizbońską, a mieszkanie określane jest w Europejskiej Karcie Mieszkaniowej jako dobro pierwszej potrzeby, podstawowe prawo socjalne leżące u podstaw europejskiego modelu społecznego oraz element godności ludzkiej (Parlament Europejski, Komisja Rozwoju Regionalnego, 2006). W Europejskiej Karcie Mieszkaniowej znajduje się również zapis dotyczący budownictwa społecznego, finansowanego z funduszy strukturalnych, oraz potwierdzający prawo do mieszkań o wysokiej jakości, w tym mieszkań socjalnych. Artykuł 31 Europejskiej Karty Społecznej, podpisanej w Strasburgu 3 maja 1995 roku, mówi o popieraniu dostępu do mieszkań o odpowiednim standardzie; zapobieganiu i ograniczaniu bezdomności, w celu jej stopniowego likwidowania; uczynieniu kosztów mieszkań dostępnymi dla osób, które nie mają wystarczających zasobów.

W Opinii Komitetu Regionów Polityka Mieszkaniowa i Regionalna możemy przeczytać, że głównym warunkiem spójności społecznej i terytorialnej Unii Europejskiej jest dostęp do budownictwa mieszkaniowego na przyzwoitym poziomie, gdyż jest ono jednym z głównych czynników walki z wykluczeniem społecznym i bezrobociem.

W artykule 34 Karty Praw Podstawowych Unii Europejskiej znajdujemy zapis mówiący, że w celu zwalczania wykluczenia społecznego oraz ubóstwa Unia uznaje i szanuje prawo do pomocy społecznej i mieszkaniowej dla zapewnienia, zgodnie z zasadami ustanowionymi w prawie wspólnotowym oraz ustawodawstwach i praktykach krajowych, godnej egzystencji wszystkim osobom pozbawionym wystarczających środków. W Opinii Europejskiego Komitetu Ekonomiczno-Społecznego w sprawie: „Mieszkalnictwo, a polityka regionalna" zwrócono uwagę na potrzebę utworzenia różnorodnej oferty mieszkań w celu umożliwienia konkurencyjności wszystkich obszarów i ułatwienia mobilności pracowników. Ponadto wskazano, że pomoc mieszkaniowa powinna uwzględniać specyficzne cechy mieszkalnictwa socjalnego, w tym słabą wypłacalność mieszkańców. Regulacje krajowe dotyczące prawa do mieszkania wynikają bezpośrednio z konstytucji poszczególnych państw europejskich. W ustawach zasadniczych państw takich jak Belgia, Austria, Holandia, Hiszpania, Portugalia, Polska, Ukraina, Słowacja i Rosja zapisane są ogólne kryteria prawa do zamieszkania. W Polsce prawo do zamieszkania reguluje artykuł 75 Konstytucji RP oraz Ustawa z dnia 21 czerwca 2001 roku o ochronie praw lokatorów, mieszkaniowym zasobie gminy i o zmianie Kodeksu Cywilnego, gdzie wskazane jest, że tworzenie warunków do zaspokajania potrzeb mieszkaniowych wspólnoty samorządowej należy do zadań własnych gminy oraz że gmina, na zasadach i w przypadkach określonych w ustawie, zapewnia lokale w ramach najmu socjalnego i lokale zamienne, a także zaspokaja potrzeby mieszkaniowe gospodarstw domowych o niskich dochodach. Tak więc prawo do zamieszkania jest fundamentem polityki mieszkaniowej i jednym z głównych czynników, który określa jej cele i instrumenty.

4. Cele polityki mieszkaniowej

Formułując główne cele polityki mieszkaniowej, można podzielić je według założeń ich oddziaływania na rynek mieszkaniowy. Wyróżnia się dwie grupy podstawowych celów. Są to:

1) cele odnoszące się do strony popytowej, rozumianej jako popyt efektywny oraz potrzeby mieszkaniowe, czyli formułowanie warunków do nabywania lub najmu mieszkań przez wszystkich obywateli;
2) cele skupiające się na stronie podażowej, związane z zagwarantowaniem odpowiedniej dostępności i jakości zasobów mieszkaniowych.

Określając główny cel polityki mieszkaniowej, związany z zaspokojeniem potrzeb mieszkaniowych, uwidacznia się problem tych gospodarstw domowych, które nie są w stanie zrealizować go samodzielnie. Stąd też w ramach polityki mieszkaniowej została wyodrębniona społeczna polityka mieszkaniowa, która ma na celu zapewnienie gospodarstwom mieszkaniowym o najniższych dochodach dostępu do zasobów mieszkaniowych w sposób najbardziej efektywny ekonomicznie. W ramach społecznej polityki mieszkaniowej pomoc obejmuje również takie grupy jak osoby bezdomne i osoby starsze. Z punktu widzenia społecznej polityki mieszkaniowej można wyróżnić mechanizmy bezpośredniego wpływu państwa. W tej grupie instrumentów znajduje się społeczne budownictwo mieszkaniowe. Do powszechnych form społecznego budownictwa zalicza się przede wszystkim budownictwo przeznaczone na wynajem, w szczególności budownictwo komunalne, realizowane przez władze lokalne, społeczne budownictwo czynszowe, zależne lub niezależne od władz lokalnych oraz społeczne budownictwo własnościowe, realizowane przez inwestorów prywatnych (Lis, 2018: 73).

Poszczególne kraje Unii Europejskiej formułują swoje cele polityki mieszkaniowej na podstawie różnych aspektów ekonomiczno-gospodarczych. Większość państw europejskich formułuje cele polityki państwa poprzez stronę popytową rynku mieszkaniowego. Podażowe podejście w kształtowaniu polityki państwa reprezentują kraje takie jak Hiszpania, Dania, Litwa oraz Słowenia. Promocję mieszkań własnościowych, jako główny cel polityki wobec budownictwa mieszkaniowego, podkreślają takie kraje jak Niemcy, Francja, Belgia oraz Malta (Lis, 2018). Politycy formułujący cele priorytetowe polityki mieszkaniowej często nie mają wystarczającej wiedzy z zakresu ekonomii mieszkalnictwa oraz modeli opisujących funkcjonowanie gospodarki, w szczególności sektora mieszkaniowego, stąd niemożliwe jest utrzymanie i nadanie stałych celów polityki mieszkaniowej oraz ich realizacji w tym samym czasie (Acocella, 2002: 214).

5. Instrumenty polityki mieszkaniowej

Instrumentami polityki gospodarczej, w tym także polityki mieszkaniowej, powinny być tylko te zmienne, które spełniają trzy podstawowe warunki. Po pierwsze, politycy mogą kontrolować daną zmienną. Po drugie, zmienna, której wartość została ustalona przez polityków, powinna mieć wpływ na zmienne uznane za cele polityki (Lis, 2005: 12). Po trzecie, konieczne jest odróżnienie danej zmiennej od innych instrumentów pod względem możliwości kontrolowania i skuteczności (Acocella, 2002: 214). Mnogość środowisk decyzyjnych w zakresie instrumentów polityki mieszkaniowej stanowi podstawowy problem. Powoduje to nieprawidłowości w weryfikowaniu instrumentów oraz komplikacje w zakresie ich skuteczności.

Prezentując instrumenty polityki mieszkaniowej, można przyjąć dwa kryteria podziału:

1) według rodzaju oddziaływania,
2) według kierunku oddziaływania.

Zgodnie z pierwszym założeniem zespół instrumentów polityki mieszkaniowej tworzą narzędzia polityki fiskalnej, strony przychodowej, strony rozchodów oraz instrumenty bezpośredniego oddziaływania. Do narzędzi polityki fiskalnej można zaliczyć przychody oraz rozchody instytucji rządowych i władz lokalnych, a do narzędzi strony przychodowej: podatki, uwzględniając podatek od nieruchomości, podatek od czynności cywilnoprawnych, podatek od przychodu ze sprzedaży nieruchomości, podatek od dochodu z najmu. Do narzędzi strony rozchodów można zaliczyć: transfery obejmujące gospodarstwa domowe, inwestycje rządowe oraz inwestycje władz lokalnych, dofinansowania przeznaczone dla przedsiębiorstw sektora budowlanego. Do instrumentów bezpośredniego oddziaływania można zaliczyć wszelkie nakazy i zakazy administracyjne, do których zaliczają się: regulacje instytucji, które finansują sektor mieszkaniowy, regulacje prawa do własności nieruchomości, regulacje dotyczące zagospodarowania przestrzennego, regulacje dotyczące finansowania firm z obszaru budownictwa mieszkaniowego. Do instrumentów zalicza się również tworzenie i regulowanie instytucji ułatwiających lub wspomagających zarządzanie obszarem z zakresu budownictwa mieszkaniowego.

6. Analiza porównawcza sytuacji gospodarczo-społecznej w Łodzi i Krakowie w 2006 i 2017 roku

Kształt sytuacji mieszkaniowej jest wynikiem uwarunkowań gospodarczych i społecznych, stanu zasobu mieszkaniowego oraz zmian kulturowych. Niejednokrotnie obserwuje się wyraźne różnice wynikające z zaszłości historycznych i sposobu percepcji potrzeb mieszkaniowych. Kraków i Łódź to miasta zbliżone pod względem liczby ludności. W Łodzi zauważalna jest tendencja do wyludniania się miasta, Kraków zyskuje mieszkańców, co jest wyjątkowe wśród największych polskich miast (poza Warszawą). Sytuację gospodarczo-społeczną w omawianych miastach ilustrują: stopa bezrobocia, przeciętne wynagrodzenie brutto ogółem, dochody budżetu gmin i miast na prawach powiatu, wydatki budżetu gmin i miast na prawach powiatu oraz spółki handlowe z udziałem kapitału zagranicznego.

W badaniu uwzględniono również zmiany liczby ludności w Łodzi i Krakowie. Według stanu na 31 marca 2006 roku ludność Łodzi liczyła 765,8 tys. osób, tj. 29,7% ogółu ludności województwa łódzkiego. W 2006 roku ludność Krakowa liczyła 756,3 tys. mieszkańców, tj. 23,1% ogółu ludności województwa małopolskiego. Na 30 grudnia 2017 roku Łódź liczyła jedynie 693,79 tys. mieszkańców, co stanowiło około 28% populacji województwa łódzkiego. W końcu grudnia 2017 roku populacja Krakowa wynosiła 767,35 tys. mieszkańców, co stanowiło 22,6% ogólnej liczby ludności województwa małopolskiego. Obserwowany od wielu lat ubytek liczby ludności w Łodzi to rezultat ujemnego przyrostu naturalnego oraz stopnia migracji mieszkańców na pobyt stały, głównie w Łódzkim Obszarze Metropolitalnym.

Analizując stan bezrobocia, należy przypomnieć, że czynnikiem kształtującym sytuację gospodarczo-społeczną jest poziom bezrobocia. W maju 2004 roku, po wstąpieniu Polski do Unii Europejskiej, poziom bezrobocia w naszym kraju należał do najwyższych wśród państw Wspólnoty. W kolejnych latach nastąpiło ożywienie na rynkach pracy oraz rozpoczęła się emigracja zarobkowa Polaków do krajów Unii Europejskiej, którą obserwujemy do dziś.

Tabela 1. Stopa bezrobocia w Łodzi i Krakowie w 2006 i 2017 roku ogółem (w proc.)

Lata	Łódź	Kraków
2006	11,70	5,50
2017	6,30	2,70

Źródło: opracowanie własne na podstawie BDL GUS

Tabela 2. Bezrobotni pozostający bez pracy dłużej niż rok w Łodzi i Krakowie

Lata	Łódź	Kraków
2006	19 087	9 739
2017	10 140	5 301

Źródło: opracowanie własne na podstawie BDL GUS

Zarówno w 2006 roku, jak i w 2017 roku stopa bezrobocia w Krakowie była niższa niż w Łodzi. Było to spowodowane ożywieniem na rynku pracy obu miast (patrz Tabele 1 i 2).

Kolejnym czynnikiem wpływającym na sytuację społeczno-gospodarczą danego miasta są zwiększające się zarobki jego mieszkańców. Uzyskiwane wynagrodzenie ma bezpośredni wpływ na możliwość zaspokajania potrzeb mieszkaniowych i wyraźniej uwidoczniło się to w Krakowie niż w Łodzi (patrz Tabele 3 i 4).

Tabela 3. Przeciętne wynagrodzenie brutto ogółem (zł) w Łodzi i Krakowie

Lata	Łódź	Kraków
2006	2500,84	2753,21
2017	4462,50	4966,20

Źródło: opracowanie własne na podstawie BDL GUS

Tabela 4. Przeciętne miesięczne wynagrodzenie brutto (zł) w relacji do średniej krajowej (Polska = 100%) (w proc.)

Lata	Łódź	Kraków
2006	94,80	104,40
2017	98,60	109,70

Źródło: opracowanie własne na podstawie BDL GUS

Sytuacja ekonomiczna badanych miast może być wyrażona dochodami budżetowymi na jednego mieszkańca. Różnica pomiędzy badanymi miastami w 2006 i 2017 roku jest widoczna z analizy (patrz Tabele 3 i 4). W obu miastach znacząco wzrosły dochody na jednego mieszkańca, szczególnie w Krakowie.

Tabela 5. Dochody budżetu gmin i miast na prawach powiatu – dochody na jednego mieszkańca

Lata	Polska	Łódź	Kraków
2006	2431,28	2761,06	3075,30
2017	4937,72	5616,32	6469,60

Źródło: opracowanie własne na podstawie BDL GUS

Tabela 6. Dochody budżetu gmin i miast na prawach powiatu – dochody na jednego mieszkańca w relacji do średniej krajowej (Polska = 100%) (w proc.)

Lata	Łódź	Kraków
2006	114	126
2017	114	131

Źródło: opracowanie własne na podstawie BDL GUS

Z analizy powyższych dochodów budżetu gmin i miast na prawach powiatu na jednego mieszkańca wynika, że w Łodzi zarówno w 2006, jak i w 2017 roku powyższe dochody przewyższały średnią krajową o 14%. Natomiast w Krakowie dochody budżetu gmin i miast na prawach powiatu na jednego mieszkańca przewyższały średnią krajową o 26% w 2006 roku, a w 2017 roku o 31%. Oznacza to wzrost o 5% w stosunku do 2006 roku.

Tabela 7. Wydatki budżetu gmin i miast na prawach powiatu – wydatki na jednego mieszkańca

Lata	Polska	Łódź	Kraków
2006	24 760,00	2 810,02	3 249,06
2017	4 957,80	5 681,41	6 540,49

Źródło: opracowanie własne na podstawie BDL GUS

Tabela 8. Wydatki budżetu gmin i miast na prawach powiatu – wydatki na jednego mieszkańca w relacji do średniej krajowej (Polska = 100%) (w proc.)

Lata	Łódź	Kraków
2006	113	131
2017	114	132

Źródło: opracowanie własne na podstawie BDL GUS

Analiza danych GUS zebranych w Tabelach 7 i 8 wskazuje, że w Łodzi wydatki budżetu gmin i miast na prawach powiatu na jednego mieszkańca przewyższały średnią krajową o 13% w 2006 roku, a w 2017 roku o 14%, wzrosły więc o 1% w stosunku do 2006 roku. W Krakowie wydatki budżetu gmin i miast na prawach powiatu na jednego mieszkańca przewyższały średnią krajową o 31% w 2006 roku, a w 2017 roku o 32%, wzrosły więc o 1% w stosunku do 2006 roku. Należy podkreślić, że dochody i wydatki budżetów badanych miast na jednego mieszkańca w analizowanych latach były wyższe w Krakowie niż w Łodzi.

Innym czynnikiem wpływającym na aktywność gospodarczą są spółki z kapitałem zagranicznym.

Tabela 9. Spółki handlowe z udziałem kapitału zagranicznego według rejestru REGON

Lata	Łódź	Kraków
2006	1475	2086
2017	2172	4965

Źródło: opracowanie własne na podstawie BDL GUS

Tabela 10. Spółki handlowe według rejestru REGON

Lata	Łódź	Kraków
2006	6 702	10 742
2017	13 272	26 160

Źródło: opracowanie własne na podstawie BDL GUS

Podobne zjawisko obserwujemy, analizując liczbę wpisanych spółek handlowych według rejestru REGON w Łodzi i Krakowie. W Krakowie zauważalna jest zdecydowanie większa liczba spółek handlowych. Widoczny jest również większy przyrost liczby spółek handlowych w roku 2017 w porównaniu do 2006 w Krakowie niż w Łodzi (patrz Tabela 9).

7. Ocena zasobu mieszkaniowego w Łodzi i Krakowie w latach 2006 i 2017

Gmina powinna podejmować wszelkie działania sprzyjające efektywnemu wykorzystaniu zasobu mieszkaniowego oraz tworzyć i wcielać w życie konkretne strategie zasiedlania mieszkań i zrozumiałe procedury ich przydziałów oraz monitorować ich realizację. Podstawowe przepisy prawne dotyczące obowiązków i zadań gmin z zakresu mieszkalnictwa znajdują się w różnych aktach prawnych. Artykuł 75 Konstytucji Rzeczypospolitej Polskiej stanowi, iż władze publiczne prowadzą politykę sprzyjającą zaspokojeniu potrzeb mieszkaniowych obywateli, a w szczególności przeciwdziałają bezdomności, wspierają rozwój budownictwa socjalnego oraz popierają działania obywateli zmierzające do uzyskania własnego mieszkania. W Ustawie z dnia 8 marca 1990 r. o samorządzie gminnym w wykazie zadań powierzonych gminie wymienione zostało między innymi komunalne budownictwo mieszkaniowe. Zasadniczym źródłem prawa w zakresie ochrony praw lokatorów oraz zadań gminy w zakresie mieszkalnictwa jest ustawa o ochronie praw lokatorów, mieszkaniowym zasobie gminy i o zmianie Kodeksu cywilnego. Należy dodać, że dodatkowym dokumentem prawnym odnoszącym się do zadań gminy z pogranicza pomocy społecznej i mieszkalnictwa jest Ustawa z dnia 21 czerwca 2001 r. o dodatkach mieszkaniowych, regulująca kwestię wspomagania finansowego utrzymania lokali mieszkalnych przez gminy.

Kolejną ustawą jest Ustawa z dnia 12 marca 2004 r. o pomocy społecznej, odnosząca się do kwestii udzielania czasowego schronienia dla bezdomnych, a także prowadzenia i zapewnienia miejsc w mieszkaniach chronionych. Tworzenie warunków do zaspokojenia potrzeb mieszkaniowych powinno być jednym z priorytetowych celów gminy. Nakładając na gminy obowiązek zaspokajania potrzeb mieszkaniowych, w szczególności zapewnienia najmu socjalnego, ustawodawca umożliwił obywatelom nabywanie lokali z mieszkaniowego zasobu gminy (Ustawa z dnia 21 czerwca 2001 r. o ochronie praw lokatorów…). Tworzenie takiego zasobu nie jest jednak obowiązkiem, a jedynie prawem gminy. Gmina może wynajmować lokale wchodzące w skład zasobu mieszkaniowego, jednakże umowy dotyczące tych lokali muszą być zawierane na czas nieoznaczony. W celu zaspokajania potrzeb mieszkaniowych gmina może także wynajmować lokale od innych właścicieli i podnajmować je osobom, których gospodarstwa domowe osiągają niski dochód. Takie lokale nie wchodzą w skład zasobu mieszkaniowego gminy.

Rozwój zasobów mieszkaniowych w gminie można rozpatrywać w dwóch ujęciach. Pierwsze z nich to budowa nowych mieszkań, drugie skupia się zaś na inwestowaniu w istniejące już zasoby.

Tabela 11. Zasoby mieszkaniowe ogółem – mieszkania, 2006 r.

Łódź	Zasoby mieszkaniowe ogółem – 2006 rok	332 692
Kraków		299 754

Źródło: opracowanie własne na podstawie BDL GUS

Tabela 12. Zasoby mieszkaniowe według form własności – zasoby gminne (komunalne), 2006 r.

Łódź	Zasoby gminne (komunalne) – 2006 rok	72 045
Kraków		23 658

Źródło: opracowanie własne na podstawie BDL GUS

Zasady gospodarowania mieszkaniowym zasobem gminy miejskiej Kraków zostały zawarte w Wieloletnim programie gospodarowania mieszkaniowym zasobem gminy miejskiej Kraków na lata 2006–2010. W dokumencie tym założono, że racjonalne gospodarowanie zasobem mieszkaniowym wymaga jego zinwentaryzowania.

Tabela 13. Zasoby mieszkaniowe ogółem – mieszkania, 2017 r.

Łódź	Zasoby mieszkaniowe ogółem – 2017 rok	356 350
Kraków		380 088

Źródło: opracowanie własne na podstawie BDL GUS

Tabela 14. Zasoby mieszkaniowe gmin – lokale socjalne

Lata	Miasto	Wyszczególnienie	Wartość
2017	Łódź	Lokale socjalne	2345
2006			1731
2006	Kraków		715*
2017			3521

* Według stanu na 31 grudnia 2006 roku w ramach posiadanego zasobu gmina miejska Kraków wykorzystywała 715 lokali na cele socjalne.

Źródło: opracowanie własne na podstawie danych Zarządu Budynków Komunalnych w Krakowie oraz BDL GUS (dane dotyczące Łodzi).

Z analizy powyższych danych wynika, że w latach 2006 i 2017 liczba lokali socjalnych w Łodzi zwiększyła o około 35%. W roku 2017 w Łodzi kontynuowano działania zmierzające do zmiany koncepcji zarządzania zasobem lokalowym.

W skład mieszkaniowego zasobu gminy Łódź wchodziło 45 015 lokali mieszkalnych i 7054 niezasiedlone lokale mieszkalne. W 2017 roku kontynuowano prace dotyczące powiększenia zasobu lokali socjalnych w tym mieście. Zasób lokali socjalnych w 2017 roku wynosił 1490 sztuk o powierzchni użytkowej 38,3 tys. m², co stanowiło 3,8% powierzchni użytkowej wszystkich lokali gminnych (2016 – 393 lokale o powierzchni 7,9 tys. m²).

8. Stan zasobu komunalnego w Łodzi i Krakowie w latach 2006 i 2017

Zarządzanie komunalnymi zasobami mieszkaniowymi jest działaniem skierowanym na majątek mieszkaniowy gminy, obejmującym planowanie i podejmowanie decyzji w zakresie budowy nowych mieszkań oraz bieżącej eksploatacji i utrzymania w odpowiednim stanie technicznym istniejących już zasobów. Gminne zasoby mieszkaniowe zostały zdefiniowane w ustawie o ochronie praw lokatorów, zgodnie z którą mieszkaniowy zasób gminy tworzą mieszkania stanowiące własność gminy albo gminnych osób prawnych lub spółek handlowych utworzonych z udziałem gminy, z wyjątkiem towarzystw budownictwa społecznego (TBS). Jest to zatem zasób mieszkaniowy przeznaczony na wynajem, z założenia dla osób, które nie są w stanie zaspokoić swoich potrzeb mieszkaniowych. Mieszkanie komunalne pełni funkcję pomocniczą, a jednocześnie stanowi część majątku gminy, który może być przez nią wykorzystywany w celu pomnażania swoich dochodów. Gmina może zatem mieszkania komunalne wynająć lub sprzedać, a dochody – zarówno z wynajmu mieszkania, jak i jego sprzedaży – są dochodami własnymi gminy.

Mieszkania komunalne stanowią zasób mieszkaniowy należący do gminy i są częścią zasobów publicznych (Sikora-Fernandez, 2011: 181). Zasób mieszkaniowy Łodzi jest słaby pod względem jakościowym. Decyduje o tym nie tylko niekorzystna struktura wiekowa, ale również długoletnie zaniedbania jego stanu technicznego. W połowie lat dziewięćdziesiątych ubiegłego wieku Łódź miała duży zasób mieszkaniowy – na poziomie 92,7 tys., co stanowiło 28,5% komunalnego zasobu mieszkaniowego miasta. W 2000 roku liczba mieszkań będących wyłączną własnością gminy wynosiła 81,2 tys., a w 2005 roku już tylko 72,3 tys. Przeprowadzany proces prywatyzacji przyczynił się do zmniejszania zasobu mieszkań komunalnych o ponad 15 tys. w latach 1995–2002. Udział łódzkiego zasobu komunalnego w zasobach ogółem w 2006 roku wynosił 20,85%.

Tabela 15. Udział zasobu komunalnego w całkowitym zasobie mieszkaniowym w 2006 i 2017 roku w Łodzi

Wyszczególnienie	2006	2017
Zasób mieszkaniowy (ogółem)	332 692	356 350
Zasób komunalny mieszkania	69 379	Brak danych
Procentowy udział zasobu komunalnego w zasobie ogółem	21	Brak danych
Powierzchnia użytkowa mieszkań ogółem (m²)	17 607 396	19 317 806
Powierzchnia użytkowa mieszkań komunalnych (m²)	2 912 933	Brak danych
Procentowy udział zasobu komunalnego w zasobie ogółem	16,54	Brak danych

Źródło: opracowanie własne na podstawie BDL GUS

Zmiany ilościowe w zasobie mieszkaniowym gminy Łódź spowodowane były między innymi procesem prywatyzacji zasobów mieszkaniowych, wyburzeniami i oddaniem nowych mieszkań do użytku przez gminę.

Zasady gospodarowania zasobem komunalnym gminy miejskiej Łódź w 2017 roku były określone na podstawie Wieloletniego program gospodarowania mieszkaniowym zasobem miasta Łodzi na lata 2016–2020. Podstawowym założeniem programu było rozdzielenie obszaru polityki mieszkaniowej zasobu socjalnego od polityki mieszkaniowej pozostałego zasobu komunalnego. Oznacza to zdecydowane wyodrębnienie funkcji komercyjnej zasobu komunalnego od jego funkcji socjalnej (Uchwała Nr XXIV/572/16 Rady Miejskiej w Łodzi, 2016).

Tabela 16. Struktura gminnego zasobu mieszkaniowego w Łodzi w 2006 i 2007 roku

Wyszczególnienie	2006	2007
Budynki mieszkalne ogółem (szt.)	6631	6631
Budynki mieszkalne będące w 100% własnością gminy (szt.)	3631	3631
Lokale mieszkalne ogółem (szt.)	69 379	69 379

Źródło: Kucharska-Stasiak, Załęczna, Żelazowski, 2011

Prognoza ilościowa zasobu mieszkaniowego gminy miejskiej Łódź zakładała stworzenie zasobu lokali socjalnych i pomieszczeń tymczasowych ze wskazaniem na budynki o niskim standardzie, w których zwalniane lokale będą wynajmowane tylko jako lokale socjalne oraz pomieszczenia tymczasowe.

Kraków zarządzał w 2006 roku, poprzez Zarząd Budynków Komunalnych (ZBK), 990 budynkami, 52% z nich było budynkami mieszkalnymi, z czego 62% było własnością miasta, Skarbu Państwa oraz własnością mieszaną, 3% własnością wspólnot mieszkaniowych w zarządzie ZBK, a 35% własnością prywatną. W 2006 roku koszty związane z mieszkaniami komunalnymi przewyższyły o 19% dochody z mieszkań komunalnych ogółem. Zasób mieszkaniowy zarządzany przez ZBK to 568 budynków. Zmniejszanie liczby budynków pozostających w zarządzie ZBK, w tym również tych najstarszych, było głównie wynikiem zwrotu nieruchomości prywatnym właścicielom, przekazywania zarządu budynkami wspólnotom w związku z wykupem lokali przez najemców oraz sprzedaży przez gminę miejską Kraków całych budynków.

Tabela 17. Struktura własnościowa budynków w zarządzie ZBK w 2017 i 2018 roku w Krakowie

Własność	Budynki mieszkalne 2017	Budynki użytkowe 2017	Budynki mieszkalne 2018
Budynki w zarządzie ZBK, z tego:	282	304	326
Własność GMK i/lub Skarbu Państwa	190	289	388
Własność prywatna	35	7	–21
Współwłasność (GMK, własność prywatna)	57	8	–41

Źródło: opracowanie własne na podstawie danych Zarządu Budynków Komunalnych w Krakowie

Tabela 18. Liczba i struktura własności mieszkań położonych w budynkach zarządzanych przez ZBK i w budynkach wspólnot mieszkaniowych znajdujących się poza zarządem ZBK w 2006 roku w Krakowie

Własność budynku	Rodzaj własności mieszkań w zarządzie komunalnym	2006
Liczba mieszkań ogółem zarządzanych przez ZBK, w tym:		5 639
Komunalna	Mieszkanie komunalne	2 770
Prywatna	Mieszkanie prywatne	1 869
Liczba mieszkań komunalnych położonych w budynkach wspólnot mieszkaniowych poza zarządem ZBK		19 969

Źródło: opracowanie własne na podstawie danych Zarządu Budynków Komunalnych w Krakowie

Struktura własności budynków mieszkalnych Łodzi różni się od struktury budynków mieszkalnych w Krakowie, co utrudnia bezpośrednią analizę i wnioskowanie.

9. Zasady polityki czynszowej i zaległości czynszowe w Łodzi i Krakowie w latach 2006 i 2017

Wartość użytkową lokali mieszkalnych różnicuje się w zależności od indywidualnych cech danego lokalu. Średnia stawka w zasobie komunalnym Krakowa wynosiła w 2006 roku 3,84 zł/m². Stawka czynszu za lokal socjalny wynosiła 0,58 zł/m², natomiast w zasobach mieszkaniowych czterech krakowskich Towarzystw Budownictwa Społecznego (TBS-ów) – od 8,46 do 9,15 zł/m². Średnia miesięczna stawka czynszu komunalnego w 2017 roku w Krakowie była niższa w stosunku do lat poprzednich i wyniosła 5,47 PLN/m². W 2017 roku stawka czynszu za tymczasowe pomieszczenie ustalona była w takiej samej wysokości jak za lokal socjalny. W gminie miejskiej Łódź średnia stawka czynszu w kwocie bazowej w okresie od 1.05.2005 do 31.07.2007 roku była równa 3,35 zł/m². Stawka czynszu za lokal socjalny była równa 0,50 zł/m².

Tabela 19. Czynsz komunalny w 2006 roku – Kraków

Wyszczególnienie	2006
Średnia stawka dla zasobu komunalnego (zł/m²/miesiąc)	3,84
Minimalna i maksymalna wartość czynszu komunalnego (zł/m²/miesiąc)	1,16–4,42

Źródło: opracowanie własne na podstawie danych Wydziału Mieszkalnictwa Urzędu Miasta Krakowa

Z analizy danych zamieszczonych w Tabelach 19 i 21 wynika, że średnia stawka czynszu w Łodzi i w Krakowie w 2006 roku różniła się o 0,29 zł/m². Jeśli przyjąć, że średnia jest wielkością reprezentatywną, to różni się o około 7,55%, co nie jest znaczącą różnicą.

Tabela 20. Zaległości czynszowe w zasobie komunalnym gminy miejskiej Kraków (stan na 31 grudnia 2006 r.)

Wyszczególnienie	2006
Zaległości czynszowe z lokali mieszkalnych ogółem + media	82 181 tys.
Ściągalności należności czynszowych mieszkań	87,29%

Źródło: opracowanie własne na podstawie danych Wydziału Mieszkalnictwa Urzędu Miasta Krakowa

Tabela 21. Czynsz w gminie miejskiej Łódź w 2006 roku

Wyszczególnienie	2006
Średnia stawka czynszu w kwocie bazowej w okresie od 1.05.2005 do 31.07.2007 r.	3,35 zł/m²
Stawka czynszu za lokal socjalny	0,50 zł/m²

Źródło: Kucharska-Stasiak, Załęczna, Żelazowski, 2011

Na przestrzeni lat 2005–2011 w Łodzi stawka bazowa wzrosła o 49,9%, stawka maksymalna o 25%, a minimalna o 20,8%.

Tabela 22. Zaległości czynszowe w zasobie lokali mieszkalnych i użytkowych w Łodzi w 2006 roku

Wyszczególnienie	2006
Zaległości czynszowe w lokalach mieszkalnych (mln zł)	140 015 686

Źródło: Kucharska-Stasiak, Załęczna, Żelazowski, 2011

W Krakowie zaległości czynszowe z lokalu mieszkalnego ogółem wraz z mediami były równe na koniec 2006 roku 57% dochodów gminy ogółem (rok wcześniej 52%). Poziom ściągalności należności czynszowych mieszkań będących w zarządzie ZBK w 2006 roku wynosił 89%. W Łodzi w 2006 roku utrzymywał się również wysoki poziom zaległości czynszowych.

Tabela 23. Czynsze w budynkach komunalnych w Krakowie w 2017 roku

Wyszczególnienie	2017
Średnia stawka czynszu komunalnego, według przypisu, w PLN/m²/miesiąc	5,47
Minimalna i maksymalna stawka czynszu komunalnego w PLN/m²/miesiąc	3,08–7,87

Źródło: opracowanie własne na podstawie danych Zarządu Budynków Komunalnych w Krakowie

Tabela 24. Zaległości oraz ściągalność zaległości czynszowych w komunalnych lokalach mieszkalnych w Krakowie w 2017 roku

Wyszczególnienie	2017
Zaległości z tytułu czynszów i opłat za media (w tys. PLN)	240 149
Ściągalność należności czynszowych (w proc.)	93,34

Źródło: opracowanie własne na podstawie danych Zarządu Budynków Komunalnych w Krakowie

Zaległości czynszowe w zasobach spółdzielni mieszkaniowych Krakowa zwiększały się w tempie większym niż w Łodzi, średnio o około 1,23 raza, mimo że tempo wzrostu wynagrodzeń w Krakowie było większe o 1,09.

10. Zasady i proces prywatyzacji w Łodzi i Krakowie jako jeden z kierunków gospodarki mieszkaniowej

Prywatyzacja powinna być poprzedzona odpowiednio przygotowaną długotrwałą koncepcją, nie powinna być nastawiona na bieżące wpływy (Wojciechowski, 2003: 82). Prywatyzacja zasobów mieszkaniowych była powszechnym zjawiskiem od czasu transformacji ustrojowej, która nastąpiła w Polsce po 1989 roku.

Zgodnie z ustawą o samorządzie gminnym oraz ustawą o gospodarce nieruchomościami gminny zasób nieruchomości może być przedmiotem sprzedaży. Zmiany ilościowe w zasobie mieszkań komunalnych w Łodzi zostały spowodowane głównie prywatyzacją zasobów mieszkaniowych. Wyjście ze wspólnoty mieszkaniowej może polegać na zaoferowaniu lokalu zamiennego najemcy, który nie chce skorzystać z możliwości wykupu lokalu. Pierwszeństwo w prywatyzacji powinny mieć lokale w budynkach, w których gmina ma udział do 25%. Jeżeli najemcy w tych budynkach nie wyrażą chęci wykupienia lokalu, powinni być oni przeprowadzeni do lokali zamiennych, a pozyskane lokale gmina powinna sprzedać na wolnym rynku.

Gmina, jako gospodarz miasta, musi dbać o stan techniczny i estetyczny budynków w sposób pośredni, na przykład przez egzekwowanie wymogów prawa budowlanego, co może okazać się bardzo trudne w sytuacji niewydolności finansowej właścicieli budynków (Korniłowicz, 1995: 6). W gminie miejskiej Łódź proces prywatyzacji nie został przeprowadzony prawidłowo, ponieważ – podobnie jak i w innych miastach – przyjęty został model tzw. prywatyzacji rozproszonej, inicjowanej przez najemców, w miejsce prywatyzacji selektywnej, czyli prywatyzacji w wytypowanych przez gminę budynkach. Ten model przyniósł negatywne skutki, ponieważ o wykupienie mieszkania wystąpili najemcy w najlepszych budynkach pod względem lokalizacji, stanu technicznego i standardu użytkowego. W zasobie gminy pozostały mieszkania w najgorszym stanie technicznym i lokalizacyjnym. Ponadto nie wszyscy najemcy wykupili mieszkania w danym budynku, co przyczyniło się do tego, że w budynkach gmina stała się członkiem wspólnoty mieszkaniowej, uczestnicząc w procesie zarządzania tym zasobem. Oznaczało to zobowiązanie wpłacania zaliczki na pokrycie kosztów utrzymania, co prowadziło do rozpraszania posiadanych przez nią środków. Negatywnym zjawiskiem jest wciąż rosnąca liczba wspólnot, w których gmina ma swój udział. Przyjęcie przez miasto programu rewitalizacji wpłynęło na zahamowanie sprzedaży mieszkań na obszarze objętym rewitalizacją.

Należy dążyć do takiego prowadzenia polityki mieszkaniowej, aby prywatyzacja nie stała się przejściową modą, powinny być analizowane korzyści i koszty z nią związane (Kamerschen, McKenzie, Nardinelli, 1991: 95–96). Realizując przyjętą przez gminę Łódź koncepcję prywatyzacji zasobu lokalowego w obrębie polityki mieszkaniowej gminy na lata 2016–2020, w której jako jeden z celów operacyjnych został wymieniony cel „Miasto o rosnącym udziale mieszkań prywatnych", prowadzono następujące działania zmierzające do zmiany w strukturze własności zasobów mieszkaniowych:

1) sprzedaż lokali mieszkalnych i użytkowych na rzecz obecnych najemców,
2) zupełne wyjście z małych wspólnot mieszkaniowych, a następnie ze wspólnot, w których gmina ma mniejszościowy udział,
3) rozpoczęcie procesu prywatyzacji w budynkach całkowicie należących do miasta, w których chęć wykupu zadeklarują najemcy dysponujący co najmniej 50-procentowym udziałem w nieruchomości, przy czym wszystkie lokale są lokalami samodzielnymi.

Podjęta uchwała w sprawie Polityki Miasta Łodzi dotyczącej gminnego zasobu mieszkaniowego 2020+ umożliwiała zmiany zasad sprzedaży lokali mieszkalnych i domów jednorodzinnych na rzecz ich najemców, w związku z planowanym podwyższeniem bonifikat udzielanych przez miasto od wartości określonych w wycenach.

Program „Miasto o rosnącym udziale mieszkań prywatnych" zakłada, że wprowadzone zmiany pozwolą na poszerzenie grupy najemców mogących nabyć lokal mieszkalny lub dom jednorodzinny na własność. Umożliwi to odstąpienie od uzależnienia możliwości nabycia lokalu mieszkalnego przez najemców ze zwiększoną bonifikatą od terminu złożenia wniosku o jego sprzedaż i ogłoszenia w BIP wykazu nieruchomości przeznaczonych do prywatyzacji. Spowoduje także odstąpienie od uzależnienia wysokości bonifikaty przy sprzedaży lokali mieszkalnych na rzecz najemców od okresu najmu lokalu przez osobę ubiegającą się o jego nabycie. Liczba wspólnot mieszkaniowych Łodzi na koniec 2017 roku wyniosła 2116 i zmniejszyła się w stosunku do roku poprzedniego o 36, z czego w 100% sprywatyzowano 35 wspólnot mieszkaniowych.

Prywatyzacja nie powinna zagrażać realizacji ważnych celów publicznych, co upoważnia do wskazywania jej granic (Savas, 1992: 130). Po zaprzestaniu prywatyzacji zasobu będą podejmowane działania zmierzające do osiągnięcia liczby lokali w zasobie, która umożliwi pełną realizację obligatoryjnych zadań ustawowych gminy. Po osiągnięciu tego stanu dopuszcza się wznowienie procesu prywatyzacji lokali, przy czym część zasobu zapewniająca pełną realizację obligatoryjnych zadań gminy zostanie na stałe wyłączona spod prywatyzacji (Wieloletni program gospodarowania mieszkaniowym zasobem gminy miejskiej Kraków oraz zasobem tymczasowych pomieszczeń na lata 2012–2017).

Do czasu osiągnięcia postulowanego stanu prywatyzacja jest dopuszczalna jedynie wyjątkowo, w wypadku zaistnienia konieczności zapewnienia racjonalnego gospodarowania lokalami wchodzącymi w skład zasobu, w szczególności w przypadku wysokich kosztów utrzymania. Celem sprzedaży lokali mieszkalnych stanowiących własność gminy jest racjonalne gospodarowanie zasobem. Podsumowując, sprzedaż mieszkań musi być poprzedzona akcją wyjaśniającą przyszłym nabywcom istotę nabywanych praw oraz obowiązków (Bogusz, 1993: 41). Gminy w dalszym ciągu dążą do prywatyzacji swoich zasobów mieszkaniowych. Większość badanych jednostek samorządu terytorialnego cechuje pasywność inwestycyjna w zakresie gospodarki mieszkaniowej. Długofalowo może to prowadzić do pogłębienia problemu niezaspokojenia potrzeb mieszkaniowych, zwłaszcza wśród osób w trudnej sytuacji materialnej.

Polityka prywatyzacyjna gminy miejskiej Kraków dąży do wstrzymania sprzedaży z bonifikatą lokali mieszkalnych wchodzących w skład zasobu mieszkaniowego. Celem tych działań jest zapobieżenie masowej wyprzedaży lokali mieszkalnych z zasobu za niewielki procent ich wartości rynkowej. Istnieje również możliwość wprowadzenia prywatyzacji zasobu mieszkaniowego, gdy w jego skład będzie wchodziła liczba lokali umożliwiająca pełną realizację obligatoryjnych zadań ustawowych gminy miejskiej Kraków. Część jej zasobu służąca temu celowi ma być całkowicie wyłączona z prywatyzacji. Natomiast przed osiągnięciem proponowanego stanu prywatyzacja jest dopuszczalna jedynie wtedy, gdy wystąpi konieczność zapewnienia racjonalnego gospodarowania lokalami będącymi częścią zasobu gminy miejskiej Kraków.

Z inicjatywy prezydenta Krakowa 27 sierpnia 2003 roku lokale mieszkalne sprzedawane były po cenach preferencyjnych na podstawie uchwały Rady Miasta Krakowa w sprawie zasad zbywania lokali mieszkalnych stanowiących własność Miasta Krakowa (Nr LVIII/471/00). Miało to na celu umożliwienie jak największej liczbie mieszkańców wykupu zajmowanych lokali mieszkalnych.

Praktyka wykazała, że znaczna dekoncentracja udziałów Krakowa we wspólnotach mieszkaniowych i innych współwłasnościach powodowała zwiększenie kosztów utrzymania zasobu mieszkaniowego. Dzięki działaniom polegających na wyprowadzaniu własności miasta z budynków wspólnot mieszkaniowych będzie można między innymi zmniejszyć koszty utrzymania mieszkaniowego zasobu miasta. Stopniowe zmniejszanie liczby wspólnot z udziałem miasta Krakowa jest umotywowane faktem, że gospodarowanie gminnymi udziałami w budynkach wchodzących w skład wspólnot mieszkaniowych nie zależy wyłącznie od chęci miasta. Jak pokazuje praktyka, pomijając wysokość udziału w nieruchomości wspólnej, miasto ma jeden głos w podejmowaniu uchwał przez właścicieli.

Ograniczanie liczby wspólnot przez miasto Kraków realizowane jest dzięki proponowanym zamianom pojedynczych lokali mieszkalnych we wspólnotach i sprzedaży uzyskanych w ten sposób wolnych lokali w drodze przetargu. W sytuacji wysokich nakładów potrzebnych na remonty i stopniowego podwyższania stawek czynszu niezbędne jest, aby środki przekazywane z budżetu gminy nie były mniejsze niż wpływy do budżetu pozyskane z gospodarki mieszkaniowej, w skład której wchodzą między innymi czynsze z lokali mieszkalnych oraz z lokali użytkowych.

11. Wnioski z przeprowadzonej analizy polityki mieszkaniowej w badanych miastach

W artykule poddano analizie lokalną politykę mieszkaniową Łodzi i Krakowa na podstawie potrzeb mieszkaniowych danych społeczności, na tle obowiązujących przepisów wewnętrznych i zewnętrznych państwa. Porównano zmienne charakteryzujące zasoby materialne i społeczne badanych miast. Z przeprowadzonych analiz wynika, że:

1) zarówno w 2006, jak i w 2017 roku liczba ludności w badanych miastach utrzymywała się na podobnym poziomie;
2) w 2006 i w 2017 roku stopa bezrobocia w Krakowie była niższa niż w Łodzi;
3) sytuacja ekonomiczna badanych miast, wyrażona dochodami budżetowymi na jednego mieszkańca, wskazuje, że w obu miastach znacząco wzrosły dochody na jednego mieszkańca (z przewagą w Krakowie);
4) dochody i wydatki budżetów badanych miast na jednego mieszkańca w analizowanych latach były wyższe w Krakowie niż w Łodzi;
5) w Łodzi zarówno w 2006, jak i w 2017 roku dochody budżetu gmin i miast na prawach powiatu na jednego mieszkańca przewyższały średnią krajową o 14%, a w Krakowie odpowiednio w 2006 roku o 26%, a w 2017 roku o 31%, co daje wzrost dochodów o 5% w obu miastach;
6) aktywność gospodarcza w Łodzi i Krakowie pod kątem spółek z kapitałem zagranicznym pokazała przewagę liczby spółek handlowych w Krakowie oraz ich przyrost w 2017 roku;
7) w latach 2006 i 2017 liczba lokali mających status lokali socjalnych w Łodzi zwiększyła się o około 35%;
8) udział łódzkiego zasobu komunalnego w zasobach ogółem w 2006 roku wynosił 20,85%;
9) zmiany ilościowe w zasobie mieszkaniowym gminy Łódź spowodowane były między innymi procesem prywatyzacji zasobów mieszkaniowych, wyburzeniami oraz oddaniem nowych mieszkań do użytku przez gminę; natomiast zmniejszanie liczby budynków w zarządzie ZBK w Krakowie, również tych najstarszych, było głównie wynikiem zwrotu nieruchomości prywatnym właścicielom, przekazywania zarządu budynkami wspólnotom w związku z wykupem lokali przez najemców oraz sprzedaży przez gminę miejską Kraków całych budynków;

10) w obu miastach obserwuje się trend zmniejszania się liczby zasobów spółdzielni mieszkaniowych z tą samą intensywnością; w ostatnich latach intensywniej zmniejszają się zasoby w Łodzi niż w Krakowie, jednak struktura własności budynków mieszkalnych Łodzi różni się od struktury budynków mieszkalnych w Krakowie, co utrudnia bezpośrednią analizę i wnioskowanie.

Z przeprowadzonej analizy czynników lokalnej polityki mieszkaniowej Łodzi i Krakowa wynika, że pomimo podobnej liczby ludności w latach 2006 i 2017 uwarunkowania gospodarcze w wielu aspektach znacząco różnią się na korzyść Krakowa.

Rozłożenie działań polityki mieszkaniowej badanych miast wynika z konieczności ekonomicznych, finansowych oraz gospodarczych. Zakładane przedsięwzięcia mogą nie być w pełni zrealizowane z powodu zupełnie innej sytuacji demograficznej, gospodarczej, ekonomicznej miast, niż zakładają to wieloletnie programy gospodarowania zasobem mieszkaniowym. Powinny one być zgodne z rzeczywistym stanem polityki mieszkaniowej miasta. Dlatego przy ocenie polityki mieszkaniowej dużo większe znaczenie niż odrębny dokument ma suma działań podejmowanych przez władze na wszystkich szczeblach (Twardoch, 2015: 22).

Bibliografia

Acocella N. (2002), *Zasady polityki gospodarczej. Wartości i metody analizy*, Wydawnictwo Naukowe PWN, Warszawa.

Andrzejewski A. (1987), *Polityka mieszkaniowa*, Polskie Wydawnictwo Ekonomiczne, Warszawa.

Bank Danych Lokalnych GUS, https://bdl.stat.gov.pl/BDL/dane/podgrup/teryt (dostęp: 23.03.2020).

Bogusz J. (1993), *Kierunki prywatyzacji zasobu mieszkaniowego w krajach Europy Środkowej i Wschodniej*, Instytut Gospodarki Mieszkaniowej, Warszawa.

Cyran R. (2013), *Zintegrowana polityka mieszkaniowa jako instrument rozwoju lokalnego*, Uniwersytet Ekonomiczny w Katowicach, Katowice.

Declaration on Social Progress and Development Proclaimed by General Assembly resolution 2542 (XXIV) of 11 December 1969, https://www.ohchr.org/Documents/ProfessionalInterest/progress.pdf (dostęp: 23.03.2020).

Deklaracja Światowej Konferencji ONZ Habitat II, https://undocs.org/A/CONF.165/14 (dostęp: 23.03.2020).

Europejska Karta Społeczna – wersja zrewidowana, podpisana w Strasburgu dnia 3 maja 1996 r., https://rm.coe.int/europejska-karta-spoleczna/16808b6383 (dostęp: 23.03.2020).

Kamerschen D., McKenzie R.B., Nardinelli C. (1991), *Ekonomia*, Fundacja Gospodarcza NSZZ „Solidarność", Gdańsk.

Karta Praw Podstawowych Unii Europejskiej, http://oide.sejm.gov.pl/oide/?option=com_content&view=article&id=14428&Itemid=422 (dostęp: 23.03.2020).

Konstytucja Rzeczypospolitej Polskiej z dnia 2 kwietnia 1997 r. (Dz.U. z 1997 r. Nr 78, poz. 483 ze zm.), https://www.sejm.gov.pl/prawo/konst/polski/kon1.htm (dostęp: 23.03.2020).

Konwencja o prawach dziecka przyjęta przez Zgromadzenie Ogólne Narodów Zjednoczonych dnia 20 listopada 1989 roku, http://prawo.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU19911200526 (dostęp: 23.03.2020).

Korniłowicz J. (1995), *Problemy prywatyzacji zasobów mieszkaniowych w Polsce*, II Konferencja Naukowa „Reformowanie Mieszkalnictwa w Polsce", Warszawa, czerwiec.
Kucharska-Stasiak E., Załęczna M., Żelazowski K. (2011), *Stan gospodarki mieszkaniowej w gminie Łódź – pożądane kierunki zmian*, Urząd Miasta Łodzi, Łódź.
Kulesza H., Nieciuński W. (red.) (1989), *Mieszkalnictwo: zagrożenia i szanse rozwoju (zbiór opracowań)*, Państwowe Wydawnictwo Ekonomiczne, Warszawa.
Lis P. (2005), *Koncepcje polityki mieszkaniowej*, „Zeszyt Naukowy" nr 31, Katedra Polityki Gospodarczej i Planowania Rozwoju, Akademia Ekonomiczna w Poznaniu, Poznań, https://docplayer.pl/48577153-Koncepcje-polityki-mieszkaniowej.html (dostęp: 23.03.2020).
Lis P. (2017), *Zrównoważony rozwój rynków mieszkaniowych jako współczesne wyzwanie dla polityki mieszkaniowej państwa*, „Problemy Polityki Społecznej. Studia i Dyskusje", nr 37(2), s. 53–71.
Lis P. (2018), *Od uniwersalności do selektywności w społecznej polityce mieszkaniowej*, „Acta Universitatis Lodziensis. Folia Oeconomica", nr 2(334), s. 67–81.
Międzynarodowy Pakt Praw Gospodarczych, Społecznych i Kulturowych (Dz.U. z 1977 r. Nr 38, poz. 169).
Opinia Europejskiego Komitetu Ekonomiczno-Społecznego w sprawie: „Mieszkalnictwo a polityka regionalna" (Dz.U. UE 2007/C161/03).
Opinia Komitetu Regionów Polityka mieszkaniowa i Regionalna (Dz.U. UE 2007/C146/02).
Parlament Europejski, Komisja Rozwoju Regionalnego, Dokument roboczy w sprawie mieszkalnictwa i polityki regionalnej, 8 listopada 2006 r., sprawozdawca: A. Andria, PE 380.837v01–00.
Powszechna Deklaracja Praw Człowieka, rezolucja Zgromadzenia Ogólnego ONZ, nr 217A(III) z dnia 10 grudnia 1948 r., https://amnesty.org.pl/wp-content/uploads/2016/04/Powszechna_Deklaracja_Praw_Czlowieka.pdf (dostęp: 23.03.2020).
Savas E.S. (1992), *Prywatyzacja. Klucz do lepszego rządzenia*, Polskie Wydawnictwo Ekonomiczne, Warszawa.
Sikora-Fernandez D. (2011), *Budownictwo socjalne na tle zrównoważonego rozwoju miast*, „Acta Universitatis Lodziensis. Folia Oeconomica", nr 261, s. 433–441.
Twardoch A. (2015), *Centralna, regionalna i lokalna polityka mieszkaniowa w kontekście prognozowanych zmian demograficznych. Wybrane skutki przestrzenne, społeczne i gospodarcze*, „Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach", nr 222, s. 21–31.
Uchwała nr XXIV/572/16 Rady Miejskiej w Łodzi z dnia 3 lutego 2016 r. w sprawie przyjęcia Wieloletniego programu gospodarowania mieszkaniowym zasobem Miasta Łodzi na lata 2016–2020.
Ustawa z dnia 8 marca 1990 r. o samorządzie gminnym (Dz.U. z 1990 r. Nr 16, poz. 95).
Ustawa z dnia 21 czerwca 2001 r. o dodatkach mieszkaniowych (Dz.U. z 2001 r. Nr 71, poz. 734).
Ustawa z dna 21 czerwca 2001 r. o ochronie praw lokatorów, mieszkaniowym zasobie gminy i zmianach Kodeksu cywilnego (Dz.U. z 2001 r. Nr 71, poz. 733).
Ustawa z dnia 12 marca 2004 r. o pomocy społecznej (Dz.U. z 2004 r. Nr 64, poz. 593).
Wieloletni program gospodarowania mieszkaniowym zasobem Gminy Miejskiej Kraków oraz zasobem tymczasowych pomieszczeń na lata 2012–2017, http://www.bip.krakow.pl/_inc/rada/posiedzenia/show_pdfdoc.php?id=64188 (dostęp: 23.03.2020).
Wojciechowski E. (2003), *Zarządzanie w samorządzie terytorialnym*, Wydawnictwo Difin, Warszawa.
Wydział Mieszkalnictwa Urzędu Miasta Krakowa, https://www.bip.krakow.pl (dostęp: 23.03.2020).
Zarząd Budynków Komunalnych w Krakowie, https://zbk-krakow.pl/ (dostęp: 23.03.2020).

Assessment of the Local Housing Policy of Łódź and Kraków in the Light of Selected Aspects

Abstract: The article carried out a study on the local housing policy of Łódź and Kraków in the year 2006 and 2017. A comparative analysis was carried out in terms of population, housing stock (including social and municipal flat), unemployment rate, income and expenditure of the budgets of the cities studied, the economic situation expressed in income budget per capita. The results of the quantitative analyzes carried out show the trend of decreasing the amount of housing stock in the examined cities. In recent years housing stock in Łódź have been decreasing more intensively than in Kraków, and the ownership structure of Łódź residential buildings is different from that of residential buildings in Krakow.

Keywords: municipal and social housing, local housing policy, municipal and social housing stock, demand

JEL: R21

OPEN ACCESS

© by the author, licensee Łódź University – Łódź University Press, Łódź, Poland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license CC-BY (https://creativecommons.org/licenses/by/4.0/)

Received: 2020-04-06; verified: 2020-07-18. Accepted: 2020-08-24

COPE Member since 2018 JM13703

This journal adheres to the COPE's Core Practices https://publicationethics.org/core-practices

Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663
www.czasopisma.uni.lodz.pl/foe/
4(349) 2020

DOI: http://dx.doi.org/10.18778/0208-6018.349.03

Aleksandra Kupis-Fijałkowska
University of Łódź, Faculty of Economics and Sociology, Department of Statistical Methods
aleksandra.kupis-fijalkowska@uni.lodz.pl

Selected Problems of Quality Assessment in Internet Surveys – a Statistical Perspective

Abstract: The paper presents selected problems related to the quality assessment from the statistical perspective of survey data based on Internet sources. Internet access is consequently expanding all over the world. In parallel with the running development of other new technologies, it is pervading daily life and business activities more and more. It also has influenced surveys practice to a large extent as a research tool for collecting both primary and secondary data, and it also challenges surveys to research the Internet population. Moreover, as the Internet and its entities are able to register all activities that are performed on the web, issues related to big data and organic data processing as well as their applications arise. As a result of decreasing response rates and increasing survey costs, Internet data collection is constantly growing. Due to many advantages, Internet surveys are used widely and this process seems to be inevitable. However, it needs to be emphasised that Internet surveys are developing in practice faster than the methodology in this area. Hence, a lot of problems can be identified, especially when considering the quality of data based on Internet sources. The following issues are discussed as the most far-reaching in the prism of statistical survey methodology: determination of the sampling frame, self-selection and related estimates bias, as well as under/over-coverage.

Keywords: Internet survey, online survey, survey quality, survey error

JEL: C8

1. Introduction

Internet coverage and its penetration rate are constantly growing. In parallel, interest in and usage of Internet sources and resources are increasing. This includes scientists, researchers, students, and occasional users. There are many postulates regarding today's surveys, however, the requirement of providing up to date data, delivered as fast as possible with the lowest possible cost, is the main reason that seems to encourage support for offline modes with Internet data collection or transfer of surveys completely to the web space (Bethlehem, 2010; Bethlehem, Biffignandi, 2011; Tourangeau, Conrad, Couper, 2013; Schonlau, Couper, 2017; Kalton, 2018). "The use of Internet for collecting survey-type data has grown enormously in recent years. [...] However, the quality of the estimates produced is questionable" as G. Kalton wrote (Kalton, 2018: S12). And the theory of statistics is challenged to assess that quality as well as to make recommendations what statistical methods can be applied to improve the quality of the results. As it is invertible, the methodology must commensurate to the progress that occurs in practice. Researchers and recipients must be aware of its properties and a great deal of attention should be paid to the quality assessment (Szreder, 2017). The issue is complex, as it affects many areas of survey methodology (Schonlau, Cooper, 2017; de Leeuw, 2018).

2. Internet coverage & Internet population – introduction and influence on surveys

Internet coverage is constantly growing all over the world, its penetration is becoming wider and deeper. The continuous development of new technologies strengthens the effect of omnipresence of the Internet. Its applications and meaning are expanding for both individuals and corporate users. As the development of the information society is progressing, data demand is increasing. The Internet has had a significant impact on surveys by providing broader possibilities in the data collection process with lower costs. It has become a communication tool, a medium, and an easily accessible source of data. It is a social and business space now: individual users, social media, e-commerce; banking and accounting portals; news; government, public institutions, non-government organisations; corporations and enterprises. A hitherto unknown new dimension of human and business life has been created and the boundary between reality and virtuality is blurred now. The term of virtual society (understood as a sub-population of entities that have and use Internet access) has been introduced and, from the scientific point of view, a new collectivity has come to life: the Internet population – the population of Internet users. When studying the Official Statistics reports and different organisations' elaborations dedicated to the Internet and Internet surveys, it can be observed that as Internet coverage is rising, also the surveys conducted via and on the Internet are gaining in popularity. To illustrate it based on an example, a case of Poland will be presented. Currently, 84% households in Poland have Internet access. In Table 1 detailed statistics are listed for EU countries.

Table 1. Households – level of Internet access [%] in EU countries in 2010–2018

GEO/TIME	2010	2011	2012	2013	2014	2015	2016	2017	2018
European Union	70	73	76	79	81	83	85	87	89
Belgium	73	77	78	80	83	82	85	86	87
Bulgaria	33	45	51	54	57	59	64	67	72
Czechia	61	67	73	73	78	79	82	83	86
Denmark	86	90	92	93	93	92	94	97	93
Germany	82	83	85	88	89	90	92	93	94
Estonia	67	69	74	79	83	88	86	88	90
Ireland	72	78	81	82	82	85	87	88	89
Greece	46	50	54	56	66	68	69	71	76
Spain	58	63	67	70	74	79	82	83	86
France	74	76	80	82	83	83	86	86	89
Croatia	56	61	66	65	68	77	77	76	82
Italy	59	62	63	69	73	75	79	81	84
Cyprus	54	57	62	65	69	71	74	79	86
Latvia	60	64	69	72	73	76	77	79	82
Lithuania	61	60	60	65	66	68	72	75	78
Luxembourg	90	91	93	94	96	97	97	97	93
Hungary	58	63	67	70	73	76	79	82	83
Malta	70	75	77	78	80	81	81	85	84
Netherlands	91	94	94	95	96	96	97	98	98
Austria	73	75	79	81	81	82	85	89	89
Poland	63	67	70	72	75	76	80	82	84
Portugal	54	58	61	62	65	70	74	77	79
Romania	42	47	54	58	61	68	72	76	81
Slovenia	68	73	74	76	77	78	78	82	87
Slovakia	67	71	75	78	78	79	81	81	81
Finland	81	84	87	89	90	90	92	94	94
Sweden	88	91	92	93	90	91	94	95	92
United Kingdom	80	83	87	88	90	91	93	94	95
Iceland	92	93	95	96	96	na	na	98	99
Norway	90	92	93	94	93	97	97	97	96

Source: Eurostat, 2018

According to Statistics Poland's report "Information society in Poland. Results of statistical surveys in the years 2014–2018", in 2018 in Poland: 97.4% households with children and 95.6% of all enterprises had Internet access. Figure 1 presents how the Internet coverage has been growing in Poland since 2000.

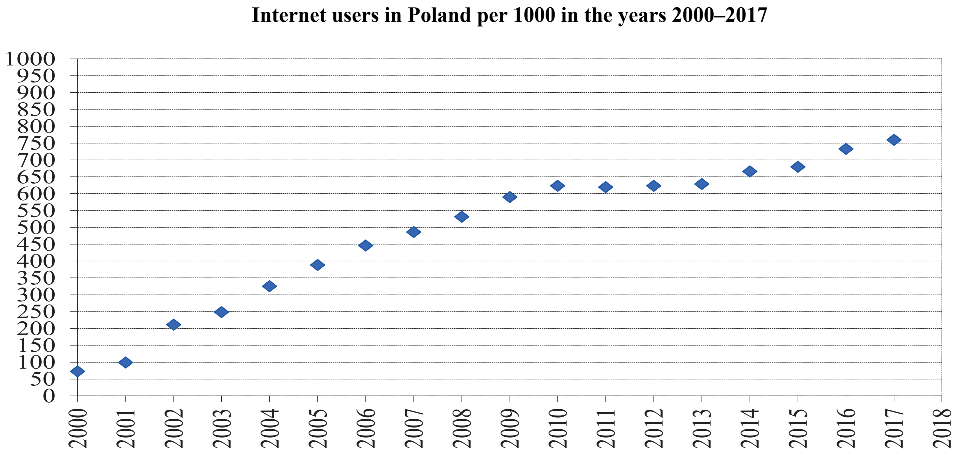


Figure 1. Internet users in Poland per 1000 population from 2000 to 2017

Source: own elaboration based on Statistics Poland data

The presented above descriptive statistics for Poland and different world regions regarding Internet access provide a good proof of the growing power of the Internet. There are no doubts that era of digitisation has come, and it is a natural consequence that surveys have to reach into online sources (Bethlehem, 2010; Callegaro, Manfreda, Vehovar, 2015).

It is important to mention smartphone users statistics here, as these devices connect to the web, which intensifies Internet penetration. The Polish Internet Survey by Gemius S. A. (2018) provides monthly data about Internet users as well as most popular websites and applications. According to the November 2018 report, 23.4 million Internet users in Poland were connecting via smartphone. And, according to the already mentioned report of Statistics Poland "Information society in Poland. Results of statistical surveys in the years 2014–2018", in 2018, in Poland, 47.4% of individuals had access to the Internet via a mobile phone or smartphone.

As Internet access expands, the ratio of research based on Internet surveys is constantly growing, The Polish Society of Market and Opinion Researchers in its Yearbooks presents each year statistics about the situation of market and opinion research in Poland.

In 2008, 2.9% respondents in the market and opinion research in Poland were contacted by CAWI, 5 years later in 2012 it was nearly 25% and in 2017 the figure exceeded 50%, so more than half of all respondents were interviewed this way. In comparison to CATI, it can be observed that in the years 2008–2014 this mode was oscillating around 1/3 of all modes, and in 2015 a trend change occurred and it decreased to 27%, and then respectively to 25% in 2016 and to 21% in 2017.

Table 2. The CAWI and CATI modes by respondents [%] in the Polish market research and opinion sector from 2008 to 2017

Year/Method	CAWI	CATI
2008	2.9	29.6
2009	7.1	36.1
2010	18.3	33.4
2011	21.3	33.0
2012	24.6	32.5
2013	23.5	30.0
2014	30.3	31.0
2015	36.0	27.0
2016	48.0	25.0
2017	53.0	21.0

Source: own elaboration based on Yearbooks 2011/2012 to 2018/2019 – data of the Polish Society of Market and Opinion Researchers

The Official Statistics also recognises a great opportunity to conduct research on the Internet. For example, in Poland, in the National Census 2011, Statistics Poland used a mixed mode for its data collection process and Poles had the ability to decide on online self-interviewing (CAII) – all together around 12% respondents preferred this way of contact. Internet surveys sources, same as Big Data, have a huge potential to support official statistics, probably complementarily (Szreder, 2015; de Leeuw, 2018), however, using Internet based data sources for official statistics purposes at the moment is under discussion: scientists and statistical experts working groups are investigating the potential of available e-sources (Beręsewicz, Szymkowiak, 2015). Methodological studies are being carried out how to merge this type of sources to the official statistics area and how it could work with current legal regulations and good practices.

In summary, the development of new technologies has already influenced the survey execution process, and it is expected by many authors that a deeper influence will be observed in the future (de Leeuw, 2018; Kalton, 2018).

The terms "Internet survey" and "web survey" can be used interchangeably or can be understood differently, as they may be considered in different context/meanings, i.e. the mode of contact, the mode of response, or they may refer to the population of Internet users. Bethlehem and Biffignandi (Bethlehem, Biffignandi, 2011) proposed the following definitions:

> ***Internet survey*** is a general term for various forms of data collection via the Internet (i.e. a web survey, an e-mail survey), also all forms of data collection that use the Internet to transfer questionnaires and collected data between entities of interest;

> ***Web survey*** is a form of data collection via the Internet in which respondents complete questionnaires on the World Wide Web, the questionnaire is accessed by means of a link to a web page.

For the purpose of this article, the definitions given by Bethlehem and Biffignandi apply.

Bethlehem and Biffignandi (2011) also introduced the definition of self-selection survey, which will be referred to later in this paper, as:

> ***Self-selection survey*** is a survey for which the sample has been recruited by means of self-selection, hence users can decide whether or not to participate in the survey.

Many approaches can be found in the literature in the context of the mentioned definitions of the analysed terms (Bethlehem, Biffignandi, 2011; Tourangeau et al., 2013; Fielding, Lee, Blank, 2017), and new concepts, more detailed, are proposed as well. For example, in the prism of self-selection issue and entity responsible for data maintenance, Beręsewicz (Beręsewicz, 2015; 2017) introduced the following Internet data source (IDS) definition:

> ***Internet data source (IDS)*** is a self-selected (non-probabilistic) sample that is created through the Internet and maintained by entities external to NSIs and administrative regulations.

3. Internet surveys – benefits and problems

The Internet is already a successful tool for surveys, the main reason lies in many technical opportunities which it gives to researchers. It offers a broad spectrum of new tools, for example, in-time dynamic question adjustments, reaction time or mimics can be measured, or new multimedia tools are available: animations, movies, sound, high contrast interface, or online eye tracking. Regarding conducting surveys not by but on the Internet, its popularity is caused, as already mentioned, by growing Internet coverage and by the phenomenon that a large part of human life moves to the web relations building, shopping, paying bills, e-medicine, e-pharmacy, watching nature and entertainment places via cameras, voting, etc. Also, modern business depends more and more on the web and a lot of enterprises cooperate more online than offline.

Some advantages and disadvantages were already mentioned but here is a synthetic list of the most important ones (based on: Fricker, Schonlau, 2002; Bethlehem, Biffignandi, 2012; Tourangeau et al., 2013; Fielding, Lee, Blank, 2017).

The most visible benefits:

- quicker and cheaper data collection (at all stages of the data collection process);
- simplicity in comparison to other modes and attractive multimedia forms;
- quick respondent selection on the basis of required features (questionnaires can be filled with already available information, i.e. digital traces);
- no interviewer effect, higher individualisation;
- less intrusive and suffer less from social desirability effects;
- immediately sent and answered questionnaires, quick follow-ups and reminders;
- dynamic sequences of questions adapted to the specific respondent, which results in lower respondent burden and introduction of small modifications;
- reduction of the number of missing responses and partial answers as well as data entry errors;
- lower time and space respondent burden, the response burden can be easily monitored as server-side and client-side information is available;
- a new understanding of individual's anonymity and intimacy (it allows researchers to reach niche populations' opinions easier and investigate rare features more effectively).

And, respectively the list of the most visible disadvantages looks as follows:

- inability to construct a comprehensive sampling frame (can't identify all members of the Internet population and hence unable to apply the assignment rule[1]) that results in sample selection limitation as well as a lack of representativity and biased estimations;
- self-selection;
- coverage problems;
- low response rates;
- problem with bias measurement and quality assessment;
- technological exclusion and problem with respondents' computer skills;
- technical problems can occur;
- inability to confirm respondents' identity;
- "professional" respondents, multiple participation;
- unusual real-time situations can create problems resulting in discontinuation of answering.

In summary, from the statistical point of view, a lack of representativity (from the perspective of the probabilistic survey theory) is the main cause of reducing quality: the inability to define the sampling frame means that selection methods are extremely reduced, and in majority of cases the target population differs from the survey population (coverage problems).

1 It is possible only for specific websites and if the page administrator keeps a registry of users.

4. Internet surveys data quality – a statistical perspective

To reach the most possible reliable information, it is crucial to make a solid research design as well as to choose and apply data collecting methods properly. It allows researchers to know all the details through the survey realisation process and be aware of all existing complications as well as possible error sources. Preceding the further considerations, the classic theory of survey sampling should be presented. Generally, in the early fifties of the twentieth century, the methodology of survey sampling was completed and became a common practice for the official statistics systems, as well as scientific and private sector research (Bethlehem, 2009). If Internet surveys are taken into consideration, fundamental principles of probability sampling and survey theory are not applied (Bethlehem, 2009), which results in the lack of representativity that generates low quality data. Especially, in the context of growing web surveys popularity, the obtained results are published frequently and their recipients are getting more familiarised with this type of surveys, so the results might be perceived as reliable, while they are not. It is observed that full information about the data collection process, problems, and their consequences is not revealed. In the context of probability sampling approach attributes, there are a lot of methodological issues to be solved in the nearest and further future. There are three main problems from the statistical point of view: Internet under/over-coverage, determination of the sampling frame and respondents' self-selection. All of the aforementioned issues result in a lack of (full) representativity, and thereby do not reflect the exact nature of the phenomena studied, so the quality is not sufficient. At the same time, some statistical tools exist and their implementation can improve the quality by toning down discrepancies, low precision and poor accuracy effects.

Probability sampling is crucial to obtaining the most possible reliable information. Selection of data collecting methods and a high quality survey execution process are crucial as well. A lot of surveys suffer from a lack of representativity, which causes the reliability of the collected data to be lower than it could.

The first three of the disadvantages listed above are the main methodological problems in web surveys from the statistical perspective, due to the generated bias: estimations based on the collected material differ significantly from the population parameters and no valuable inferences can be drawn about the researched phenomenon. Hence, the main objective of the conducted survey – obtaining reliable information – is not achieved. The bias in general can be caused by many errors that can occur in the survey execution process (Figure 2).

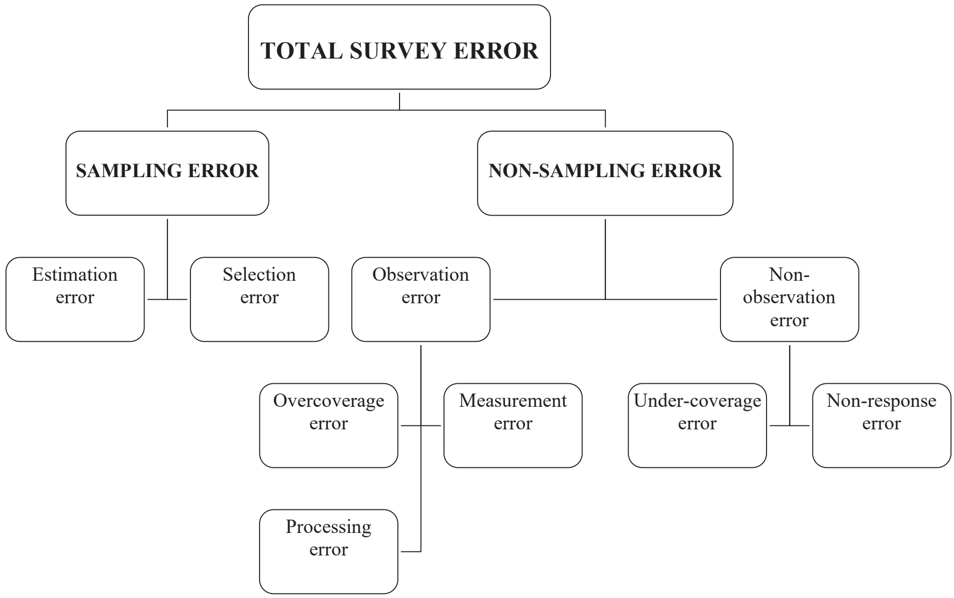


Figure 2. Taxonomy of survey errors
Source: Bethlehem, 2010: 164

Let us consider selected problems concerning data quality when using the Internet for collecting survey-type data that have the most far-reaching consequences from the statistical perspective.

From the prism of the statistical survey theory, it should be done in the context of the presented above breakdown of errors: undercoverage and selection errors that occur here. The first type of errors is the consequence of the inability to build the sampling frame, so no proper selection method can be applied. Hence, basically no proper random sample is selected and a self-selection situation occurs. It means that the respondent has to be aware of the existence of the questionnaire and has to decide to fill it. The other error source is obvious: not all elements of the target population have Internet access. Hence, there is no chance those units can be contacted and interviewed (Bethlehem, 2010).

A short statistical investigation will be introduced now in order to present how the bias caused by undercoverage error can be measured (Bethlehem, 2010). Let us consider the target population of N fully identifiable elements (each element k is labelled; $k = 1, 2, 3, \ldots, N$) and the target variable Y, where for each element k, a value Y_k exists. Let us assume that the web survey aims to estimate the value of the population simple mean for the target variable Y given as:

$$\overline{Y} = \frac{1}{N}\sum_{k=1}^{N} Y_k \ . \tag{1}$$

The population U is divided into two subpopulations, U_I – all elements with Internet access and U_{NI} – all elements without Internet access. Let each element k be characterised by the I_k indicator which:

$$I_k = \begin{cases} 1 \text{ for } k \in U_I \\ 0 \text{ for } k \in U_{NI} \end{cases} . \tag{2}$$

Hence, the number of U_I (Internet population) is equal to:

$$N_I = \sum_{k=1}^{N} I_k . \tag{3}$$

Respectively N_{NI} denotes the U_{NI} (non-Internet population) number, where:

$$N = N_{NI} + N_I . \tag{4}$$

The mean of the target variable for the U_I population is equal to:

$$\bar{Y}_I = \frac{1}{N_I} \sum_{k=1}^{N} I_k Y_k \tag{5}$$

and the mean of the target variable for the U_{NI} population is equal to:

$$\bar{Y}_{NI} = \frac{1}{N_{NI}} \sum_{k=1}^{N} (1 - I_k) Y_k . \tag{6}$$

Let us assume now that the sampling frame can be constructed for the Internet population and a random sample (simple random sampling scheme without replacement) represented by the following series is selected:

$$s_1, s_2, s_3, \dots, s_{N-1}, s_N \tag{7}$$

of N indicators, where the k^{th} indicator s_k assumes 1 if element k is selected and 0 if it is not, for $k = 1, 2, 3, \dots, N - 1, N$. Hence, the sample size is equal to:

$$n_I = s_1 + s_2 + s_3 + \dots + s_{N-1} + s_N = \sum_{k=1}^{N} s_k . \tag{8}$$

The first-order inclusion probability of the k^{th} element is defined by the following expected value:

$$\pi_k = E(s_k). \tag{9}$$

The Horvitz-Thompson estimator for the mean of the U_I population is defined by:

$$\overline{y}_{HT} = \frac{1}{N_I}\sum_{k=1}^{N} s_k I_k \frac{Y_k}{\pi_k}. \tag{10}$$

The inclusion probability π_k for all elements outside the Internet population is equal to 0:

$$\pi_k = 0. \tag{11}$$

When we deal with a simple random sample from the Internet population, all inclusion probabilities are equal to:

$$\pi_k = \frac{n}{N_I}. \tag{12}$$

Hence, expression (10) reduces to:

$$\overline{y}_I = \frac{1}{n}\sum_{k=1}^{N} s_k I_k Y_k. \tag{13}$$

Expression (13) represents an unbiased estimator of the mean $\overline{Y}_I$ given by expression (5), but not necessarily of the mean $\overline{Y}$ given by expression (1).

Let us denote $B(\overline{y}_{HT})$ as the estimator bias, in the discussed situation, it is equal to:

$$B(\overline{y}_{HT}) = E(\overline{y}_{HT}) - \overline{Y} = \overline{Y}_I - \overline{Y} = \frac{N_{NI}}{N}(\overline{Y}_I - \overline{Y}_{NI}). \tag{14}$$

Expression (14) shows that the magnitude of this bias is determined by the following two factors:

– the relative size of $\frac{N_{NI}}{N}$ of the U_{NI} population, and the larger this proportion is, the higher bias occurs;

– the difference $(\overline{Y}_I - \overline{Y}_{NI})$, and the larger this difference is, the higher bias occurs.

As not everyone has web access, two sub-populations exist: the Internet population and the non-Internet population. Their structures can differ, for example, while considered through the prism of age, structures of the U_I and U_{NI} populations can be much different.

The next quality issue that should be discussed is the self-selection problem (Bethlehem, 2010). As the participation requires the awareness of the existence of the survey, and then the decision whether to participate in it or not, this means that each element k (k = 1, 2, 3, …, N – 1, N) of the Internet population has unknown probability ρ_k of individuals participating in the survey. The responding elements are denoted by a vector:

$$r_1, r_2, r_3, \ldots, r_{N-1}, r_N, \tag{15}$$

where $r_k = 1$ if the k^{th} element responds and $r_k = 0$ if it does not, for k = 1, 2, 3, …, N – 1, N. Let the probability of response of element k be given as the expected value $\rho_k = E(r_k)$.

Considering U_{NI}, all response probabilities for elements in the non-Internet population are 0.

The obtained sample size is denoted by:

$$n_S = r_1 + r_2 + r_3 + \ldots + r_{N-1} + r_N = \sum_{k=1}^{N} r_k . \tag{16}$$

If every element in the Internet population had the same probability of being included in the sample, then the estimator for the population mean would be expressed as:

$$\overline{y}_S = \frac{1}{n_S} \sum_{k=1}^{N} r_k Y_k \tag{17}$$

and its expected value would be approximately equal to:

$$E(\overline{y}_S) \approx \overline{Y}_I^* = \frac{1}{N_I \overline{\rho}} \sum_{k=1}^{N} \rho_k I_k Y_k , \tag{18}$$

where $\overline{\rho}$ is the mean of all response propensities in the Internet population.

It can be shown (Bethlehem, 2010) that the bias of the estimator given by (17) can be expressed as:

$$B(\bar{y}_S) = E(\bar{y}_S) - \bar{Y}_I \approx \bar{Y}_I^* - \bar{Y}_I = \frac{\text{cov}(\rho, Y)}{\bar{\rho}} = \frac{R_{\rho Y} SD_\rho SD_Y}{\bar{\rho}}, \tag{19}$$

in which the covariance between the values of the target variable and the response probabilities in the Internet population is given as:

$$\text{cov}(\rho, Y) = \frac{1}{N_I} \sum_{k=1}^{N} I_k (\rho_k - \bar{\rho})(Y_k - \bar{Y}) \tag{20}$$

and respectively:
$\bar{\rho}$ is the average response probability;
$R_{\rho,Y}$ is the correlation coefficient between the target variable and the response behaviour;
SD_ρ is the standard deviation of the response probabilities;
SD_Y is the standard deviation of the target variable.

In the case of self-selection, the bias is determined by the following factors:
- the average response probability;
- the variance of response probabilities;
- the relationship between the target variable and the response behaviour.

As the general population is considered, the bias of the sample mean consists of under-coverage and self-selection biases and can be expressed as:

$$B(\bar{y}_S) = E(\bar{y}_S) - \bar{Y} = E(\bar{y}_S) - \bar{Y}_I + \bar{Y}_I - \bar{Y} = \frac{N_{NI}}{N}(\bar{Y}_I - \bar{Y}_{NI}) + \frac{\text{cov}(\rho, Y)}{\bar{\rho}}. \tag{21}$$

There are different methods to reduce the bias of the estimates in such cases and increase informativity of Internet survey results (Bethlehem, 2010). The most popular ones are weighting adjustment methods, including post-stratification weighting, weighting adjustment with a reference sample, propensity score adjustment, and rim weighting. However, it should be emphasised that only from the theoretical point of view those methods should be sufficient to deal with the bias. In practice, the application of those techniques does not result in the bias elimination but only allows for some reduction of it (Bethlehem, Biffignandi, 2012).

Internet surveys, in general, suffer from a problem of nonresponse (unit nonresponse or item nonresponse). It is the most recognised source of errors from the statistical point of view (Schouten et al., 2012). In the case of web surveys, Bethlehem (2012) has shown that the expression for the bias in the case of random sample affected by nonresponse is identical as (19), as the magnitude of the nonresponse bias is equal to:

$$B(\bar{y}_R) = \tilde{Y} - \bar{Y} = \frac{\mathrm{cov}(\rho, Y)}{\bar{\rho}} = \frac{R_{\rho Y} SD_{\rho} SD_Y}{\bar{\rho}}. \tag{22}$$

This means that in the case of web surveys, the bias generated by self-selection corresponds to the non-response one.

The non-response is recognised as a serious source of survey errors. The related bias of estimates is determined by two factors (Skinner et al., 2009):

– how respondents and non-respondents differ, on average, with respect to the target variable (the contrast between response and non-response);
– the number of responses in the survey (the response rate sets a bound to the maximal impact of non-response).

To assess the effects of non-response on the quality of estimators, both the response rate itself and the contrast (between respondents and non-respondents) should be investigated. It is discussed in the literature (Groves, Peytcheva, 2008; Schouten, Cobben, Bethlehem, 2009) that response rates by themselves are not sufficient indicators of the non-response bias. Schouten Cobben and Bethlehem (2009) found that increases in response rates due to follow-up efforts did not significantly improve response representativeness.

To complete the quality assessment based on the response rate, supplemental survey quality measures are proposed, including: R-indicators (Representativeness indicators), bias reduction indicators, Mahalanobis distance, response rates for key domains, or tracking key survey estimates.

Currently, in the context of Internet surveys, the R-indicators concept seems to be the most widely discussed in the literature as a supplemental quality measure to the response rate (Shlomo et. al., 2008). Although the response rate should be treated as the core indicator of the survey quality, it does not necessarily express all the aspects that influence the representativity of the survey results suffering from non-response. In this paper, the R-indicator as a measure based upon the variance of estimated response probabilities (Cobben, Schouten, 2005; 2007; Schouten, Cobben, Bethlehem, 2009) will be discussed.

Let us suppose that a sample survey is undertaken where a sample s is selected from a finite population U. The sizes of s and U are denoted n and N, respectively. The units in U are: $i = 1, 2, \ldots, N$. The sample is assumed to be drawn by the probability sampling design $p(.)$ where the sample s is selected with probability $p(s)$.

Let us denote s_i as the 0–1 sample indicator (if unit i is sampled, it takes the value 1 and 0 otherwise), r_i as the 0–1 response indicator for the unit i (if unit i is sampled and did respond, it takes the value 1 and 0 otherwise), so the set of respondents is given as r ($r \subset s \subset U$) and π_i as the first-order inclusion probability of unit i. Let us assume that no-response occurs.

Let ρ_i be the probability that the unit *i* responds when it is sampled. Let us consider that response propensity is motivated by a variable *X* (more than one could be assumed), then the expected conditional response propensity is given as:

$$\rho_i = \rho_X(x_i) = E(R_i \setminus x_i) . \tag{23}$$

In respect to the survey response, two definitions of representativeness (as a wide concept of the response representativeness, not in the understanding of the sampling theory) were introduced (Schouten, Cobben, Bethlehem, 2009) strong and week.

Definition (strong): A response subset is representative with respect to the sample if the response propensities ρ_i are the same for all units in the population:

$$\forall i \quad \rho_i = P[r_i = 1 \mid s_i = 1] = \rho \tag{24}$$

and if the response of a unit is independent of the response of all other units.

If a missing-data mechanism satisfies the strong definition, then the mechanism will correspond to Missing-Completely-at-Random (MCAR) with respect to all survey questions. The validity of the strong definition cannot be verified in practice, so a weak definition was proposed (Schouten, Cobben, Bethlehem, 2009).

Definition (weak): A response subset is representative of a categorical variable *X* with *H* categories if the average response propensity $\overline{\rho}$ over the categories is constant:

$$\overline{\rho} = \frac{1}{N_h} \sum_{k=1}^{N_h} \rho_{hk} = \rho, \text{for } h = 1, 2, ..., H, \tag{25}$$

where:

N_h is the population size of category *h*;

ρ_{hk} is the response propensity of the unit *k* in the class h and summation is over all units in this category.

The week definition corresponds to MCAR with respect to *X*, as distinguishing respondents from nonrespondents based on knowledge of *X* is not possible. Hence, regarding a week definition, the response propensities can be estimated within corresponding strata based on *X*, so the assumption of weak representativity can be verified in practice.

Schouten, Cobben and Bethlehem (2009) introduced the R-indicator for the evaluation of a representative response as a measure based upon the variance of estimated response probabilities.

Let us consider the hypothetical situation with all individual response propensities known – a strong definition could be tested and measurement of variability in the response propensities would be easy, and the more variation, the less representativity in the context of the strong definition.

Let $\rho = (\rho_1, \rho_2, \ldots, \rho_N)'$ be a vector of response propensities, let $1 = (1, 1, \ldots, 1)'$ be the N – vector of "1", and let $\rho_0 = 1 \times \rho$ be the vector of the average population propensity:

$$\bar{\rho} = \frac{1}{N}\sum_{i=1}^{N} \rho_i \,. \tag{26}$$

Any distance function d in $[0, 1]^N$ would suffice in order to measure the deviation from the strong representative response (the strong definition) by measuring the distance $d(\rho, \rho_0)$.

The Euclidean distance can be applied to the distance $d(\rho, \rho_0)$ and the measure proportional to the standard deviation of the response probabilities is given as:

$$SD(\rho) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} (\rho_i - \bar{\rho})^2} \,. \tag{27}$$

When fixing the average response probability $\bar{\rho}$, the maximum possible variance value is obtained by letting $\bar{\rho}N$ of the response probabilities be equal to 1 and respectively $(1-\bar{\rho})N$ to value 0 (Cobben, 2009), hence:

$$SD(\rho) \le \sqrt{\bar{\rho}(1-\bar{\rho})} \,. \tag{28}$$

Moreover, for $\bar{\rho} = \frac{1}{2}$:

$$SD(\rho) \le \sqrt{\bar{\rho}(1-\bar{\rho})} \le \frac{1}{2} \,. \tag{29}$$

The R-indicator proposed by Schouten, Cobben and Bethlehem (2009) takes values in the interval [0, 1] with the value 1 being strong representativeness and the value 0 being the maximum deviation from the strong representativeness. The following indicator was defined:

$$R(\rho) = 1 - 2SD(\rho) \;\; \Rightarrow \;\; R(\rho) = 1 - 2\sqrt{\frac{1}{N-1}\sum_{i=1}^{N} (\rho_i - \bar{\rho})^2} \,. \tag{30}$$

The minimum value of (29) depends on the response rate, it has the 0 value for $\overline{\rho}=\frac{1}{2}$ and the 1 value for $\overline{\rho}=0$ or $\overline{\rho}=1$, as there is no variation observed in the response rate then. The R-indicator may be considered as a lack of association measure. From the quality perspective, it should be discussed as a measure of extent to which the survey response deviates from the representative response. R-indicators can be used to compare representativeness of different surveys, but cannot be used for identifying subgroups that are over and under represented. However, they can be supplemented by partial R-indicators corresponding to the weak definition (Schouten, Cobben, Bethlehem, 2009).

Let us denote estimated response propensity for each element *i* as $\hat{\rho}_i$.

Let $\hat{\overline{\rho}}$ be denoted as the weighted sample average of the estimated response propensities given as:

$$\hat{\overline{\rho}}=\frac{1}{N}\sum_{i=1}^{N}\hat{\rho}_i\frac{s_i}{\pi_i}, \tag{31}$$

where the inclusion weights are applied. If $\hat{\overline{\rho}}$ is introduced to the R formula given as (30), the following partial indicator can be defined:

$$\hat{R}(\rho)=1-2\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\frac{s_i}{\pi_i}(\hat{\rho}_i-\hat{\overline{\rho}})^2}. \tag{32}$$

It should be emphasised that representativity is considered here in the sense of the representative response concept (Schouten, Cobben, Bethlehem, 2009), not the statistical sampling theory. Especially for the surveys based on Internet sources, this approach might be satisfying in assessing quality by measuring if and to what extent answers from a given survey are representative in the context of the entire population. The main advantages of R-indicators are: a simple scale of measurement, the assessment of sample representativeness and the nonresponse bias, as well as the identification of subgroups for nonresponse follow-up. The main limitations are: the auxiliary data availability as well as the fact that comparisons require identical auxiliary variables and that threshold values are not identified. However, the R-indicators seem to be successfully used as a quality assessment tool in tandem with the response rate, and they can help to improve quality during data collection as well as help to compare data representativeness in different modes. Partial R-indicators can be used to determine which subgroup(s) are contributing the most to a lack of sample representativeness, which can significantly support the adaptive survey approach.

References

Beręsewicz M. (2015), *On the representativeness of Internet data sources for the real estate market in Poland*, "Austrian Journal of Statistics", vol. 4(2), pp. 45–57.

Beręsewicz M. (2017), *A Two-Step Procedure to Measure Representativeness of Internet Data Sources*, "International Statistical Review", vol. 85(3), pp. 473–493.

Beręsewicz M., Szymkowiak M. (2015), *Big data w statystyce publicznej – nadzieje, osiągnięcia, wyzwania i zagrożenia*, "Ekonometria", vol. 2(48), pp. 9–21.

Bethlehem J. (2009), *The Rise of Survey Sampling*, Discussion Paper 09015, Statistics Netherlands, The Hague/Heerlen.

Bethlehem J. (2010), *Selection bias in web surveys*, "International Statistical Review", vol. 78(2), pp. 161–188.

Bethlehem J. (2012), *Using response probabilities for assessing representativity*, Discussion paper 201212, Statistics Netherlands, Voorburg/Heerleen.

Bethlehem J., Biffignandi S. (2011), *Handbook of Web Surveys*, Wiley, New York.

Callegaro M., Manfreda K., Vehovar V. (2015), *Web survey methodology*, Sage, London.

Cobben F. (2009), *Nonresponse in sample surveys: methods for analysis and adjustment*, Statistics Netherlands, The Hague/Heerlen.

Cobben F., Schouten B. (2005), *Bias measures for evaluating and facilitating flexible fieldwork strategies*, Paper presented at 16th International Workshop on Household Survey Nonresponse, August 28–31, Tällberg.

Cobben F., Schouten B. (2007), *An empirical validation of R-indicators*, Discussion paper, Statistics Netherlands, Voorburg.

Cochran W.G. (1977), *Sampling Techniques*, John Wiley & Sons, New York.

Eurostat (2018), *Households – level of internet access*, http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_ci_in_h&lang=en (accessed: 20.12.2018).

Fielding N.G., Lee R.M., Blank G. (2017), *The SAGE Handbook of Online Research Methods*, Sage, London.

Fricker R., Schonlau M. (2002), *Advantages and disadvantages of internet research surveys: Evidence from the literature*, "Field Methods", no. 15, pp. 347–367.

Gemius S.A. (2018), *Wyniki badania Gemius/PBI za listopad 2018*, https://www.gemius.pl/reklamodawcy-aktualnosci/wyniki-badania-gemiuspbi-za-listopad–2018.html (accessed: 28.12.2018).

Groves R.M., Peytcheva E. (2008), *The Impact of Nonresponse Rates on Nonresponse Bias*, "Public Opinion Quarterly", no. 72, pp. 1–23.

Kalton G. (2019), *Developments in Survey Research over the Past 60 Years: A Personal Perspective*, "International Statistical Review", no. 87(S1), pp. S10–S30.

Leeuw E.D. de (2018), *Mixed-Mode: Past, Present, and Future*, "Survey Research Methods", vol. 12(2), pp. 75–89.

Polish Society of Market and Opinion Researchers (2011), *Yearbook of Polish Society of Market and Opinion Researchers 2011/2012*, Polskie Towarzystwo Badaczy Rynku i Opinii, Warszawa.

Polish Society of Market and Opinion Researchers (2012), *Yearbook of Polish Society of Market and Opinion Researchers 2012/2013*, Polskie Towarzystwo Badaczy Rynku i Opinii, Warszawa.

Polish Society of Market and Opinion Researchers (2013), *Yearbook of Polish Society of Market and Opinion Researchers 2013/2014*, Polskie Towarzystwo Badaczy Rynku i Opinii, Warszawa.

Polish Society of Market and Opinion Researchers (2014), *Yearbook of Polish Society of Market and Opinion Researchers 2014/2015*, Polskie Towarzystwo Badaczy Rynku i Opinii, Warszawa.

Polish Society of Market and Opinion Researchers (2015), *Yearbook of Polish Society of Market and Opinion Researchers 2015/2016*, Polskie Towarzystwo Badaczy Rynku i Opinii, Warszawa.

Polish Society of Market and Opinion Researchers (2016), *Yearbook of Polish Society of Market and Opinion Researchers 2016/2017*, Polskie Towarzystwo Badaczy Rynku i Opinii, Warszawa.

Polish Society of Market and Opinion Researchers (2017), *Yearbook of Polish Society of Market and Opinion Researchers 2017/2018*, ORA & Funksters, Warszawa.

Polish Society of Market and Opinion Researchers (2018), *Yearbook of Polish Society of Market and Opinion Researchers 2018/2019*, Polskie Towarzystwo Badaczy Rynku i Opinii, Warszawa.

Schonlau M., Couper M.P. (2017), *Options for Conducting Web Surveys*, "Statistical Science", vol. 32(2), pp. 279–292.

Schouten B., Cobben F., Bethlehem J.G. (2009), *Indicators for the Representativeness of Survey Response*, „Survey Methodology", vol. 35(1), pp. 101–113.

Schouten B., Bethlehem J., Beullens K., Kleven O., Loosveldt G., Luiten A., Rutar K., Shlomo N. (2012), *Evaluating, Comparing, Monitoring and Improving Representativeness of Survey Response Through R-indicators and Partial R-indicators*, "International Statistical Review", vol. 80(3), pp. 382–399.

Shlomo N., Skinner C.J., Schouten B., Bethlehem J., Zhang L. (2008), *Statistical Properties of R-indicators*, Work Package 3, Deliverable 3.1, RISQ Project, 7th Framework Programme (FP7) of the European Union.

Skinner C., Shlomo N., Schouten B., Zhang L., Bethlehem J. (2009), *Measuring Survey Quality Through Representativeness Indicators Using Sample and Population Based Information*, NTTS Conference, 18–20 February 2009, Brussels.

Statistics Poland (2018a), *Information society in Poland. Results of statistical surveys in the years 2013–2017*, https://stat.gov.pl/en/topics/science-and-technology/information-society/information-society-in-poland-results-of-statistical-surveys-in-the-years–20132017,1,4.html (accessed: 20.12.2018).

Statistics Poland (2018b), *Information society in Poland. Results of statistical surveys in the years 2014–2018*, https://stat.gov.pl/en/topics/science-and-technology/information-society/information-society-in-poland-results-of-statistical-surveys-in-the-years–20142018,1,5.html (accessed: 20.12.2018).

Statistics Poland (2018c), *Local Data Bank*, https://bdl.stat.gov.pl/BDL/start (accessed: 17.12.2018).

Szreder M. (2015), *Big data wyzwaniem dla człowieka i statystyki*, "Wiadomości Statystyczne", vol. 8(651), pp. 1–11.

Szreder M. (2017), *Nowe źródła informacji i ich wykorzystywanie w podejmowaniu decyzji*, "Wiadomości Statystyczne", vol. 7(674), pp. 5–17.

Tourangeau R., Conrad F., Couper M. (2013), *The science of web surveys*, Oxford University Press, New York.

Wybrane problemy oceny jakości w badaniach internetowych – perspektywa statystyczna

Streszczenie: W artykule przedstawiono – z perspektywy statystycznej – wybrane problemy związane z oceną jakości badań opartych na źródłach internetowych.

Dostęp do internetu konsekwentnie poszerza się na całym świecie. Równolegle, wraz z rozwojem innych nowych technologii, przestrzeń internetowa przenika coraz bardziej codzienne życie społeczeństwa, a także funkcjonowanie firm. Wszechobecny internet wywarł także wpływ na badania rynku i opinii: jako narzędzie badawcze do zbierania danych pierwotnych i wtórnych oraz w kontekście badania populacji internetowej. Ponadto, ponieważ internet i jego podmioty rejestrują wszystkie działania podejmowane w sieci, pojawiła się kwestia związana z wykorzystaniem i analizą big data i danych organicznych. W połączeniu z problemem malejących stóp odpowiedzi w badaniach i z rosnącymi ich kosztami źródła internetowe, ze względu na wiele zalet, są w powszechnym użyciu. Coraz szersze wykorzystanie internetu i jego zasobów wydaje się nieuniknione. Należy jednak podkreślić, że w praktyce proces realizacji badań na podstawie źródeł internetowych wyprzedził prace metodologiczne. Można wskazać wiele problemów, szczególnie w kwestii jakości uzyskiwanych danych. Artykuł prezentuje wybrane z nich, istotne zwłaszcza z punktu widzenia statystyki: kwestie związane z poprawnym zdefiniowaniem operatu losowania, samodoborem, nadmiernym/niedostatecznym pokryciem i powiązanymi z nimi obciążeniami estymatorów.

Słowa kluczowe: badanie internetowe, badanie on-line, jakość badania, błędy w badaniach

JEL: C8

OPEN ACCESS	© by the author, licensee Łódź University – Łódź University Press, Łódź, Poland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license CC-BY (https://creativecommons.org/licenses/by/4.0/)
	Received: 2019-01-18; verified: 2020-02-24. Accepted: 2020-09-01
COPE Member since 2018 JM13703	This journal adheres to the COPE's Core Practices https://publicationethics.org/core-practices

Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663
www.czasopisma.uni.lodz.pl/foe/
4(349) 2020
DOI: http://dx.doi.org/10.18778/0208-6018.349.04

Wojciech Gamrot
University of Economics in Katowice, College of Management, Department of Statistics, Econometrics and Mathematics, wojciech.gamrot@ue.katowice.pl

A Stopping Rule for Simulation-Based Estimation of Inclusion Probabilities

Abstract: Design-based estimation of finite population parameters such as totals usually relies on the knowledge of inclusion probabilities characterising the sampling design. They are directly incorporated into sampling weights and estimators. However, for some useful sampling designs, these probabilities may remain unknown. In such a case, they may often be estimated in a simulation experiment which is carried out by repeatedly generating samples using the same sampling scheme and counting occurrences of individual units. By replacing unknown inclusion probabilities with such estimates, design-based population total estimates may be computed. The calculation of required sample replication numbers remains an important challenge in such an approach. In this paper, a new procedure is proposed that might lead to the reduction in computational complexity of simulations.

Keywords: Horvitz-Thompson estimator, inclusion probabilities, simulation, precision

JEL: C83, C63

1. Introduction

Following Särndal, Swensson and Wretman (1992: 5), we shall represent a finite population as a set of unit indices $U = \{1, \ldots, N\}$. Values of a fixed characteristic for corresponding population units are represented by a vector $\boldsymbol{y} = [y_1, \ldots, y_N]'$. The parameter under study is the population total (Hedayat, Sinha, 1991: 2):

$$t = \sum_{i \in U} y_i = 1'\boldsymbol{y}. \tag{1}$$

The unordered sample space may be represented by a matrix:

$$\boldsymbol{A} = \left[a_{ij}\right] = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & 1 \end{bmatrix} \tag{2}$$

whose each i-th row:

$$\boldsymbol{a}_i = \left[a_{i1}, \ldots, a_{\mathrm{iN}}\right] \tag{3}$$

represents one possible sample with $a_{ij} = 1$ when this sample contains the j-th unit and $a_{ij} = 0$ otherwise. The matrix $\boldsymbol{A}$ has N columns and $Z = 2^{\mathrm{N}}$ rows representing all possible sequences of zeros and ones of the length N, including an empty sample represented by a sequence of N zeros and a *census* represented by N ones. A vector of corresponding sample sizes may be calculated as:

$$n = \left[n_{1,} \ldots, n_Z\right]' = \boldsymbol{A}1. \tag{4}$$

Let an unordered sample: $s \subseteq U$ be drawn from U. The sample composition may be characterised by a vector of sample membership indicators (Tillé, 2006: 8):

$$\boldsymbol{I}(s) = \left[I_1(s), \ldots, I_N(s)\right], \tag{5}$$

where

$$I_i(s) = \begin{cases} 0 & \text{for} \quad i \in s \\ 1 & \text{for} \quad i \notin s \end{cases} \tag{6}$$

The sampling is equivalent to choosing a certain (say i-th) row of $\boldsymbol{A}$ so that $\boldsymbol{I}(s) = \boldsymbol{a}_i$. It may be done according to a sampling design:

$$\boldsymbol{P} = \left[P_1, \ldots, P_Z\right]', \tag{7}$$

which associates a selection probability $P_i \in [0,1]$ with every row $\boldsymbol{a}_i$ for $i \in U$, so that $\sum P_i = 1$. The expectation of the vector $\boldsymbol{I}$ may be expressed as:

$$\boldsymbol{\pi} = [\pi_1, ..., \pi_N]' = \boldsymbol{P}'\boldsymbol{A}. \tag{8}$$

Elements of the vector π are called *first-order inclusion probabilities* because $\pi_i = Pr\{i \in s\}$ for $i \in U$. Let us also introduce a vector of corresponding weights:

$$\boldsymbol{d} = \left[\pi_1^{-1}, ..., \pi_N^{-1}\right]'. \tag{9}$$

One may also define a matrix of second-order inclusion probabilities (Tillé, 2006: 17) as:

$$\Pi = \begin{bmatrix} \pi_{11} & \cdots & \pi_{1N} \\ \vdots & \ddots & \vdots \\ \pi_{N1} & \cdots & \pi_{NN} \end{bmatrix} = \boldsymbol{A}' diag(\boldsymbol{P}) \boldsymbol{A}, \tag{10}$$

where $\pi_{ij} = Pr\{i, j \in s\}$. This lets us express the covariance matrix of the vector $\boldsymbol{I}$ as:

$$\boldsymbol{C} = \Pi - \pi\pi'. \tag{11}$$

The size of the sample s may be expressed as:

$$\boldsymbol{n}(s) = 1'\boldsymbol{I}(s). \tag{12}$$

Denote sampled elements as:

$$s = \{i_{1,} ..., i_{n(s)}\}. \tag{13}$$

For any vector $\boldsymbol{u} = [u_1, \dots, u_N]$, let:

$$\boldsymbol{u}(s) = \left[u_{i_1}, ..., u_{i_{n(s)}}\right]. \tag{14}$$

This lets us define sample vectors: $\boldsymbol{y}(s)$, $\boldsymbol{\pi}(s)$, $\boldsymbol{d}(s)$ which are obtained by omitting elements corresponding to zeros in $\boldsymbol{I}(s)$ respectively in $\boldsymbol{y}$, $\boldsymbol{\pi}$, $\boldsymbol{d}$. For known $\boldsymbol{\pi}$, the design-unbiased Horvitz-Thompson (HT) estimator of t may be expressed in the form (cf. Narain, 1951; Horvitz, Thompson, 1952):

$$\hat{t}(s) = \boldsymbol{d}' diag(\boldsymbol{I}(s)) \boldsymbol{y} \tag{15}$$

or equivalently:

$$\hat{t}(s) = \boldsymbol{d}'(s) \boldsymbol{y}(s). \tag{16}$$

2. Simulation-based estimation

To calculate the HT estimator, first order inclusion probabilities are needed. However, many sampling procedures are too complicated to calculate them. In particular, this is true for spatial sampling (Barabesi, Fattorini, Ridolfi, 1997; Fattorini, Ridolfi, 1997), order sampling schemes, especially the Pareto scheme (Rosén, 1997), rejective sampling (Wywiał, 2003; Boistard, Lopuhaä, Ruiz-Gazen, 2012; Yu, 2012), and sequential sum-quota sampling schemes (Pathak, 1976; Kremers, 1985). A particular example is the greedy sampling scheme (Gamrot, 2014: 223) where costs of sampling individual units vary but are known in advance, and the survey budget is restricted. Individual units are drawn to the sample sequentially, one-by-one, with equal probabilities, from a gradually shrinking pool of still-affordable units. In the most pessimistic case, the calculation of inclusion probabilities would require analysing all permutations of units, which is unfeasible. If inclusion probabilities do not depend on sample observations, then Fattorini (2006; 2009) proposes to perform a simulation experiment. It is carried out by generating a large number R of sample replications $\breve{s}_1, \ldots, \breve{s}_R$. Empirical counts of unit occurrences are then calculated as:

$$\boldsymbol{m} = [m_1, \ldots, m_N]' = \sum_{i \in \{1 \ldots R\}} \boldsymbol{I}(\breve{s}_i). \tag{17}$$

This enables the calculation of empirical inclusion probabilities:

$$\hat{\pi} = \left[\hat{\pi}_1, \ldots, \hat{\pi}_N\right]' = \frac{m}{R} \tag{18}$$

and empirical weights:

$$\hat{\boldsymbol{d}} = \left[\hat{\pi}_1^{-1}, \ldots, \hat{\pi}_N^{-1}\right]'. \tag{19}$$

By omitting elements corresponding to non-sampled units respectively in $m, \hat{\pi}, \hat{d}$ one may then obtain empirical quantities $m(s), \hat{\pi}(s), \hat{d}$ (s) associated with the realised sample s. This leads to the calculation of the empirical HT estimator in the form:

$$\hat{\hat{t}}(s) = \hat{\boldsymbol{d}}' diag(\boldsymbol{I}(s))\, \boldsymbol{y} \tag{20}$$

or equivalently:

$$\hat{\hat{t}}(s) = \hat{\boldsymbol{d}}'(s)\, \boldsymbol{y}(s). \tag{21}$$

3. Setting up the stopping rule

In order to establish a sufficient value of replication number R that guarantees a required precision of simulation-based estimates, Fattorini (2006; 2009) proposes the accuracy criterion:

$$Q(R) = Pr\left\{\left|\hat{\hat{t}}(s) - \hat{t}(s)\right| < \varepsilon \hat{t}(s)\right\} \tag{22}$$

and finds its upper bound on the basis of Bennet's inequality. On this basis, he proposes a formula for the sufficient value of R. Later, Gamrot (2013) attempted to improve over that using asymptotic approximations based on a normal distribution, Chernoff-Hoeffding inequality, and pre-calculated tables of exact probabilities for the restricted maximum likelihood estimator. However, the relative deviation of the empirical HT estimator $\hat{\hat{t}}(s)$ from its "true" value $\hat{t}(s)$ that would be calculated for known inclusion probabilities has a complex distribution. The construction of an upper bound for it requires the pessimistic assumption of possible high correlation among sample membership indicators. This leads to very conservative replication numbers, which results in long calculation time.

In what follows, it is demonstrated that these pessimistic assumptions are often overly conservative, and may be improved upon. The value of the empirical HT estimator depends on the count vector m. Let Ω be a set of such values of this vector for which the condition

$$\left|\frac{\hat{\hat{t}}(s) - \hat{t}(s)}{\hat{t}(s)}\right| < \varepsilon \tag{23}$$

is satisfied so that:

$$Q(R) = Pr\{\boldsymbol{m} \in \text{Ù}\}\ . \tag{24}$$

Hence, instead of examining the scalar distribution of $\hat{\hat{t}}(s)$ one may investigate a much simpler, multivariate distribution of $\boldsymbol{m}$. As an introductory example, let us consider a population of size $N = 3$, with the sample space and sampling design:

$$\boldsymbol{A}_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad \boldsymbol{P}_0 = \begin{bmatrix} p_{000} \\ p_{001} \\ p_{010} \\ p_{011} \\ p_{100} \\ p_{101} \\ p_{110} \\ p_{111} \end{bmatrix}. \tag{25}$$

When a sample s corresponding to the sample indicator vector $\boldsymbol{I}(s) = [0, 1, 1]$ is drawn, all the columns of the matrix $\boldsymbol{A}$ which contain zeros and correspond to non-sampled units may be disregarded in our analysis because the HT estimator does not depend on these units and corresponding inclusion probabilities. Hence, it is sufficient to consider a reduced sample space and sampling design:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \boldsymbol{P} = \begin{bmatrix} p_{00} \\ p_{01} \\ p_{10} \\ p_{11} \end{bmatrix} = \begin{bmatrix} p_{000} + p_{100} \\ p_{001} + p_{101} \\ p_{010} + p_{110} \\ p_{011} + p_{111} \end{bmatrix} \tag{26}$$

with the reduced sample indicator vector $\boldsymbol{I}(s) = [1, 1]$. Such reduction may be carried out for populations and samples of any size. However, in a bivariate case, the distribution of the vector $\boldsymbol{m} = [m_1, m_2]'$ for R sample replications is particularly simple and takes the form (Beyer, 1987: 532):

$$Pr\left(m_1, m_2\right) = \sum_a D\left(a\right) p_{11}^{a} p_{10}^{(m_1 - a)} p_{01}^{(m_2 - a)} p_{00}^{R - m_1 - m_2 + a}, \tag{27}$$

where:

$$D\left(a\right) = \frac{R!}{a!\left(m_1 - a\right)!\left(m_2 - a\right)!\left(R - m_1 - m_2 + a\right)!}. \tag{28}$$

In the following examples, we will now discuss in more detail some interesting special cases of such a bivariate distribution.

Example 1. Let the sample s of size 2 corresponding to the indicator vector $\boldsymbol{I}(s) = [1, 1]$ be drawn from the population according to the sampling design $\boldsymbol{P} = [0.1, 0.3, 0.5, 0.1]'$. Let us consider four possible sample outcomes: $\mathbf{y} = [1, 4]'$, $\mathbf{y} = [1, 2]'$, $\mathbf{y} = [1, 1]'$ and $\mathbf{y} = [2, 1]'$. Assume that $\varepsilon = 0.1$ and $R = 100$. Distributions of the count vector $\boldsymbol{m}$ for all four designs along with resulting values of the accuracy criterion Q(R) are shown in Figure 1. It is clearly visible that the region Ω shifts when values of the study variable change. This results in higher or lower probabilities Q(R).

The dependency of Q(R) on sample observations of the study variable are not the only important effect. The following example illustrates another one:

Example 2. Let us consider four sampling designs $\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_3, \boldsymbol{P}_4$ given by Table 1, together with corresponding values of the correlation coefficient ρ between the two sample membership indicators corresponding to the first and second unit. Let the sample s of size 2 corresponding to the sample membership indicator vector $\boldsymbol{I}(s) = [1, 1]$ be drawn from the population and let the following values of the study variable be observed: $\mathbf{y} = [1, 1]$. Assume that $\varepsilon = 0.1$ and $R = 100$. Distributions of the count vector $\boldsymbol{m}$ for all four designs along with resulting values of the accuracy criterion Q(R) are shown in Figure 2.

Table 1. Distributions of the count vector and correlation coefficients of sample membership indicators for certain sampling designs

	$\boldsymbol{P}_1$	$\boldsymbol{P}_2$	$\boldsymbol{P}_3$	$\boldsymbol{P}_4$
p_{00}	0.49	0.35	0.25	0.10
p_{01}	0.01	0.15	0.25	0.40
p_{10}	0.01	0.15	0.25	0.40
p_{11}	0.49	0.35	0.25	0.10
ρ	0.96	0.40	0.00	–0.6

Source: own elaboration

Despite its simplicity, the example shows that the probability Q(R) depends on the correlation ρ between sample membership indicators. It is high for the least favourable case extreme of positive correlation that is tacitly assumed in the derivation of known stopping rules. In reality, however, it is often much lower. Taking this effect into account, one might construct tighter bounds for Q(R) and obtain a stopping rule which gives lower required replication numbers.

This effect remains in force when more than two elements are drawn to the sample, as shown in the last example.

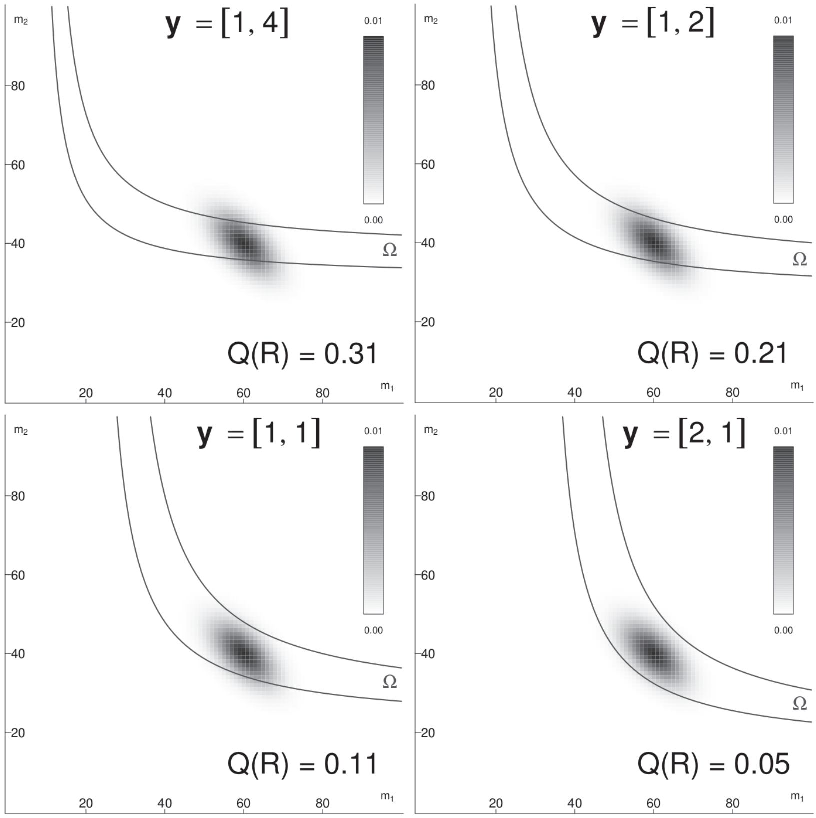


Figure 1. Location of the Ω region for various sample observations of the study variable

Source: own elaboration

Example 3. Consider the population of size $N = 3$ and the general sampling design $\boldsymbol{P} = [b, a, a, a, a, a, a, b]'$, where $b = (1 - 6a)/2$, that is parameterised by the constant $a \in (0,1/6)$. The special feature of this design is that the correlation coefficient between all sample membership indicators takes the same value (say ρ). Probabilities of drawing all the possible samples obtained for varying values of a and resulting ρ-values are shown in Table 2. Let us assume that all the three population elements are drawn to a sample resulting in the sample observation of the study variable: $\mathbf{y} = [1, 2, 3]$. Let us also assume that $\varepsilon = 0.15$ and $R = 50$. The empirical distribution of 5000 realisations of the count vector $\boldsymbol{m}$, along with boundaries of the Ω region for $a = 0.01, 0.05, 0.125, 0.166$, are shown in Figure 3. It is clearly visible that the negative correlation among sample membership indicators – and even lack thereof – improves the accuracy criterion as compared with the worst case of perfect positive correlation.

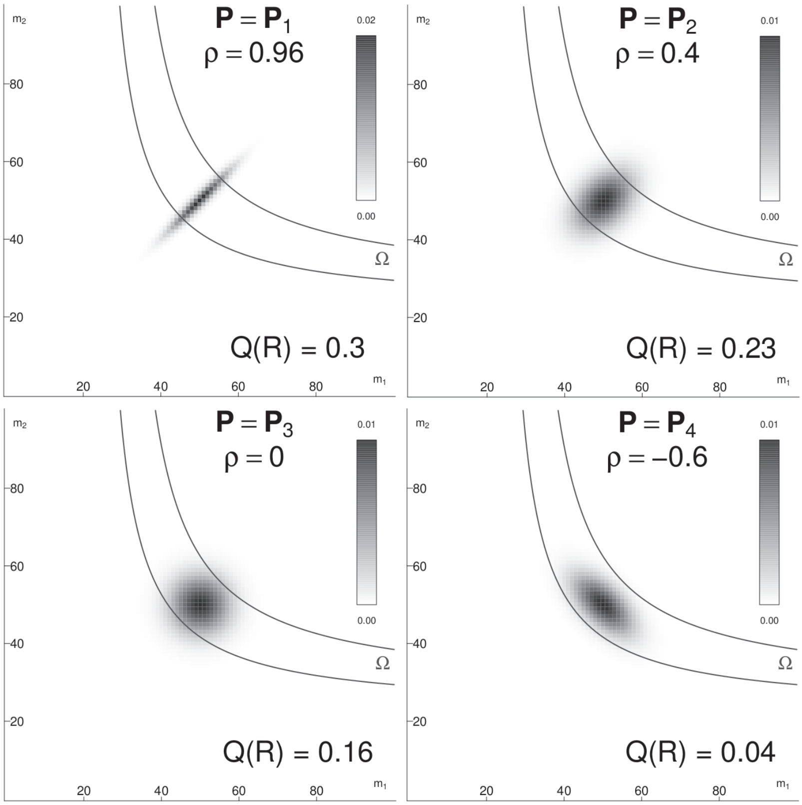


Figure 2. Distribution of the simulation count vector ***m*** for various sampling designs and $n(s) = 2$

Source: own elaboration

Table 2. Sample drawing probabilities and correlation coefficients between sample membership indicators for various values of the parameter *a*

a	0.01	0.05	0.125	0.166
p_{000}	0.47	0.35	0.125	0.002
p_{001}	0.01	0.05	0.125	0.166
p_{010}	0.01	0.05	0.125	0.166
p_{011}	0.01	0.05	0.125	0.166
p_{100}	0.01	0.05	0.125	0.166
p_{101}	0.01	0.05	0.125	0.166
p_{110}	0.01	0.05	0.125	0.166
p_{111}	0.47	0.35	0.125	0.002
ρ	0.92	0.6	0.00	–0.328

Source: own elaboration

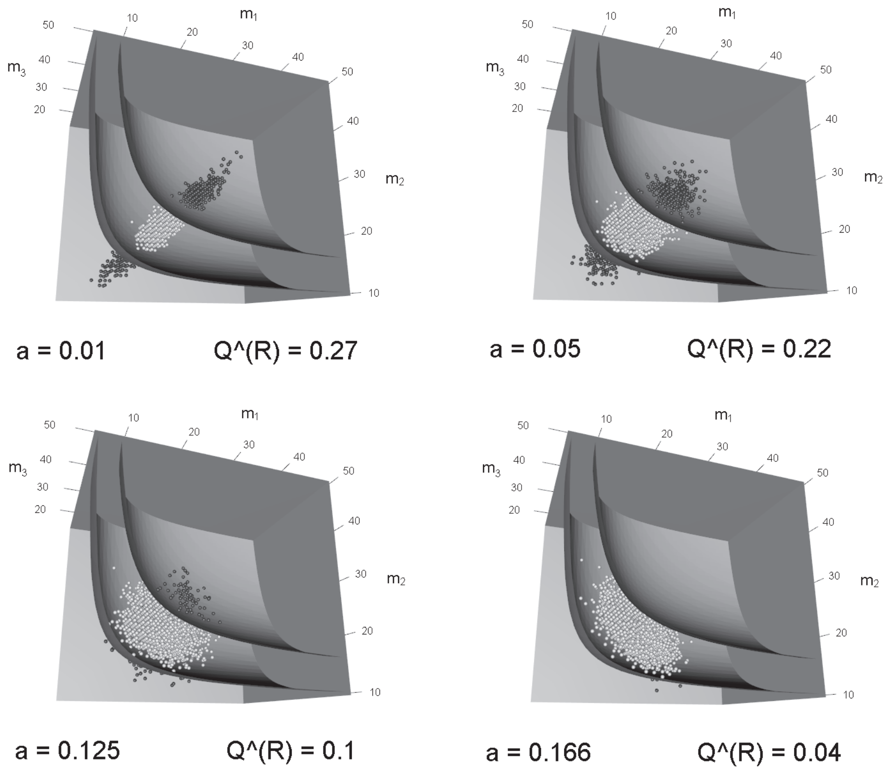


Figure 3. Distribution of the simulation count vector ***m*** for various sampling designs and $n(s) = 3$
Source: own elaboration

4. The proposed stopping rule

The accuracy criterion depends on correlations among sample membership indicators. However, it is not reasonable to expect these correlations to be known when inclusion probabilities (defined as moments of their distributions) remain unknown. To account for correlations, the simulation may be divided into two phases. In the first phase, R_1 sample replications $\breve{s}_1, \ldots, \breve{s}_{R_1}$ are generated. Occurrences of individual units and pairs of units are counted resulting in the following counts:

$$\boldsymbol{m}_1 = \sum\nolimits_{i\in\{1\ldots R_1\}} \boldsymbol{I}\left(\breve{s}_i\right). \tag{29}$$

$$\boldsymbol{M}_1 = \begin{bmatrix} m_{11} & \cdots & m_{1N} \\ \vdots & \ddots & \vdots \\ m_{N1} & \cdots & m_{NN} \end{bmatrix}' = \sum\nolimits_{i \in \{1 \ldots R_1\}} \boldsymbol{I}\left(\breve{s}_i\right) \boldsymbol{I}'\left(\breve{s}_i\right). \tag{30}$$

Then the estimates of first and second order inclusion probabilities are obtained as:

$$\hat{\pi}_1 = \frac{\boldsymbol{m}_1}{R_1}, \tag{31}$$

$$\hat{\Pi}_1 = \frac{\boldsymbol{M}_1}{R_1} \tag{32}$$

This enables estimation of the covariance matrix $\boldsymbol{C}$ which characterises a joint distribution of sample membership indicators as:

$$\hat{\boldsymbol{C}} = \hat{\Pi}_1 - \hat{\pi}_1 \hat{\pi}_1' \tag{33}$$

while $\hat{\pi}_1$ is obviously an estimate of its expectation vector $\boldsymbol{\pi}$. In the second phase of simulation, R_2 sample replications $\breve{s}_{R_1+1}, \ldots, \breve{s}_{R_1+R_2}$ are generated. Hence, in the whole simulation, a total of $R = R_1 + R_2$ sample replications $\breve{s}_1, \ldots, \breve{s}_R$ are generated, which leads to the calculation of the final count vector $\boldsymbol{m}$ and the empirical HT estimator according to expressions (17)–(21). Capabilities of contemporary computers make it possible to set R_1, R_2 and R quite large (in the order of millions) without much effort so that the distribution of $\boldsymbol{m}$ tends to multivariate normal: $N\left(R\pi, R^2\boldsymbol{C}\right)$ as shown by Krzyśko (2000: 31). After the first phase of the simulation, it may be approximated by $N\left(R\hat{\pi}, R^2\hat{\boldsymbol{C}}\right)$. Realisations of this distribution may be easily and quickly generated in large quantities, for example, by using algorithms described by Zieliński and Wieczorkowski (1997). This enables the estimation of the probability Q(R) associated with *any* value of R by counting what percentage of these pseudo-random realisations falls outside the Ω region (with the unknown 'true' statistic $\hat{t}\left(s\right)$ approximated by $\hat{\hat{t}}\left(s\right)$ based on R_1 replications). Such estimation is easily repeated for various candidate values of R because generated replications of multivariate normal distribution may be reused. The re-calculation boils down to a few matrix operations (addition, multiplication, division of corresponding elements) which are easily serialised and optimised. The well-known golden-section or Newton-Raphson algorithms may hence be applied to find the minimum sufficient number R before the second phase of simulation is initiated.

5. Conclusions

According to the approach sketched in the last section, one relatively simple, but very time-consuming simulation, is replaced with a more complex but potentially faster procedure. Instead of calculating a conservative number of replications and then generating them all, a more subtle approach is proposed. After initially generating some R_1 sample replications, the auxiliary nested but fast simulation is executed to establish the required total number R of replications accounting for correlations among sample membership indicators. Then the second, possibly quite a small batch of $R - R_1$ replications, is generated, and the empirical HT estimator may finally be evaluated. The nested fast simulation step may eventually be repeated more times when more and more replications are available to make the initial assessment of R more reliable. It may also be done after all replications are generated to verify that their number is indeed sufficient. Further studies are needed to confirm whether the proposed procedure indeed produces substantial speeding-up of the whole simulation process.

Acknowledgements

This paper was presented during the MSA 2019 conference which financed its publication. Organisation of the international conference "Multivariate Statistical Analysis 2019" (MSA 2019) was supported from resources for popularisation of scientific activities of the Minister of Science and Higher Education in the framework of agreement No 712/P-DUN/202019.

References

Barabesi L., Fattorini L., Ridolfi G. (1997), *Two-Phase Surveys of Elusive Populations*, [in:] *Proceedings of the Statistics Canada Symposium 97: New Directions of Surveys and Censuses*, Statistics Canada, Ottawa, pp. 285–287.

Beyer W.H. (1987), *CRC Standard Mathematical Tables*, CRC Press, Boca Raton.

Boistard H., Lopuhaä H.P., Ruiz-Gazen A. (2012), *Approximation of rejective sampling inclusion probabilities and application to higher order correlations*, "Electronic Journal of Statistics", no. 6, pp. 1967–1983.

Fattorini L. (2006), *Applying the Horvitz-Thompson Criterion in Complex Designs: A Computer-Intensive Perspective For Estimating Inclusion Probabilities*, "Biometrica", vol. 93(2), pp. 269–278.

Fattorini L. (2009), *An Adaptive Algorithm for Estimating Inclusion Probabilities and Performing the Horvitz-Thompson Criterion in Complex Designs*, "Computational Statistics", vol. 24(4), pp. 623–639.

Fattorini L., Ridolfi G. (1997), *A Sampling Design for Areal Units Based on Spatial Variability*, "Metron", no. 55, pp. 59–72.

Gamrot W. (2013), *On exact computation of minimum sample sizes for restricted estimation of a binomial parameter*, "Journal of Statistical Planning and Inference", no. 143, pp. 852–866.

Hedayat A.S., Sinha B.K. (1991), *Design and Inference in Finite Population Sampling*, Wiley, New York.

Horvitz D.G., Thompson D.J. (1952), *A Generalization of Sampling Without Replacement from a Finite Universe*, "Journal of the American Statistical Association", no. 47, pp. 663–685.

Kremers W.K. (1985), *The Statistical Analysis of Sum-Quota Sampling*, unpublished PHD thesis, Cornell University, Ithaca.

Krzyśko M. (2000), *Wielowymiarowa analiza statystyczna*, Uniwersytet Adama Mickiewicza, Poznań.

Narain R.D. (1951), *On sampling without replacement with varying probabilities*, "Journal of the Indian Society of Agricultural Statistics", no. 3, pp. 169–175.

Pathak K. (1976), *Unbiased estimation in Fixed-Cost Sequential Sampling Schemes*, "Annals of Statistics", vol. 4(5), pp. 1012–1017.

Rosén B. (1997), *On Sampling with Probability Proportional to Size*, "Journal of Statistical Planning and Inference", no. 62, pp. 159–191.

Särndal C.E., Swensson B., Wretman J.H. (1992), *Model Assisted Survey Sampling*, Springer-Verlag, New York.

Tillé Y. (2006), *Sampling Algorithms*, Springer-Verlag, New York.

Wywiał J. (2003), *On conditional sampling strategies*, "Statistical Papers", no. 44, pp. 397–419.

Yu Y. (2012), *On the inclusion probabilities in some unequal probability sampling plans without replacement*, "Bernoulli", vol. 18(1), pp. 279–289.

Zieliński W., Wieczorkowski R. (1997), *Komputerowe generatory liczb losowych*, WNT, Warszawa.

Reguła stopu dla estymacji prawdopodobieństw inkluzji na drodze symulacyjnej

Streszczenie: Estymacja parametrów populacji skończonych i ustalonych, prowadzona w ramach podejścia randomizacyjnego, zazwyczaj wymaga znajomości prawdopodobieństw inkluzji charakteryzujących schemat losowania próby. Są one bezpośrednio wykorzystywane w celu wyznaczenia wag przypisanych poszczególnym wylosowanym jednostkom i uwzględniane podczas obliczania estymatorów. Jednak dla pewnych użytecznych schematów losowania pozostają nieznane. W takim wypadku możliwe jest ich wyznaczenie na drodze symulacyjnej, poprzez wielokrotne losowanie prób z wykorzystaniem tego samego schematu losowania i zliczanie wystąpień poszczególnych jednostek populacji. Zastępując nieznane prawdopodobieństwa inkluzji oszacowaniami uzyskanymi w wyniku takiego eksperymentu, otrzymuje się oszacowania wartości globalnej badanej cechy populacji. Szczególnym wyzwaniem podczas takiego postępowania jest wyznaczenie liczby replikacji próby, zapewniającej wymaganą precyzję estymacji. W niniejszym artykule proponowana jest nowa procedura, która może przyczynić się do zmniejszenia złożoności obliczeniowej eksperymentu symulacyjnego.

Słowa kluczowe: estymator Horvitza-Thompsona, prawdopodobieństwa inkluzji, symulacja, precyzja

JEL: C83, C63

© by the author, licensee Łódź University – Łódź University Press, Łódź, Poland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license CC-BY (https://creativecommons.org/licenses/by/4.0/)

Received: 2020-04-27; verified: 2020-06-12. Accepted: 2020-09-17

COPE Member since 2018 JM13703

This journal adheres to the COPE's Core Practices
https://publicationethics.org/core-practices

Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663

www.czasopisma.uni.lodz.pl/foe/
4(349) 2020

DOI: http://dx.doi.org/10.18778/0208-6018.349.05

Dominik Kubacki
University of Łódź, Faculty of Economics and Sociology, Department of Banking
dominik.kubacki@uni.lodz.pl

Robert Kubacki
robertkubacki@o2.pl

Examining Selected Theoretical Distributions of Life Expectancy to Analyse Customer Loyalty Durability The Case of a European Retail Bank

Abstract: One of the key elements related to calculating Customer Lifetime Value is to estimate the duration of a client's relationship with a bank in the future. This can be done using survival analysis. The aim of the article is to examine which of the known distributions used in survival analysis (Weibull, Exponential, Gamma, Log-normal) best describes the churn phenomenon of a bank's clients. If the aim is to estimate the distribution according to which certain units (bank customers) survive and the factors that cause this are not so important, then parametric models can be used. Estimation of survival function parameters is faster than estimating a full Cox model with a properly selected set of explanatory variables. The authors used censored data from a retail bank for the study. The article also draws attention to the most common problems related to preparing data for survival analysis.

Keywords: survival analysis, customer lifetime value, banking, parametric models, Kaplan–Meier estimator

JEL: C34, M31, G21

1. Introduction

Nowadays, there is an increasing need to measure the effectiveness of marketing activities. One of the indicators that synthetically describes a client's value for a company is CLTV (Customer Lifetime Value). This ratio, apart from the revenues and costs incurred so far, also includes future cash flows. It differs from NPV (Net Present Value) in that it also takes into account the probability of customers who will leave (Jeffery, 2010: 167).

$$CLTV = -AC + \sum_{n=1}^{N} \frac{\left(M_n - C_n\right) p^n}{\left(1+r\right)^n}, \tag{1}$$

where:
AC – cost of customer acquisition,
M_n – margin achieved on transactions with clients in period n,
C_n – the cost of marketing and customer service activities in period n,
p – the probability that the client will not cease cooperation within the next year[1],
N – total number of years or other periods.

Estimating the survival probability of the client population is crucial in calculating a client's value over time. Survival analysis methods can be used for this purpose.

Survival time data measure the time to a certain event, such as failure, death, response, relapse, parole, divorce, or the development of a disease. These times are subject to random variations, and like any random variables, they form a distribution (Balicki, 2006: 17).

Let T denote the survival time. The distribution of T can be characterised by three equivalent functions: the survival function, the cumulative survival function, and the cumulative hazard function. The survival function, denoted by $S(t)$, is defined as the probability that an individual will survive longer than t:

$$S(t) = P(T > t), 0 < t < \infty. \tag{2}$$

Here, $S(t)$ is a nonincreasing function of time t. The probability of surviving at time zero is 1, while the probability of surviving up to infinity is 0. The cumulative distribution function $F(t)$ is defined as the probability that an individual will fail before t:

$$F(t) = P(T \le t), 0 < t < \infty. \tag{3}$$

1 Probability may be proportional to the duration of the relationship with the bank or it may vary depending on the client's seniority.

The hazard function (t) of survival time T gives the conditional failure rate. This is defined as the probability of failure during a very small time interval, assuming that the individual has survived to the beginning of the interval, or as the limit of the probability that an individual will fail within a very short interval, $t + \Delta t$, given that the individual has survived till time T:

$$h(t) = \lim_{\Delta t \to 0}[\frac{P(t \le T < (t + \Delta t) / T \ge t}{\Delta t}] = \frac{f(t)}{S(t)}. \quad (4)$$

The cumulative hazard function is defined as:

$$H(t) = -\log(S(t)) = \int_0^t h(u)\,du. \quad (5)$$

Given any one of them, the other two can be derived:

$$S(t) = 1 - F(t) = \exp(-H(t)). \quad (6)$$

A parametric survival model is one in which survival time, thus the outcome, is assumed to follow a known distribution. By reviewing the literature about modelling survival data, it can be seen that the Exponential, Gamma, Log-normal, and Weibull probability distribution functions are commonly used in survival analysis. The $f(t)$ probability density function, $S(t)$ survival function, and mean lifetime, denoted by the $E(t)$ form of these distribution models, can be summarised below (Erişoğlu, Erişoğlu, Erol, 2011: 545):

Exponential Distribution:

$$f_{exp}(t) = \frac{1}{\lambda} e^{-\frac{t}{\lambda}}, \text{ t>0, λ>0,} \quad (7)$$

$$S_{exp}(t) = 1 - e^{-\frac{t}{\lambda}}, \quad (8)$$

$$E_{exp}(t) = \lambda. \quad (9)$$

The exponential model is a parametric model. It assumes that the baseline hazard is constant over time. The probability of surviving another time unit does not depend on how long the object has lived so far.

Gamma Distribution:

$$f_{gm}(t) = t^{\alpha_1 - 1} \frac{e^{-t/\beta_1}}{\beta_1^{\alpha_1 \Gamma(\alpha_1)}},\ t \text{ and } \alpha_1, \beta_1 > 0, \tag{10}$$

$$S_{gm}(t) = 1 - \frac{\Gamma_x(\alpha_1)}{\Gamma(\alpha_1)}, \tag{11}$$

$$E_{gm}(t) = \alpha_1 \beta_1, \tag{12}$$

$$\Gamma_x(\alpha_1) = \int_0^x t^{\alpha_1 - 1} e^{-t} dt, \tag{13}$$

$$\Gamma(t) = P(T \le t), 0 < t < \infty. \tag{14}$$

Weibull Distribution:

$$f_{wbi}(t) = \frac{\beta_2}{\alpha_2} \left(\frac{t}{\alpha_2} \right)^{\beta_2 - 1} e^{-\left(\frac{t}{\alpha_2} \right)^{\beta_2}},\ t \text{ and } \alpha_2, \beta_2 > 0, \tag{15}$$

$$S_{wbi}(t) = e^{-\left(\frac{t}{\beta_2} \right)^{\alpha_2}}, \tag{16}$$

$$E_{wbi} = \beta_2 \Gamma\left(1 + \frac{1}{\alpha_2} \right). \tag{17}$$

The Weibull distribution can also be viewed as a generalisation of the exponential distribution. It reduces to the exponential distribution when the shape parameter $\beta_2 = 1$. When the shape parameter is greater than 1, the hazard function increases; otherwise, it decreases.

Log-normal Distribution:

$$f(t) = \frac{\exp(-\frac{1}{2}\left(\frac{lnt - \mu}{\sigma} \right)^2)}{t\sigma\sqrt{2\pi}}, t > 0, \mu, \sigma > 0, \tag{18}$$

$$S(t) = 1 - \phi \frac{lnt - \mu}{\sigma}, \tag{19}$$

$$E(T) = \exp\left(\mu + \frac{\sigma^2}{2}\right), \tag{20}$$

where Φ is the cumulative distribution function of the standard normal distribution function and is defined by $\phi\left(\frac{lnt - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{lnt-\mu}{\sigma}} \exp\left(-\frac{u^2}{2}\right) du$ (Balicki, 2006: 131).

In order to select the appropriate distribution of the variable that characterises the survival curve, two assessment criteria can be used for the estimated models. The first criterion is the Akaike Information Criterion (Akaike, 1974: 716–723), and the other is the logLik or Maximised Log-likelihood (Jackson, 2016: 1–33).

2. Applications in retail banking

The study was conducted on a random sample of 100,000 retail clients in a bank located in Europe. The characteristics of the dataset are as follows:
1) individual customers aged 18–75,
2) right-censored data (date of last observation: 1.09.2018),
3) without clients with a planned termination agreement,
4) returning customers are treated as a continuous relationship if the interval does not exceed 12 months,
5) with a relationship with the bank longer than one month,
6) primary owners of the product,
7) response variable – duration in months of the customer's relationship with the bank between opening the first product and closing the last product.

The calculations and graphs were made using R and R Studio software. The packages used for the calculations included survival, flexsurv, and e1071[2]. One of the most important steps associated with preparinga survival analysis is properly preparing the data.

The first challenge is to determine what is considered to be the beginning of the relationship with the customer, whether it is the date of opening the first product or the date of establishing the general customer agreement. If the client had a relationship with the bank that handles him/her from the beginning, then these two dates should be the same. If the client was migrated to the bank as a result of a merger or takeover, then the date of establishing the customer file is usually the date of the operational merger of the two banks.

2 The Comprehensive R Archive Network, https://cran.r-project.org (accessed: 23.03.2019).

The case could get even more complicated if the customer had been served by both banks. For this study, the principle was adopted that we take into account the date of opening the first product, irrespective of the bank in which the relationship was initiated. Another solution would be to prepare separate survival curves for clients coming from the home bank, migrated clients (but new ones for the bank), and shared clients.

The second problem in preparing data for a survival analysis may be the client's return to the bank and the related opening of a new product when the last product under the previous relationship was closed. An estimation for this particular survival curve can be made. In this analysis, a business assumption was made that it was an existing relationship if the gap between the closing of the last product and the opening of a new product after returning to the bank does not exceed 12 months. This assumption can be accepted if customers use products that are characterised by a short time period, and they regularly buy a product with similar parameters after repaying the products. This may apply to banks that focus both on short-term deposits and cash loans. It is necessary to simplify the modelling of the phenomenon because such gaps may result from system limitations or the duration of setting up the product, not because of a customer actually leaving the bank.

The third problem that occurs is the large skewness of the data we work on. One of the ways to deal with this is to transform the variables, which will bring the distribution of the variable being analysed to a more symmetrical distribution. One of the most commonly used transformations of variables is the logarithmic transformation. When a log transformation is performed, adding a constant solves the problem of the legitimisation of zero. In the case of survival analysis, this condition is always met (Jajuga, Walesiak, 1999: 105–112).

3. The results of the empirical analyses conducted

The results of the non-parametric estimation of the survival function using the Kaplan–Meier (Kaplan, Meier, 1958: 457–481) estimator are presented in Figure 1. The curve is relatively regular from the 25th month[3]. In the initial period, i.e., around the 10th and the 20th month, there is a gradual decline in the survival function.

Using R software, theauthors estimated parametric models for the survival banking dataset. Four distributions were compared, and the best estimates for each distribution are presented in Table 1.

3 Months – means the time for which clients have maintained relations with the bank.

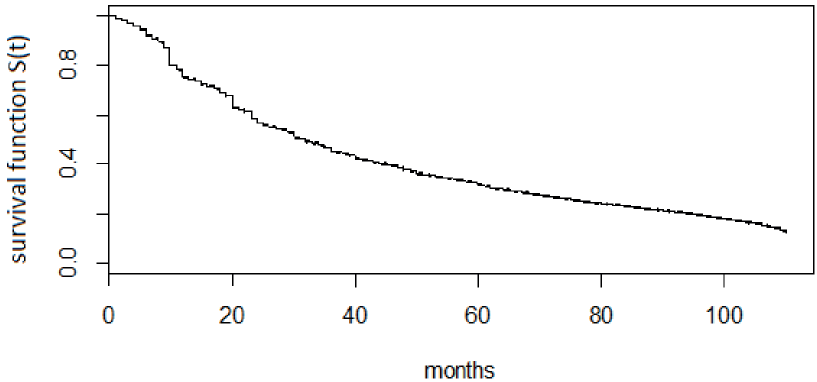


Figure 1. Survival function for months with the Kaplan–Meier estimator
Source: banking survival dataset

In Figure 2, the authors present the cumulative events for a number of months.

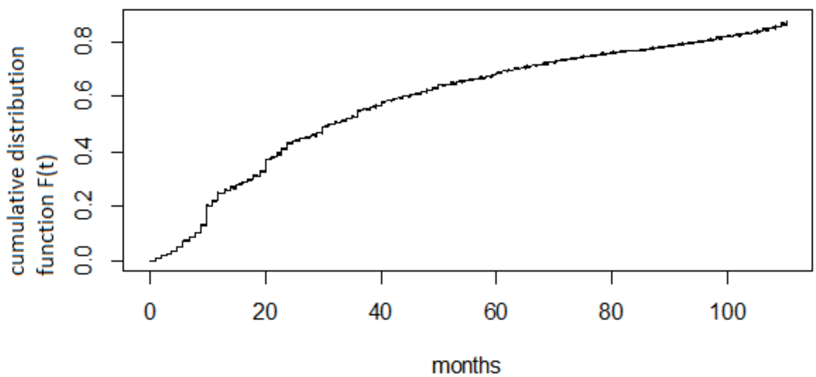


Figure 2. Cumulative events for months
Source: banking survival dataset

Table 1. Estimated parameters for the parametric survival models

Model	Parameter Estimation
Exponential	$\hat{\lambda} = 52.0833$
Gamma	$\hat{\alpha}_1 = 1.1278 \quad \hat{\beta}_1 = 44.8491$
Weibull	$\hat{\alpha}_2 = 51.9572 \quad \hat{\beta}_2 = 1.0417$
Log-normal	$\hat{\mu} = 3.488 \quad \hat{\sigma} = 1.2027$

Source: own calculation

In corresponding Table 2, the authors present the values of logLik and AIC to choose the best distribution out of the four competitors. The lowest AIC value is calculated for a log-normal distribution.

Researchers should not always focus only on the lowest AIC or logLik values. Sometimes it is better to choose a distribution with fewer parameters. This makes it easier to explain the phenomenon to business owners, as not all of them have deep statistical knowledge to interpret empirical results. Statistical significance tests can be used to check the hypothesis that the observed values do not differ from theoretical distributions.

Table 2. Values of logLik and AIC which correspond to the best-fitted distributions (variable months)

Distribution	LogLik	AIC
Exponential	–329 984	659 971
Weibull	–329 903	659 810
Gamma	–329 664	659 333
Log-normal	–326 655	653 314

Source: own calculation

Finally, in Figure 3, the authors present how a log-normal distribution fits the observed dataset. The curve fits the observed dataset. Only in regions mentioned at the beginning of the article (10th and 20th month), does the red line not fit the data.

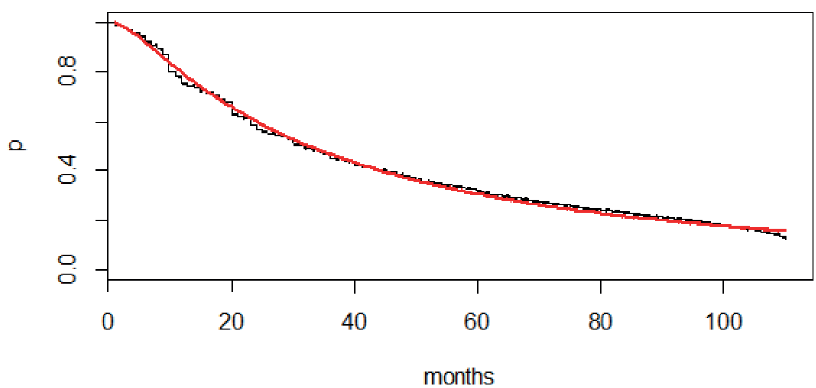


Figure 3. Log-normal survival curve for the banking dataset

Source: banking survival dataset

In Figure 4, the probability plots for the predicted and theoretical log-normal distribution are presented.

To have a good comparison between available solutions, it is sometimes worth checking other possibilities. In this study, the authors also checked how distributions for log-transformed variable months performed.

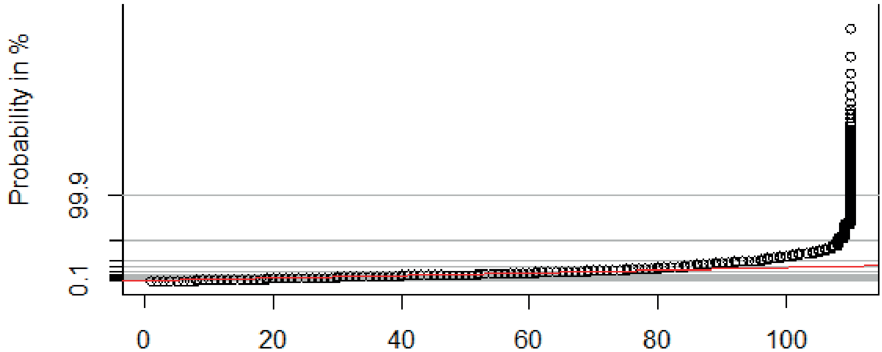


Figure 4. Probability plots for a Log-normal distribution

Source: banking survival dataset

The results and estimates obtained for these models are presented in Table 3.

Table 3. The estimated parameters for the banking survival dataset with log-transformed variable months

Model	Parameter Estimation
Exponential	$\hat{\lambda} = 4.74$
Gamma	$\hat{\alpha}_1 = 7.5347 \quad \hat{\beta}_1 = 0.4812$
Weibull	$\hat{\alpha}_2 = 3.9248 \quad \hat{\beta}_2 = 0.2865$
Log-normal	$\hat{\mu} = 1.2353 \quad \hat{\sigma} = 0.4074$

Source: own calculation

The corresponding values of logLik and AIC are presented in Table 4. The Weibull model has the lowest LogLik value. The log transformation of the data changed the winner to the best-fitted distribution.

Table 4. The values of the log-likelihood function and AIC that correspond to the best-fitted distributions (with LOG_MONTHS)

Model	LogLik	AIC
Exponential	–170 300	340 601
Weibull	–125 582	251 169
Gamma	–126 285	252 575
Log-normal	–128 409	256 822

Source: own calculation

Figure 5 presents the Weibull survival curve for the banking dataset, which fits the observed data better than the other curves using thelogLik and AIC criteria. For regions located at 2.5 and 3, there is an abrupt lowering of the survival curve.

From a business perspective, it is very interesting to investigate what type of clients end their relationship with the bank. It might be a starting point for a deeper analysis of what factors cause a customer to leave abank.

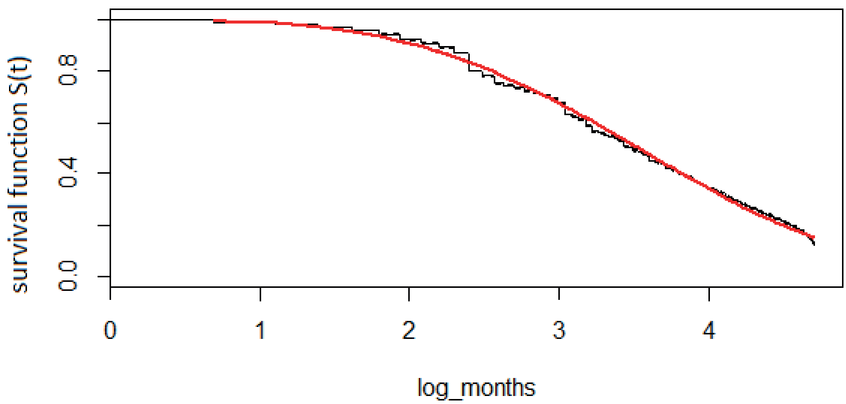


Figure 5. The Weibull survival curve for the banking dataset with a log-transformed variable
Source: banking survival dataset

Figure 6 presents the probability plots for the predicted and theoretical Weibull distribution.

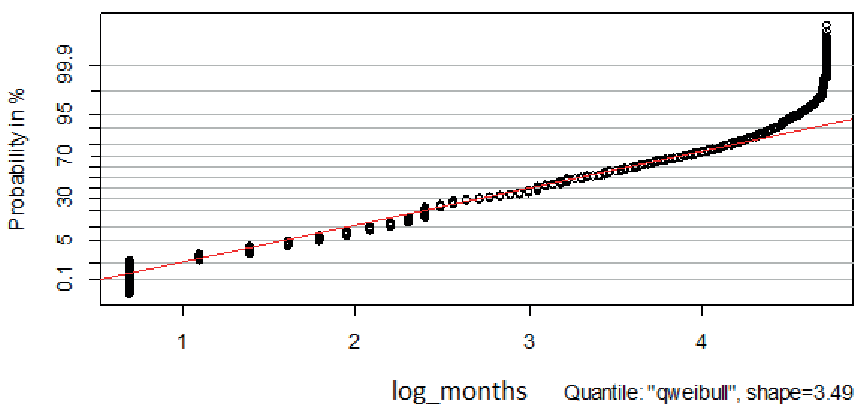


Figure 6. Probability plot for a Weibull distribution
Source: banking survival dataset

4. Conclusions

In this paper, the authors compared how observed survival data fit different theoretical distributions, such as Exponential, Weibull, Gamma, and Log-normal. The estimation of the parameters and the calculation of statistics, such as AIC and logLik, have shown that the Log-normal and Weibull distributions are best for this particular sample of clients. The results obtained in the study confirm that parametric models are valuable sources of information on the duration of customer relationships with the bank, and the model parameters themselves provide valuable knowledge of whether increased extinction occurs at the beginning of the relationship or is proportional to the examined period. The estimated parameters of survival models can be used to compare subgroups of customers that may arise from bank mergers and acquisitions. Knowing which group of customers has a steeper survival curve enables better planning of retention activities. Estimating the parameters of the survival function is simpler than building a Cox model. Gathering and preparing explanatory variables requires additional time, and not all variables that could be used in the model are available in corporate databases.

However, the analyses presented in this paper are not sufficient to extend the results to the entire banking sector. Further research is needed in this field. It would be advisable to prepare and check mixed models (the sum of two or three distributions), especially in those areas where the observed data do not perfectly fit theoretical distributions.

References

Akaike H. (1974), *A New Look at the Statistical Model Identification*, "IEEE. Transactions on Automatic Control", vol. Ac–19, no. 6, pp. 716–723.

Balicki A. (2006), *Analiza przeżycia i tablice wymieralności*, Polskie Wydawnictwo Ekonomiczne, Warszawa.

Erişoğlu Ü., Erişoğlu M., Erol H. (2011), *A Mixture Model of Two Different Distributions Approach to the Analysis of Heterogeneous Survival Data*, "World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering", vol. 5, no. 6, pp. 544–548.

Jackson C. (2016), *flexsurv: A Platform for Parametric Survival Modeling in R*, "Journal of Statistical Software", vol. 70, no. 8, pp. 1–33.

Jajuga K., Walesiak M. (1999), *Standaridisation of data set under different measurement scales*, [in]: *Classification and Information Processing at the Turn of the Millennium: Proceedings of the 23rd Annual Conference of the Gesellschaftfür Klassifikatione.V.*, University of Bielefeld, Bielefeld, pp. 105–112.

Jeffery M. (2010), *Data-Driven Marketing. The 15 Metrics Everyone in Marketing Should Know*, John Wiley & Sons, Hoboken.

Kaplan E.L., Meier P. (1958), *Nonparametric Estimation from Incomplete Observations*, "Journal of the American Statistical Association", vol. 53, no. 282, pp. 457–481.

The Comprehensive R Archive Network, https://cran.r-project.org (accessed: 23.03.2019).

Ocena wybranych rozkładów teoretycznych trwania życia do analizy lojalności klientów na przykładzie europejskiego banku detalicznego

Streszczenie: Jednym z kluczowych elementów związanych z wyliczaniem wartości klienta w czasie (*Customer Life Time Value*) jest oszacowanie długości trwania relacji klienta z bankiem w przyszłości. Można ją oszacować z wykorzystaniem metod analizy przeżycia. Celem artykułu jest sprawdzenie, który ze znanych rozkładów wykorzystywanych w analizie przeżycia (Weibulla, wykładniczy, gamma, logarytmicznie normalny) najlepiej opisuje zjawisko odejść klientów z banku. Jeśli celem jest oszacowanie rozkładu, według którego „przeżywają" określone jednostki (klienci banku), a czynniki, które to powodują, nie są aż tak istotne, to modele parametryczne mogą być wykorzystane. Oszacowanie parametrów funkcji przeżycia jest szybsze niż oszacowanie pełnego modelu Coxa z odpowiednio dobranym zestawem zmiennych objaśniających. Do badania wykorzystano dane cenzurowane banku detalicznego. W artykule zwrócono uwagę na najczęstsze problemy związane z przygotowaniem danych do analizy przeżycia.

Słowa kluczowe: analiza przeżycia, wartość życiowa klienta, bankowość, modele parametryczne, estymator Kaplana–Meiera

JEL: C34, M31, G21

OPEN ACCESS

© by the author, licensee Łódź University – Łódź University Press, Łódź, Poland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license CC-BY (https://creativecommons.org/licenses/by/4.0/)

Received: 2016-01-01; verified: 2017-01-01. Accepted: 2017-11-24

COPE Member since 2018 JM13703

This journal adheres to the COPE's Core Practices https://publicationethics.org/core-practices

Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663
www.czasopisma.uni.lodz.pl/foe/
4(349) 2020
DOI: http://dx.doi.org/10.18778/0208-6018.349.06

Grzegorz Kończak
University of Economics in Katowice, College of Management, Department of Statistics Econometrics and Mathematics, grzegorz.konczak@ue.katowice.pl

A Multivariate Extension of McNemar's Test Based on Permutations

Abstract: The purpose of this publication is to propose a permutation test to detect the departure from symmetry in multidimensional contingency tables. The proposal is a multivariate extension of McNemar's test. McNemar's test could be applied to 2 × 2 contingency tables. The proposal may be also treated as a modification of Cochran's *Q* test which is used for testing dependency for multivariate binary data. The form of the test statistics that allows us to detect the departure from counts symmetry in multidimensional contingency tables is presented in the article. The permutation method of observations was used to estimate the empirical distribution of the test statistics. The considerations were supplemented with examples of the use of a multivariate test for simulated and real data. The application of the proposed test allows us to detect the asymmetrical distribution of counts in multivariate contingency tables.

Keywords: permutation test, McNemar's test, multivariate test

JEL: C12, C14, C15

1. Introduction

McNemar's test was proposed in 1947 (McNemar, 1947). It is a statistical test used for paired nominal data. This test is applied to 2 × 2 contingency tables with a binomial outcome, with matched pairs of subjects, to determine whether there is marginal homogeneity. Typical applications involve two independent raters providing dichotomous judgments for the same set of ratings, or a panel of separate rates responding on two occasions to the same dichotomous variable. Bowker (1948) presented a generalisation of McNemar's test for k ($k > 2$) variables. The generalisations of McNemar's test for square tables larger than 2 × 2 are often referred to as the Stuart-Maxwell test (Stuart, 1955; Maxwell, 1970). Some of the modifications concern the extension of the test application to quantitative data, others to nominal polynomial data, and still others to multidimensional dependent dichotomous data.

The purpose of this publication is to propose a permutation test to detect the departure from symmetry in multidimensional contingency tables. The proposed test, like the Cochran Q test, leads to testing the null hypothesis on the independence of k ($k > 2$) binary variables. The null hypothesis is the same as in Cochran's Q test, but the alternative hypotheses in these tests are different. The use of the Cochran Q test leads to the detection of existing differences in the percentage of responses for individual variables, and the proposed test, like the Bowker test, lets us detect asymmetry of counts in multivariate contingency tables.

2. McNemar's test

Let us consider (Y_{i1}, Y_{i2}) for $i = 1, 2, \ldots, n$ paired data with the binary response: "0" and "1". There are four possible outcomes for each pair: (0, 0), (0, 1), (1, 0) and (1, 1). Let us assume that:

a) the sample of n subjects has been randomly selected from the population;
b) each of the n subjects in the contingency table is independent of the other observations;
c) the scores of subjects are in the form of a dichotomous categorical measure involving two categories;
d) the sample size should not be extremely small.

The chi-square distribution is employed to calculate the McNemar's test statistic (McNemar, 1947). When the sample size is small, some sources endorse the use of a correction for continuity, while other sources prefer the exact binomial probability for the data to be computed instead of the chi-square based statistic (Fay, 2011).

Suppose (Y_{i1}, Y_{i2}) for $i = 1, 2, \ldots, n$ are identically and independently distributed bivariate data vectors. Let the mean vector of (Y_{i1}, Y_{i2}) be (p_1, p_2) and the null hypothesis H_0: $p_1 = p_2$ against the alternative H_1: $p_1 \neq p_2$. The observable data may be arranged in a 2 × 2 contingency table (see Table 1).

Table 1. Fourfold table for the presentation of data on matched samples

(Y_1, Y_2)	$Y_2 = 0$	$Y_2 = 1$	Total
$Y_1 = 0$	a	b	$n_{0\cdot}$
$Y_1 = 1$	c	d	$n_{1\cdot}$
Total	$n_{\cdot 0}$	$n_{\cdot 1}$	n

where $n = a + b + c + d$ is a sample size.

Source: own elaboration based on McNemar (1947)

The empirical probabilities for cells π_{ij} are shown in Table 2.

Table 2. Fourfold table for the presentation of empirical cell probabilities on matched samples

(Y_1, Y_2)	$Y_2 = 0$	$Y_2 = 1$	Total
$Y_1 = 0$	π_a	π_b	$\pi_{0\cdot}$
$Y_1 = 1$	π_c	π_d	$\pi_{1\cdot}$
Total	$\pi_{\cdot 0}$	$\pi_{\cdot 1}$	1

Source: own elaboration based on McNemar (1947)

McNemar's test could be used for testing the hypothesis:

$$H_0 : \pi_b = \pi_c$$

(it means that the theoretical proportion of cell b equals the proportion of cell c in the underlying population the sample represents), against the alternative hypothesis:

$$H_1 : \pi_b \neq \pi_c .$$

The test statistic in McNemar's test has the following form (McNemar, 1947):

$$Q = \frac{(b-c)^2}{b+c}. \tag{1}$$

Under the null hypothesis, with a sufficiently large number of discordant pairs (b and c), the statistic Q has a chi-square distribution with one degree of freedom.

For small sample sizes, the modified form Q_C of the test statistic (1) with continuity-correction should be calculated. This statistic has the following form (Sheskin, 2011):

$$Q_C = \frac{\left(|b-c|-1\right)^2}{b+c}. \tag{2}$$

Statistics Q and Q_C measure the asymmetry of counts in the contingency table (see Table 1). The test leads to rejection of the null hypothesis in the case of counts asymmetry in the contingency table.

An alternative form for the McNemar's test statistic is based on a normal distribution (Sheskin, 2011). The equation (3) can be employed to compute the McNemar's test statistic (1):

$$z = \frac{b-c}{\sqrt{b+c}}. \tag{3}$$

Another form of test statistic (2) could be written as follows:

$$z = \frac{|b-c|-1}{\sqrt{b+c}}. \tag{4}$$

For the small sample size, instead of statistic (2) with continuity-correction, the exact version of McNemar's test could be used (Sheskin, 2011). In this case, the odds ratio is calculated (Fay, 2011):

$$\phi = \frac{\theta}{1-\theta} \ , \tag{5}$$

where:

$$\theta = P\left(C \geq c\right) = \sum_{i=c}^{b+c} \pi_b^{\ i} \pi_c^{\ b+c-i}.$$

3. Some modifications and extensions of McNemar's test

McNemar's test may be applied to only binary categories for each outcome. It is possible that in some cases the outcomes could be classified into k categories where k is greater than two. The paired data that result from this type of experiment can be summarised in a $k \times k$ contingency table of counts. Such data can be analysed using the Bowker test (Bowker, 1948). The use of Bowker test for binary data ($k = 2$) leads to the same results as in McNemar's test so the Bowker test is a generalisation of McNemar's test. The form of the contingency table in the Bowker test for three binary variables ($k = 3$) is presented in Table 3.

Table 3. Contingency table in the Bowker test

(Y_1, Y_2)	$Y_2 = y_{21}$	$Y_2 = y_{22}$	$Y_2 = y_{23}$
$Y_1 = y_{11}$	n_{11}	n_{12}	n_{13}
$Y_1 = y_{12}$	n_{21}	n_{22}	n_{23}
$Y_1 = y_{13}$	n_{31}	n_{32}	n_{33}

Source: own elaboration based on Bowker (1948)

The form of the null hypothesis in the Bowker symmetry test could be written as follows:

$$H_0: \pi_{ij} = \pi_{ji} \text{ for all } i \neq j \text{ versus } H_1: \pi_{ij} \neq \pi_{ji}, \text{ for some } i \neq j.$$

If n_{ij} is the count of i-th row and j-th column in the contingency table (see Table 3), then the test statistic could be written in the following form (Bowker, 1948):

$$B = \sum_{i=1}^{n} \sum_{j>i} \frac{\left(n_{ij} - n_{ji}\right)^2}{n_{ij} + n_{ji}}. \quad (6)$$

Under the null hypothesis, the test statistic B has an asymptotic chi-square distribution, with $k(k-1)/2$ degrees of freedom. In the case of $k = 3$, the statistic (6) has the following form:

$$B = \frac{\left(n_{12} - n_{21}\right)^2}{n_{12} + n_{21}} + \frac{\left(n_{13} - n_{31}\right)^2}{n_{13} + n_{31}} + \frac{\left(n_{23} - n_{32}\right)^2}{n_{23} + n_{32}}. \quad (7)$$

It is visible that only counts that are symmetric in pairs (n_{12} and n_{21}, n_{13} and n_{31}, n_{23}, and n_{32}) in the contingency table are compared.

There are some well-known extensions of McNemar's test. The generalisations of McNemar's test for square tables larger than 2×2 are often referred to as the Stuart-Maxwell test (Stuart, 1955; Maxwell, 1970). Feuer and Kessler (1989) discussed the generalisations of McNemar's test based on the case of two independent samples of paired univariate binary responses. They considered the null hypothesis that the marginal changes in each of two independently sampled tables are equal. Agresti and Klingenberg (2005) (see also Klingenberg, Agresti, 2006) considered methods for comparing two independent multivariate binary vectors. Pesarin (2001) considered methods for comparing two dependent vectors of sample proportions.

Westfall, Troendle and Pennello (2010) consider the problem of multiple comparisons of dependent proportions. They argue that multiple comparisons of dependent proportions can be made more powerful by utilising testing methods and incorporating dependence structures. They proposed a method that utilises stepwise testing and discrete characteristics for exact McNemar's test.

McNemar (1947) assumed that the data being analysed are measured on a nominal or ordinal scale. However, the experimental data may be often measured on at least an interval scale. Oyeka (2012) proposed an extension of McNemar's test which could be used for data measured on an interval or ratio scale. This modification is based on data transformation from a continuous scale to a nominal scale based on the formula:

$$u_i = \begin{cases} 1 & \text{for } y_{i2} \geq 0 \text{ and } y_{i1} < 0 \\ 0 & \text{for } y_{i2}\ y_{i1} \geq 0 \\ -1 & \text{for } y_{i2} > 0 \text{ and } y_{i1} \geq 0 \end{cases}$$

Cochran's Q test is an extension of McNemar's test to more than two matched samples (Donald, Shahren, 2018). When the Cochran's Q test statistic is computed with only $k = 2$ groups, the results are equivalent to the results obtained from McNemar's test. Cochran's Q could be also considered to be a special case of the Friedman test (Sheskin, 2011). When the responses are binary, the Friedman test becomes Cochran's Q test.

Suppose that there are k binary measurements on each of n subjects. Let y_{ij} be the binary response from the subject i in the category j ($i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, k$), with success = 1 and failure = 0. The null hypothesis for Cochran's Q test is that there are no differences between the categories (Sheskin, 2011). If the calculated probability is low (p-value is less than the selected significance level α), the null-hypothesis is rejected, and it can be concluded that proportions in at least 2 of k variables are significantly different from each other.

The null hypothesis in Cochran Q test could be written as follows:

$$H_0 : p_1 = p_2 = \cdots = p_k$$

versus $H_1 : p_i \neq p_j$, for at least one pair i, j where $i \neq j$ and $1 \leq i, j \leq k$.

The test statistic Q has the following form:

$$Q = \frac{(k-1)\left(kC - T^2\right)}{kT - R}, \tag{8}$$

where:

$T = \sum_{i=1}^{n} \sum_{j=1}^{k} y_{ij}$, $R = \sum_{i=1}^{n} \left(\sum_{j=1}^{k} y_{ij} \right)^2$ and $C = \sum_{j=1}^{k} \left(\sum_{i=1}^{n} y_{ij} \right)^2$.

Under the null hypothesis, the test statistic Q has an asymptotic chi-square distribution with $k - 1$ degrees of freedom.

4. Proposal of a multivariate extension of McNemar's test

Suppose $(Y_{i1}, Y_{i2}, \ldots, Y_{ik})$ for $i = 1, 2, \ldots, n$ and $k > 2$ are identically and independently distributed bivariate data vectors. The multivariate binary data could be arranged as in McNemar's test in a contingency table. In the case of k variables, the contingency table will be a k-dimensional contingency table. The example of a three-dimensional contingency table is shown in Table 4.

Table 4. Three-dimensional contingency table

$Y_3 = 0$			**$Y_3 = 1$**		
(Y_1, Y_2)	$Y_2 = 0$	$Y_2 = 1$	(Y_1, Y_2)	$Y_2 = 0$	$Y_2 = 1$
$Y_1 = 0$	n_{000}	n_{010}	$Y_1 = 0$	n_{001}	n_{011}
$Y_1 = 1$	n_{100}	n_{110}	$Y_1 = 1$	n_{101}	n_{111}

Source: own elaboration

The hypothesis that the theoretical probabilities in symmetric cells in a multidimensional contingency table are equal will be considered. This hypothesis could be written as follows:

$$H_0 : \pi_{i_1, i_2, \ldots, i_k} = \pi_{i_1', i_2', \ldots i_k'} \text{ for all } i_1, i_2, \ldots, i_k \in \{0,1\}^k \tag{9}$$

against the alternative:

$$H_1 : \pi_{i_1, i_2, \ldots, i_k} \neq \pi_{i_1', i_2', \ldots i_k'} \text{ for some } i_1, i_2, \ldots, i_k \in \{0,1\}^k,$$

where:

$i_s \in \{0,1\}$ for $s = 1, 2, \cdots, k$ and $i_s' = 1 - i_s$ for $s = 1, 2, \cdots, k$.

The test statistic for the proposed extension of McNemar's test is based on the idea from the Bowker test statistic which allows us to detect counts asymmetry in a contingency table. The test statistic has the following form:

$$M_k = \sum_{i_1,i_2,\ldots,i_k} \frac{\left(n_{i_1,i_2,\ldots,i_k} - n_{i'_1,i'_2,\ldots i'_k}\right)^2}{n_{i_1,i_2,\ldots,i_k} + n_{i'_1,i'_2,\ldots i'_k}}, \tag{10}$$

where:
$i_s \in \{0,1\}$ for $s = 1,2,\cdots,k$ and $i'_s = 1 - i_s$ for $s = 1,2,\cdots,k$.

The idea of the test statistic M_k is based on the test statistic (6) used in the Bowker symmetry test. The main goal is to detect counts asymmetry in a contingency table. The independent permutation of each variable is used for obtaining the empirical distribution of the M_k statistics under the null hypothesis. Permutation tests have optimum properties, which means good merit for its practical use (Oden, Wedel, 1975). The recommended number of data permutation should be $N \geq 1000$ (Pesarin, 2001; Kończak, 2016). The value of the test statistic for the sample data is denoted by M_{k0}. The estimated *p*-value is calculated as follows: $\frac{card\{i : M_{ki} \geq M_{k0}\}}{N}$, where M_{ki}(i = 1, 2, …, N) is the value of the test statistic in the *i*-th permutation and M_{k0} is the value of the test statistic for non-permuted data.

5. Multivariate extension – empirical verification

Two examples of the use of the proposed multivariate permutation extension of McNemar's test are presented. The first example is based on the simulated data and the other is based on the real data obtained from the *Diagnoza społeczna* survey (2019). The results of this test are compared to another extension – Cochran's Q test.

5.1. Empirical verification – simulation data

Let us consider the probability model (Table 5) of the three-dimensional random variable (Y_1, Y_2, Y_3).

Table 5. Three-dimensional probabilitymodel

$Y_3 = 0$		
(Y_1, Y_2)	$Y_2 = 0$	$Y_2 = 1$
$Y_1 = 0$	0.30	0.18
$Y_1 = 1$	0.18	0.00

$Y_3 = 1$		
(Y_1, Y_2)	$Y_2 = 0$	$Y_2 = 1$
$Y_1 = 0$	0.04	0.00
$Y_1 = 1$	0.00	0.30

Source: own calculations

The symmetric element in the three-dimensional probability table for the element (0, 0, 1) with probability $\pi_{001} = 0.04$ is the element (1, 1, 0) with probability $\pi_{110} = 0$. The symmetric element for the element (0, 1, 1) with probability $\pi_{011} = 0$ is the element (1, 0, 0) with probability $\pi_{100} = 0.30$. The symmetric element for the element (1, 0, 1) with probability $\pi_{101} = 0$ is the element (0, 1, 0) with probability $\pi_{010} = 0.18$.

The random sample of the size $n = 30$ of the random vector $Y = (Y_1, Y_2, Y_3)$ is shown in Table 6.

Table 6. The simulated sample data for $k = 3$

Y_1	Y_2	Y_3	n_i
0	0	0	10
0	0	1	1
0	1	0	4
0	1	1	0
1	0	0	6
1	0	1	0
1	1	0	0
1	1	1	9

Source: own elaboration

Based on the source data from Table 6, the data could be presented in a three-dimensional contingency table (see Table 7). Cochran's Q test for data presented in Table 6 leads to a decision that there is not enough evidence to reject the null hypothesis ($Q = 3.455$, p-value 0.1778).

Table 7. Three-dimensional contingency table for data from Table 6

$Y_3 = 0$		
(Y_1, Y_2)	$Y_2 = 0$	$Y_2 = 1$
$Y_1 = 0$	10	4
$Y_1 = 1$	6	0

$Y_3 = 1$		
(Y_1, Y_2)	$Y_2 = 0$	$Y_2 = 1$
$Y_1 = 0$	1	0
$Y_1 = 1$	0	9

Source: own calculations

The symmetric element in the three-dimensional contingency table for the element (0, 0, 1) with counts $n_{001} = 1$ is the element (1, 1, 0) with count $n_{110} = 0$. The symmetric element for the element (0, 1, 1) with counts $n_{011} = 0$ is the element (1, 0, 0) with count $n_{100} = 6$. The symmetric element for the element (1, 0, 1) with counts $n_{101} = 0$ is the element (0, 1, 0) with count $n_{010} = 4$.

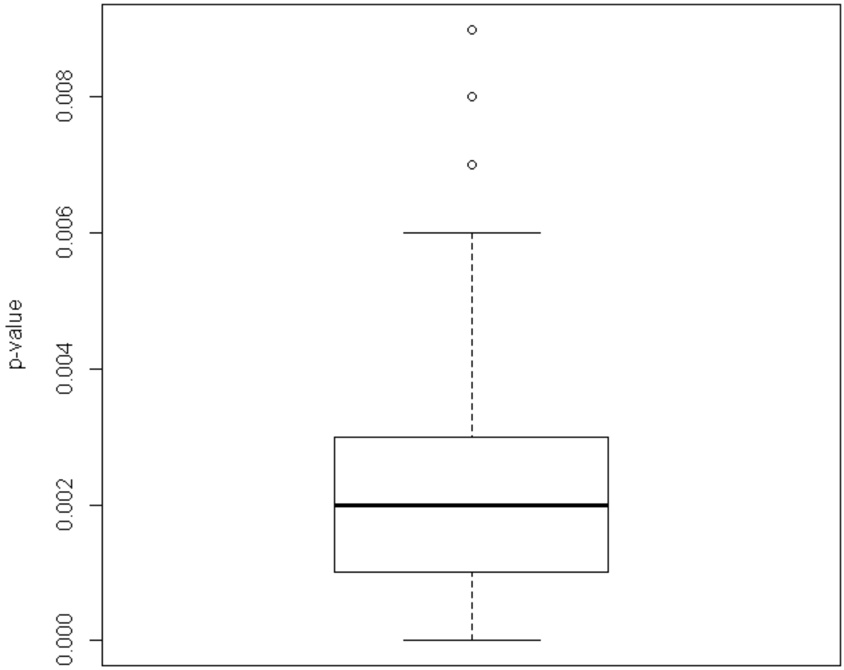


Figure 1.Distribution of *p*-values in the proposed multivariate permutation extension of McNemar's test for data from Table 6

Source: own elaboration

The proposed multivariate test uses a data permutation method to estimate the distribution of the test statistic M_k. Due to the Monte Carlo procedure, the estimated *p*-values may differ in subsequent simulations. There were made $N = 1000$ runs of the proposed test for the considered data. The *p*-values were from 0.000 up to 0.009 and the empirical distribution is shown in Figure 1. In each of the N tests, the decision was to reject the null hypothesis. This example shows the difference between the Cochran Q test and the proposed multivariate permutation extension of McNemar's test. It could be seen that in the case of counts asymmetry in the three-dimensional contingency table the proposed test leads to the rejection of the null hypothesis even if the Cochran Q test leads to the decision that we have not enough evidence to reject the null hypothesis.

5.2. Empirical verification – real data case

To show differences between the Cochran Q test and the proposed modification of McNemar's test, the data from the *Diagnoza społeczna* survey (2019) were used. *Diagnoza społeczna* is based on panel research. Researchers return to the same households every few years, with the first sample being taken in the year 2000 and the last sample in the year 2015. One of the questions asked concerns escaping into alcohol in order to deal with problems and difficulties in 2007, 2011, and 2015. The results are shown in Table 8.

Table 8. Three-dimensional contingency table for the following question "I reach for alcohol" (0 – "no", 1 – "yes")

$Y_{2015} = 0$		
(Y_1, Y_2)	$Y_{2011} = 0$	$Y_{2011} = 1$
$Y_{2007} = 0$	2204	44
$Y_{2007} = 1$	53	13

$Y_{2015} = 1$		
(Y_1, Y_2)	$Y_{2011} = 0$	$Y_{2011} = 1$
$Y_{2007} = 0$	29	22
$Y_{2007} = 1$	8	19

Source: own calculations based on data from *Diagnoza społeczna* (2019)

The percentages of "yes" answers for all variables (years 2007, 2011, 2015) are equal to 3%, 3% and 2.4% respectively in 2007, 2011 and 2015. This leads to the statement that we have not enough evidence to reject the null hypothesis in the Cochran Q test ($Q = 3.846$, p-value = 0.1462).

It is visible that counts in the three-dimensional contingency table are not symmetric. The proposal of the permutation extension of McNemar's test should detect the asymmetry in these counts. The symmetric element in the three-dimensional contingency table (see Table 7) for the element (0, 0, 1) with counts $n_{001} = 29$ is the element (1, 1, 0) with count $n_{110} = 13$. The symmetric element for the element (0, 1, 1) with counts $n_{011} = 22$ is the element (1, 0, 0) with count $n_{100} = 53$. The symmetric element for the element (1, 0, 1) with counts $n_{101} = 8$ is the element (0, 1, 0) with count $n_{010} = 44$.

The permutation test was performed $N = 1000$ times. For each series of permutation tests, there was strong evidence to reject the null hypothesis. In each case, the p-values were from 0.000 up to 0.009. The asymmetry in the three-dimensional contingency table leads to rejecting the null hypothesis by the proposed multivariate permutation extension of McNemar's test.

6. Conclusions

The permutation multivariate extension of McNemar's test was proposed in the paper. This test could be used to detect the dependency for multidimensional dependent binary variables. The test can also be considered as a modification of the Cochran Q test. The proposed test is similar to the Q Cochran test, but the test statistic is based on the test statistic as in the Bowker symmetry test. The main idea of the proposed multivariate permutation test is to detect counts asymmetry in the contingency table. The examples with the use of the simulation and the real data were presented in the paper. The presented calculations have shown that the proposal leads to effective detection of counts asymmetry in the multidimensional contingency table. A special property of the proposed test is the ability to detect asymmetry of counts in a multidimensional contingency table.

Acknowledgments

This paper was presented at the conference MSA 2019 which financed its publication. Organisation of the international conference "Multivariate Statistical Analysis 2019" (MSA 2019) was supported byresources for the popularisation of scientific activities of the Minister of Science and Higher Education in the framework of agreement No 712/P-DUN/202019.

References

Agresti A., Klingenberg B. (2005), *Multivariate tests comparing binomial probabilities, with application to safety studies for drugs*, "Applied of Statistics", vol. 54, pp. 691–816.

Bowker A.H. (1948), *A test for symmetry in contingency tables*, "Journal of American Statistical Association", vol. 43, pp. 572–574.

Diagnoza społeczna: zintegrowana baza danych, www.diagnoza.com [accessed: 23.11.2019].

Donald S., Shahren A.Z.A. (2018), *Cochran's Q with Pairwise McNemar for Dichotomous Multiple Responses Data: a Practical Approach*, "International Journal of Engineering & Technology", vol. 7, no. 3(18), pp. 4–6, https://www.sciencepubco.com/index.php/ijet/article/view/16662 [accessed: 23.11.2019].

Fay M. (2011), *Exact McNemar's Test and Matching Confidence Intervals*, https://www.researchgate.net/publication/267448346_Exact_McNemar's_Test_and_Matching_Confidence_Intervals [accessed: 23.11.2019].

Feuer E.J., Kessler L.G. (1989), *Test Statistic and Sample Size for a Two-Sample McNemar Test*, "Biometrics", vol. 45, no. 2, pp. 629–636.

Klingenberg B., Agresti A. (2006), *Multivariate Extensions of McNemar's Test*, "Biometrics", vol. 62, pp. 921–928.

Kończak G. (2016), *Testy permutacyjne. Teoria i zastosowania*, Uniwersytet Ekonomiczny w Katowicach, Katowice.

Maxwell A.E. (1970), *Comparing the classification of subjects by two independent judges*, "British Journal of Psychiatry", vol. 116(535), pp. 651–655, http://doi.org/10.1192/bjp.116.535.651

McNemar Q. (1947), *Note on the sampling error of the difference between two correlated proportions in percentages*, "Psychometrica", vol. 12, pp. 153–157.

Oden A., Wedel H. (1975), *Arguments for Fisher's Permutation Test*, "The Annals of Statistics", vol. 3, no. 2, pp. 518–520.

Oyeka I.C.A. (2012), *Modified McNemar Test*, "International Journal of Statistics in Medical Research", vol. 1, pp. 73–78.

Pesarin F. (2001), *Multivariate Permutation Tests with Applications in Biostatistics*, John Wiley & Sons, Chichester.

Sheskin D.J. (2011), *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Boca Raton.

Stuart A.A. (1955), *A test for homogeneity of the marginal distributions in a two-way classification*, "Biometrika", vol. 42(3–4), pp. 412–416, http://doi.org/10.1093/biomet/42.3–4.412

Westfall P.H., Troendle J.F., Pennello G. (2010), *Multiple McNemar Tests*, "Biometrics", vol. 66, no. 4, pp. 1185–1191.

Wielowymiarowe permutacyjne rozszerzenie testu McNemara

Streszczenie: Celem artykułu jest przedstawienie propozycji testu permutacyjnego do wykrywania odchyleń od symetrii układu liczebności w wielowymiarowej tablicy kontyngencji. Propozycja jest wielowymiarowym rozszerzeniem testu McNemara, który stosuje się do tablic o wymiarach 2 × 2. Przedstawiony test można również traktować jako modyfikację testu *Q* Cochrana, który służy do testowania zależności dla wielowymiarowych danych binarnych. Przedstawiono postać statystyki testu, która pozwala wykryć odchylenie od symetrii liczebności w wielowymiarowej tabeli kontyngencji. Do oceny rozkładu teoretycznego statystyki testowej zastosowano metodę permutacji obserwacji. Rozważania zostały uzupełnione przykładami zastosowania proponowanego testu dla danych symulowanych i rzeczywistych. Zastosowanie proponowanego testu pozwala wykryć asymetryczny rozkład liczebności w wielowymiarowych tabelach kontyngencji.

Słowa kluczowe: test permutacyjny, test McNemara, wielowymiarowy test

JEL: C12, C14, C15

OPEN ACCESS	© by the author, licensee Łódź University – Łódź University Press, Łódź, Poland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license CC-BY (https://creativecommons.org/licenses/by/4.0/)
	Received: 2020-04-06; verified: 2020-08-28. Accepted: 2020-09-30
COPE Member since 2018 JM13703	This journal adheres to the COPE's Core Practices https://publicationethics.org/core-practices

Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663
www.czasopisma.uni.lodz.pl/foe/
4(349) 2020

DOI: http://dx.doi.org/10.18778/0208-6018.349.07

Urszula Ornarowicz
Warsaw School of Economics, The Collegium of Management and Finance, uornar@sgh.waw.pl

Innovations. Market and Social Aspects

Abstract: Areas of research in economics and management science become increasingly close – they overlap and become very similar. New events, new products of people's actions, new patterns of behaviour arise with a pace unknown before. Institutionalisation of these phenomena aimed at their broad codification also takes on new forms. We live in an age of ubiquitous innovativeness. Naturally, the question arises: should innovations be perceived in the same way as in the past? Are there any new types of innovations that have appeared lately? Are the current definitions of market and social innovations still up to date?

The aim of the article is to present a change in approach to innovations over time, with particular focus on their market and social aspects. The author attempts to answer the following questions: how did technological progress visible in the networking of economy influence the understanding of social innovations, what is the role of social production and exchange which replace gradually market exchange, in the social innovation definition, to what extent is the cooperation within a community in the virtual space characteristic of a special class of social innovation?

The research method used by the author is based on literature studies on innovations and on the economics of cooperation (access, sharing, co-use). It comprises an analysis of different concepts of innovation, in particular different definitions of the name, an analysis of different approaches to cooperation economics, comparisons of the obtained results, and conclusions formulation.

The approach to innovation changes over time – from a technical, social and market approach to a differently understood today social approach. Contemporary, the criteria for innovation "society" are different. The understanding of innovation is influenced by the increased role of social production and exchange at the expense of market exchange. The networking of the information economy significantly strengthens the social aspect of innovation. Cooperation within a community, including co-creation of goods, access to them, their co-use and sharing, is an extreme example of the advantage of the social dimension of innovation over its market aspect.

Keywords: social innovations, market innovations, technical innovations

JEL: O3, M1

1. Introduction

There are many events that we are eager to call innovations. Below are some examples:

1. Nowadays in Poland an electric car is considered to be an innovation, still not very popular among Poles due to its high price and insufficient infrastructure regarding battery charging – while in the EU countries in the first half of 2016 more than 44,000 of such new cars were registered, in Poland there were only 33 of them (*Pojazd elektryczny*, 2017).
2. In May 2016, at the MOTO SHOW in Cracow, a prototype hydrogen car (to be more precise: an electric car powered with hydrogen) was called an innovation and presented for the first time in front of a wider audience – HYDROCAR PREMIER, expressing the latest Polish technical thought materialised in an innovative power system and unique design (*Pierwszy polski samochód wodorowy...*, 2016).
3. Today, in our geographical latitude, no one calls an oil lamp an innovation, however, it was considered as such in 1853 – constructed by Ignacy Łukasiewicz and ignited for the first time in Lviv in March of the same year, and on 31 July used for the first time during the night operation in the city hospital (*Lampa naftowa*, 2017).
4. Organising the World Road Championships for the first time in 1921 (in Copenhagen, the capital of Denmark) was an innovation, as well as organising and also introducing an individual time trial in 1994 to the programme of the Championships (*Mistrzostwa świata…*, 2017).
5. The sequence of innovations connected with the creation of human thought such as the circle, having the confirmation in archaeological discoveries, begins with the invention of the pottery wheel by the Sumerians about 3250 years BC. The use of wheels in the very first wheeled vehicles – chariots – is also confirmed almost a thousand years later; and in more recent times, e.g. in 1869, a patent for wheels with wire spokes, which gave rise to two-wheeled bicycles with a large front wheel and a small rear one, in other words – bicycles and their contemporary successors (Ługowik, 2017).
6. Among the latest examples of innovations, the Ice Arena in Tomaszów Mazowiecki can be pointed out, officially opened on 14 December 2017 – the largest and most modern in Poland and Europe, and meanwhile, the first indoor speed skating track in Poland, which in addition to bobsleigh operates all ice disciplines (*Prezydent Duda otworzył Arenę Lodową…*, 2017).

The examples of the events indicated above are called colloquially innovations. Innovation in the common understanding has been for the last fifty years identified with either the broadly understood novelty or in a narrower sense with a change that means either the process of introducing something new or the effect of this process, which is something recently introduced. And likewise, the term innovation is interpreted in this way in Polish language dictionaries, for example, from 1968, 1988, 2003, and also from 2017 (Skorupka, Auderska, Łępicka, 1968: 236; Szymczak, 1988: 792; Sobol, 2003: 274; *Słownik języka polskiego PWN*, 2017).

Should broadly understood innovation be identified with novelty also on scientific grounds? How broadly should we define innovation in order not to identify it with novelty? What are the typical approaches to innovation in the literature? How should we understand technical, market and social innovations in the light of broadly understood innovation, and whether they mean the same today as in the past? I will try to answer these questions in the article.

2. Research method

The interest in the term "innovation" in the literature has been steadily growing for at least 40 years, which may be reflected in the number of publications on this subject included in the business database of EBSCO publications. Four decades ago, that number increased by 2,900 positions, and in the last decade 2008–2018 by as many as 88,449; in the case of scientific literature, the figures were respectively: 1,563 and 30,266. The analysis of the number of publications on innovation obtained from this database is an important method used in my research, complementary to the qualitative analysis of the content of publications – literature studies, some of which I also conduct on the basis of this database. The next steps in my research programme on innovation and its types are as follows:

1. Identification of different meanings of "innovation" on the basis of literature (in the first stage of the research, regardless of the year of publication), comparison of these meanings, grouping them and identification of (five) typical approaches to innovation on this basis. Identification of the characteristics of relevant innovations in each approach.
2. Operationalisation of the meaning of each of the approaches – its expression using a typical set of data approaches (in particular one-piece).
3. Searching in the EBSCO database, in the general set of publications, for those subsets of publications that correspond to each of the identified types of innovation on the basis of the existence or co-existence of terms-characteristics attributed to the type of innovation, taking into account the four decades identified between 1979 and 2018.
4. Conclusion on the existence/non-existence of sets of publications corresponding to the types of innovation identified, as well as on the number of sets of publications corresponding to the identified types of innovation.
5. Providing literature examples of innovation definitions or interpretations of that meaning that correspond to each approach, taking into account the

literature described in point 1, supplemented by the ESCO database and its publications from 1979–2018 and not yet completed 2019. I would like to draw your attention to the articles in two journals on management and organisational theory which belong to the five most frequently quoted scientific journals in this discipline, i.e. “Academy of Management Journal” and “Academy of Management Review”.
6. Identification of the meaning of the terms such as “market innovation” and its different types, as well as “social innovation”. Operationalisation of these terms, separation in the ESCO database of subsets of publications devoted to the examined issues and indication of their number (similarly as in points 1–4).
7. Interpretation of the understanding of the name of social innovation in its former and contemporary meaning, taking into account the principles of a new management philosophy, i.e. the collaborative economy.

3. Innovations in general – aspects not limited to novelty

The identification of innovations with the generally understood novelty is rooted not only in colloquial language but sometimes also in the scientific work of representatives of various disciplines. I do not think that expanding the naming scope of innovation for every novelty is inappropriate. Novelty is a change. It indicates something that has not been used before – something completely, absolutely new, as it did not exist before (everything in it is new and has only new features); something new, because in this place and time it has never been used before; something new only in some respect, because something has been changed, thus it is entitled to a new feature which previously was not there. This something can be, e.g. a new product, new service, new method of operation, new event, or new natural phenomenon. However, the lastly mentioned examples could be hardly called innovations. We will not call an innovation, e.g. the phenomenon of global warming, even when it is a recognised phenomenon that has been happening for some time (I disregard assessing the validity of existing parallel and, at the same time, contradictory views on climate warming/cooling). We will also not call an innovation the observed phenomenon of increasing strength of tropical cyclones (hurricanes and typhoons)[1] and their destructive activities around the globe.

1 Tropical cyclones are extensive low-pressure systems with winds with an average speed above 120 km/h, forming over the waters of the tropical and equatorial zones, called hurricanes in the Atlantic and Eastern Pacific, and typhoons – in the Western Pacific (Popkiewicz, Malinowski, 2017).

The term "innovation" does not refer to all events that adhere to the novelty feature (with regards to the degree of its intensity and its relativisation to time and place), but only to those of them that directly or indirectly affect people's actions. For this reason, I consider the interpretation of innovation in the field of praxeology as a science of efficient operation to be particularly valuable, indicating that the action in this science refers only to a specific type of human behaviour. I recognise the praxeological interpretation of innovation to be the basis for any reflection on innovation.

In praxeology, the action of a human being (the subject of action, the causer) is a special kind of behaviour – it is a behaviour which is: 1) purposeful (aims at achieving the set goal), 2) conscious, 3) preceded by the subject's decision of behaving in a certain way, and 4) wanted by the subject (Pszczołowski, 1978: 56–59). The action is intended, approved by the causer, it is something which is under the causer's control or in the causer's power. The action is not only related to the movement of muscles, as Pszczołowski says, although this type of action, especially connected with people's physical activity, is easily observable. Thinking is also an activity, in other words: a reflection as an internal action.

Due to the fact that the term "innovation" in general refers not only to actions but also to their effects, the term of product in praxeological meaning is useful for describing innovation (Pszczołowski, 1978: 280–281). In this case, the "product" means something that is the result (effect, conclusion) of an action, that is, an object in its broad sense (all that one can talk about and think of), both material and non-material (among others: features, relationships, events – changing or not changing something, including actions and their results). The effect of an action, as explicitly stated by von Wright (1963: 39), is a change related to this action, i.e. a kinetic event, or alternatively – the final state of this change, i.e. a static event. The product is always an effect of someone's action or actions of some people at a specific time, i.e. it always refers to a specific person and time (Pszczołowski, 1978: 280). In addition, taking into account some of the criteria for product differentiation, one can indicate in their general collection, among others: individual and collective, partial and final products, as well as products of internal (mental) and external activities. A material product is one of a particular kind, different from the causer of action, in contrast to, e.g. an artist's work, which is a product of a genius, existing in the initial stage generally in an immaterial form and connected with the causer – situated in his or her mind.

Taking into consideration that the praxeological interpretation of the terms "action" and "product" when explaining the meaning of innovation, it should be noted that under the general praxeological statement that "innovations are concerned with actions (people) and their products", there are many various events related to this. Such a statement means that innovations have reference to, among others: ideas, activities, processes, methods, procedures, tools, production factors, resources, products, goods, services, etc. Due to such a large, and in principle an infinite number of events that may be the subject of innovation, I believe that it is not reasonable to attempt to list some of these events or enumerate their types when defining innovations in their general sense. Indicating in this matter a given case of events is justified only when defining a specific type of innovation, including the aforementioned market or social innovations. In addition, the fact that innovations are related to activities and their products, and are a kind of product of novelty, activities as well as products, also results from the fact that they should be considered in relation to the person (people), place and time (operations and products are characterised on the grounds of a person/persons and time, nevertheless, the new things – on account of place and time).

In addition to the already mentioned feature of innovation, which distinguishes it from novelty, namely referring novelty to actions and their products, it is necessary to indicate the second feature of innovation which I call utility or usefulness. In the praxeological sense, the term "useful" means "being used to enable or facilitate the attainment of a given goal". A useful object is a necessary object (Pszczołowski, 1978: 266). I understand the feature of innovation (its usefulness) as a distinctive combination of two other features, i.e. the assessment of the positive aspect of innovation, in other words, being valuable or priceless, its "practicality" and "applicability". Both these features in this case co-exist and depend on each other. If something is valuable, it is usually applicable, and because it finds its application, it becomes even more valuable. In the interpretation of the utility of a broadly understood subject, the emphasis is often placed on its positive assessment, sometimes on practicality, its implementation, and application. Taking into account both important features of innovations discussed here, innovation in the general sense should be understood as new and useful (i.e. valuable and applicable) human activities and their effects.

In my opinion, theses included in my current argumentation are the basis for distinguishing five typical approaches to broadly understood innovation. They are presented below in the following way: in each case the approach is briefly characterised, the resulting symbol of the set of designations of the term innovation is provided (otherwise: the symbol of the scope of this term) and exemplary definitions of innovation, typical for a given approach, are identified.

Approach 1

I consider as the most general approach to innovation the one in which innovation is identified with novelty (N), which is perceived as the only important feature of innovation.

The symbol of the collection of innovation designations: N. One of the meanings of innovation in the previously mentioned four Polish language dictionaries is simply the term "novelty" (Skorupka, Auderska, Łępicka, 1968: 236; Szymczak, 1988: 792; Sobol, 2003: 274; *Słownik języka polskiego PWN*, 2017).

Approach 2

The scope of the term innovation identified with novelty (N) is narrower, provided that it concerns the actions of people and their products (D).

The symbol of the collection of innovation designations: ND.

Examples of defining or interpreting the importance of innovation by selected authors:

1. Kotler: "Innovation refers to any good, service or idea that is perceived by someone as new". The author draws attention to the subjective aspect of innovation, adding that: "An idea may have existed for a long time, but it is an innovation for the person who perceives it as a new one" (Kotler, 1994: 322).
2. Rogers and Shoemaker: "An Innovation is defined an idea perceived as new by an individual or system" (Rogers, Shoemaker, 1971).
3. Rogers: "Innovation is an idea, practice or object that is perceived as new by an individual or a group, a group that receives it". The author adds that: "if the idea seems new to the individual, it is an innovation" (Rogers, 2003: 12).

Approach 3

Another feature of innovation, in addition to the previously indicated, that is novelty (N) regarding actions and their products (D) is the practical feature of their application (Z).

The symbol of the collection of innovation designations: NDZ.

Examples of defining or interpreting the importance of innovation by selected authors:

1. Pszczołowski: Innovation is "a new product (a novelty relative to the place and time) which through the imitation is disseminated in practice" (Pszczołowski, 1978: 83).
2. Koch: "Innovation – commercialisation of an invention: introduction of a new product or service to the market" (Koch, 1997: 89).
3. *Oslo Manual*: "Innovation is the implementation of a new or significantly improved product (or service) or process, a new marketing method or a new organisational method into the business practice, in the field of workplace organisation or relations with the environment" (*Oslo manual*..., 2005: 46).
4. Compagni, Mele and Ravasi: The book *How Early Implementations Influence Later Adoptions of Innovation* (Compagni, Mele, Ravasi, 2015) shows the implementation and application of innovation.
5. Birkinshaw, Hamel and Mol: "We define management innovation as the invention and implementation of a management practice, process, structure, or technique that is new to the state of the art and is intended to further organizational goals" (Birkinshaw, Hamel, Mol, 2008: 825).

In this approach to innovation, the authors differ significantly in their views on the subject of: which stage of "making a novelty more practical" should be considered typical for innovation. Some of the authors, such as Baruk (2009: 13; 2013: 10–11), believe that this stage is the first contact with the recipient of innovations, such as: the first introduction of a new or improved product into manufacturing and the market, the first application in the production of new or improved methods of production, and analogically, the first specific events in the case of work and production organisation, management methods or marketing. A different view is presented, for example, by the already mentioned Pszczołowski (1978: 83), for whom not only the first stage of making the novelty more practical is an innovation but also its promulgating in practice through imitation. This stage, also called diffusion, is recognised by the author as contributing to innovation. In addition to these two extreme views, there are also many others that enrich discussions around the course of the process of innovation.

Approach 4

The features of significant innovativeness, apart from the ones indicated in the second approach, e.g. the novelty (N) concerning actions and their products (D), also include the feature of value (C) – a certain assessment of events resulting from the application of novelty in practice, however, without any special emphasis on the implementing actions, "making these novelties more practical".

The symbol of the collection of innovation designations: NDC.

Examples of defining or interpreting the importance of innovation by selected authors:

1. Robbins and DeCenzo after Stevens (1999): "Innovation is the process of transforming a creative idea into a useful product, service or modus operandi" (Robbins, DeCenzo, 2002: 345).
2. Pietrasiński: "Innovations are changes deliberately introduced by man or cybernetic systems designed by a human which consist of replacing previous states with others, evaluated positively in the light of specific criteria and making up a whole for progress" (Pietrasiński, 1971: 9).
3. Drucker: "Innovation is an […] action that gives resources new opportunities to create wealth" (Drucker, 1992: 39).
4. Jia, Huang and Zhang: "Innovative activities create considerable value for the company" (Jia, Huang, Zhang, 2019: 220).
5. Utterback: "[…] the effectiveness of innovation as a dependent variable" (Utterback, 1971: 76).

Approach 5:

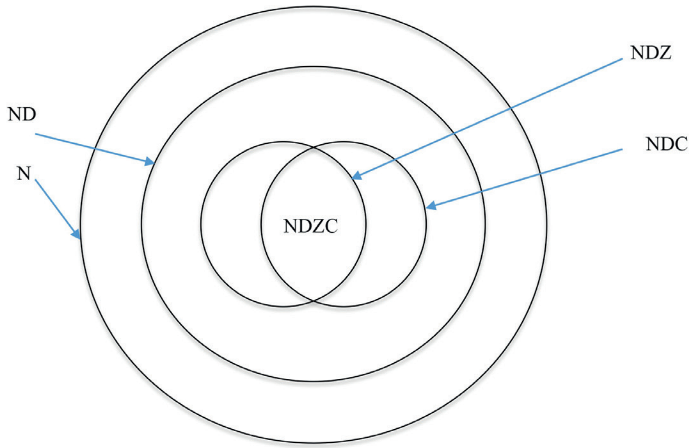
Important features of innovations are: novelty (N) of activities or their products (D), their application in practice (Z), as well as positive evaluation of subsequent events and their value (C). Referring to the sum calculus, this is a common part of the NDZ and NDC sets indicated respectively in Approaches 3 and 4.

The symbol of the collection of innovation designations: NDZC.

Examples of defining or interpreting the importance of innovation by selected authors:

1. West and Farr define innovation as "intentional introduction as part of work and application of work team or organization of ideas, processes, products or procedures that are new to this work, work team or organization and which aim to bring benefits to this work, team work or organization" (West, Farr, 1990: 9).
2. Damanpour: "The adoption of innovations is conceived to encompass the generation, development and implementation of new ideas or behaviors […] The adoption of innovation is generally intended to contribute to the performance or effectiveness of the adopting organization" (Damanpour, 1991: 556).
3. Schippers, West and Dawson: "Innovation subsumes creativity – the generation of new ideas – but, in addition, includes the implementation of the ideas […] and is seen as an important factor in organizational effectiveness and survival […]". In reference to the latter, the authors emphasise that: "innovations should also be judged on the basis of whether they prove effective in practice" (Schippers, West, Dawson, 2012: 5).
4. Deschamps does not mention innovation but innovative character. On the basis of innovation defined by him, one can conclude the meaning to be as follows: "[…] innovation means striving to change the existing state of affairs, the change that one hopes will introduce a better product, service, process or management". The author believes that "innovation is to a large extent adding value" and that "innovation is a combination of invention and implementation", for which the process consists of such stages as: immersion (submerging in market problems or the current problem), imagination (depicting potential benefits), ideation (concept creation) and initiation (project launch), as well as incubation (testing), industrialisation (production and delivery in large quantities), introduction (trial start and launch of sales), as well as installation and integration (both stages implemented at the client's site) (Deschamps, 2011: 24–25, 32–33).

The presentation of these approaches is finalised with Figure 1, where I graphically present the relationships between the terms of innovations with its five different meanings.



Essential features of innovation: N – novelty; D – relationship with actions and their creations; Z – practical application; C – value, value assignment.

Figure 1. Scopes of the term "innovation" in various approaches to innovation

Source: own elaboration

Among the identified approaches to the generally understood term of innovation, I mostly appreciate the usefulness of the last one, which is symbolically named the NDZC approach. I consider all of them to be multidimensional approaches, the most complete of the ones being analysed (as they refer to the actions of people and their products, and at the same time take into account three other important features of innovation: its novelty, practical application and positive assessment of events-effects). The understanding of innovation, typical for this approach, is consistent with the generally accepted concept of the innovation process and the stages identified within, such as: a creative idea (concept), a creative activity finalised with a new product, implementing the product for the first time, its imitation (reproduction), and dissemination (diffusion, absorption). Moreover, this approach is not only useful in the interpretation of innovation in its general sense but also in the process of extracting and displaying various special kinds of innovation. Criteria for organising innovations can be derived from distinguishing and qualifying criteria of such events as: novelties, actions, products of actions, their applications, and evaluations.

To each of the distinguished approaches, and thus types of innovations, one can assign their different subsets in the EBSCO publication database, depending on the adopted search criteria for a given subset – terms describing it. The number of elements of each of these subsets is shown in Table 1.

Table 1. Frequency of occurrence of publications on distinguished types of innovations in the EBSCO database in the years 1979–2018, in decades

Type of innovation		Decades				In total 1979–2018
Symbol	Features and methods of their expressions	1979–1988	1989–1998	1999–2008	2009–2018	
NDZ	'innovation' and 'implementation'	1	7	15	15	38
	'innovation' and 'commercialisation'	3	14	127	465	609
	'innovation' and 'spread'	–	–	–	5	5
	'innovation' and 'application'	10	380	551	635	1,576
NDZ in total		14	401	689	1,120	2,224
NDC	'innovation' and 'value'	11	37	202	852	1,102
NDZC	'innovation' and 'value' and 'implementation'	–	–	7	36	43
	'innovation' and 'value' and 'commercialisation'	–	–	1	19	20
	'innovation' and 'value' and 'spread'	–	–	2	3	5
	'innovation' and 'value' and 'application'	2	1	21	47	71
NDZC in total		2	1	31	105	139
N = ND	'innovation' in a general sense	2,900	30,247	47,451	88,449	11,467

Essential features of innovation: N – novelty; D – relationship with actions and their creations; Z – practical application; C – value, value assignment; in the examined database, the novelty feature concerns in every case actions of people and their products, hence the conventional notation: N = D.

Source: own elaboration

The data contained in Table 1 show that apart from exceptionally numerous sets of publications concerning innovation in its broadest sense, innovation is relatively often interpreted as a novelty applied in practice (NDZ symbol) – 2,224 publications, less frequently as a novelty from the point of view of its value, value for the entity using it (NDC symbol) – 1,102 publications, and the least frequently as a novelty both valuable and applied in practice (NDZC symbol) – 139 publications.

4. Technical, market and social innovations – distinction criteria – similarities and differences

Technical innovations are mainly concerned with the product of action – its result and effect. Occasionally, this product occurs in the company of the preceding process, other times, it is accompanied by the following process. There are three typical ways of defining technical innovation indicated in which innovation is subsequently interpreted as:

1. A product (or service) significant in technical terms, relativised to the place and time (e.g. the oil lamp invented by Ignacy Łukasiewicz in 1853 in Lviv, or lighting with the oil lamp the Lviv hospital for the first time in 1853).
2. A creative process that reflects the transformation of ideas into a new product, technically significant (e.g. the invention of an oil lamp in 1853).
3. A process that occurs after the appearance of a new product which is technically important i.e. the process of its usage by an external user or the creator himself, inside the company (once again, as an example, the first lighting of the oil lamp in the Lviv hospital in 1853 can be mentioned, or another example: the use of the assembly line by Henry Ford in 1913 at a car plant in Highland Park, Michigan).

Most frequently, when defining technical innovation, the third of the following approaches is used, depicted by the second example which refers to the company. This meaning is assigned to the term of technical innovation in *Leksykon naukowo-techniczny* (Czerni et al., 1984: 307), where it is interpreted as "introducing new technical inventions or improvements to the production practice which allow an increase in the quantity and quality of manufactured goods, as well as an increase in work efficiency and the level of investment". It means that in this case not only the product itself is placed in the production practice of the company but also the process of its usage and the resulting effects positively evaluated by the company.

The term that is similar to technical innovation is the term of technological innovation. This time, however, it is not so much about the product itself, but about a new way, in particular a new method of operation, a method of reaching the goal, often including this product. According to Heiskala (2007: 59), Bukowski and Rudnicki (2014: 79), speaks of technological innovations as new and more effective ways of transforming material reality. The author adds that by using them and taking into consideration their effects, we enter the sphere of economic innovation.

Technological innovations are either in the company (they apply, for example, to the structure of a new technological process), or at the interface of the enterprise and society (e.g.: customer service), or in a specific community (e.g.: students using the Internet at school). Similarly, in the case of technical innovations, the majority of examples presenting technological innovations so far have been mainly concerned with the sphere of production i.e. the interior of the company.

Market innovations, as the term implies, are related to the market, i.e. referring to the view of Wrzosek (1998) regarding the market, they are associated with all relations that take place in the process of exchange between sellers, as entities representing supply, and purchasers, representing demand.

In some cases, when defining these innovations, a special role is attributed to supply, in other cases, to demand. The first case is about push innovations, i.e. innovations "pushed" by supply, science, technology, processes (through technical and technological innovations, as previously mentioned). It is about "marketisation" of new products, and shifting them from the company to the market. The company shapes the product and when it is ready – launches it into the market. In defining push innovations, the emphasis is not put on creating innovation, but on the second phase of the innovation process, which starts with the introduction of innovation to the market.

The definitions in which innovations are simply called "commercialisation of an invention", "commercialisation of a new idea", or "launching a new or improved product" are considered typical for the supply-side approach to innovation. Koch (1997: 89), when defining innovation (in a market approach), calls it the commercialisation of invention, and explains this phrase as the introduction of a new product or service to the market.

The second type of market innovation is called pull innovation, defined as drawn by the market, or by demand. This time, demand turns out to be particularly important in interpreting innovation. Knowing market needs, the company tries to adapt its products – it responds to demand with innovation, and thus demand modifies it. As so understood, demand is an occurrence that begins and ends the process of innovation. This means that clients' contribution in the creation of innovations is in this case much larger than in the case of supply innovations. According to Drucker (1992: 42), nowadays innovation needs to be defined "in terms of demand rather than supply, i.e. as a change in the value and satisfaction of the consumer's needs through the use of specific resources".

In the next approach to innovation, this time defined as supply-demand, the relationship between demand and supply is much stronger than in the so far discussed approaches. The impact of both sides, supply on demand and demand on supply, is realised through innovation. It occurs not only in various areas of the company's operation but also at many stages of the innovation process. Among the theoretical concepts describing this issue, special attention should be paid to the developed concept of the *New Era of Innovation* by Prahalad and Krishnan (2010: 12–13), presented by the authors graphically in the form of an edifice of innovation symbolising the Greek temple. Two columns play an important role. One of them presents demand and supply, defined collectively as "co-creating the experience of personalized customers”, which means that the company shapes consumer expectations and responds to customers’ changing signals. The clients along with the company co-create value, including every single customer (hence the entry N = 1). And this, according to the authors, is the essence of innovation that is achieved thanks to the availability of global human resources (including talents), technology and finance (R = G), i.e. not so much due to owing these resources but by accessing them. The availability of resources is presented as the second column. Business processes that fill the interior of the edifice allow innovations to be achieved, i.e. the realisation of the transition from the idea to action. They are supported, using the authors’ words, with technical and social architecture, located at the base of the building and its top parts.

Social innovations, as the term itself implies, are related to specific communities, larger and smaller ones. More frequently discussed in the literature are social innovations concerning large communities presented at the mega- or macro-scale, that is, society in the broadest meaning of that word (referred to as humanity, global community, world community, or society distinguished by continental or national borders), than social innovations at the microscale (which are usually related to an enterprise).

The main reasons for introducing innovations are economic in nature, although other motives – non-economic ones – can be also pointed out. Among other authors, Kieżun (2011: 163) mentions them when discussing the scientific achievements of Kwiatkowski, with particular emphasis on the idea of innovative society, which is built mainly on the basis of Schumpeter. Innovations can take place in the technical, organisational and social spheres. These kinds of changes lie at the root of innovation (Kieżun, 2011: 164), and although they are mostly related to the production process, they do not have the greatest social and economic resonance, as pointed out by Kwiatkowski and seen in the achievements of Drucker. The innovative changes that he considers to greatly influence social effects are broadly understood spheres of services such as: education, healthcare, politics, and the arts. To the list of mentioned services, Kieżun adds public administration (Kieżun, 2011: 166). This means that these changes can be qualified as innovations in the sphere of so-called consumer and social services. Niedzielski considers consumer services to be the most sensitive in the process of being subject to innovative processes, although, as he points out, every non-material value created in the service process should be socially useful[2]. This statement should be extended to all social innovations which by definition are pro-social – have positive effects and a common weal in mind. One cannot call social innovations, as it is explicitly stated by Bukowski and Rudnicki (2014: 81), those innovations that “have negative effects or serve only business purposes”.

2 This view is expressed by Niedzielski, who is the author of a slogan “service innovations” in: Bąkowski, Głodek, Gołębiowski et al., 2005: 67–70.

Innovations are defined as social mainly due to the assessment of the effects of human actions (therefore, they can be called social innovations ex post) or in the case of intended activities – on the grounds of the objective of the action (social innovation ex ante). If these effects are qualified as social, in other words, these are changes taking place in relations or in social structures, then regardless of whether they are deliberate or unintentional adjustments (side effects of intended actions), they are called social innovations. Such a general definition of social innovations is used by the majority of authors concerned with this issue. Despite being rare, there are definitions that are more detailed. Heiskala (2007: 59, for Bukowski and Rudnicki, 2014: 80) distinguishes three types of changes in social structures: changes in regulatory, normative and cultural structures, which are expressed in new behaviours, new values, and a new interpretation of events. These three types of changes, according to Heiskala, incorporate the scope of the naming convention of social innovations.

Social effects, which constitute the essence of social innovations, may be the result of innovative changes that occur, as it is most commonly defined, in the area of technology. It also happens, although less frequently, that social effects are a consequence of organisational changes, as well as the ones taking place in the area of interpersonal relations. The changes indicated in the last of the mentioned cases can be called "pure social innovations", due to the fact that both the cause and the effect are social changes (in other words: socio-social innovations). In other cases, it is about social innovations only regarding the effect/purpose (e.g.: technical and social innovations, organisational and social innovations). This means that in the case of social innovations, the cause and effect relationship between changes of various types, forming a kind of continuum at the end of which social innovations are usually found, is visible even more than in the case of technical or market innovations.

The publications published in the EBSCO database and concerning the indicated types of innovations, i.e. social and market innovations, including supply and demand market innovations, are presented in numerical form in Table 2.

The data contained in Table 2 show that in the analysed EBSCO database there are almost 10 times more publications on social innovation than market innovations, which is expressed by the numbers 3,270 and 343 respectively. In addition, in a smaller set of publications on market innovations, the supply side is dominant (expressed by 'supply innovation' and 'push innovation') – 45 literary items, compared to 29 publications devoted to demand innovation (expressed by 'demand innovation' and 'pull innovation').

Table 2. Frequency of publications on market and social innovations in the EBSCO database in the years 1979–2018, in decades

Type and expression of innovation	Decades				In total 1979–2018
	1979–1988	1989–1998	1999–2008	2009–2018	
'market innovation' in a general sense	5	29	88	157	269
'push innovation'	–	2	12	21	35
'supply innovation'	–	–	4	6	10
'pull innovation'	–	–	–	1	1
'demand innovation'	1	1	15	11	28
market innovation in total	6	22	119	196	343
'social innovation'	8	10	736	2516	3270

Source: own elaboration

5. Social and market innovations – the past and present

Social innovations have recently been perceived differently than at least for many decades before. The social effect of social innovations is still their essential feature, but there appeared one more characteristic, also as important, namely the form in which these innovative changes are realised. This new form is known as cooperation community, also called network community or social community. It is a special form of cooperation, as Benkler defines it (2008: 76), it is a form of right to access, use and control shared resources. It is, as Ryfkin (2016: 172–174) writes, a new type of social community, a "dispersed community" of autonomous, equal and dispersed individuals, created as a result of the use of modern information technologies, as well as the emergence of the Internet and a network society (Ornarowicz, 2000).

Social innovations are not only the ones described above which have social consequences but also the innovations realised in the form of social cooperation, namely a network community. This means that the scope of social innovation has grown significantly due to the increased role of social innovations implemented in the social form, and thus also social innovations which turn out to be social because of their results and form (in the set calculus nomenclature: their common part). Figure 2 symbolically presents the scope of the term of social innovations as the sum of two sets, i.e. the set of social innovations distinguished due to their result and due to their form, considering relativisation in time.

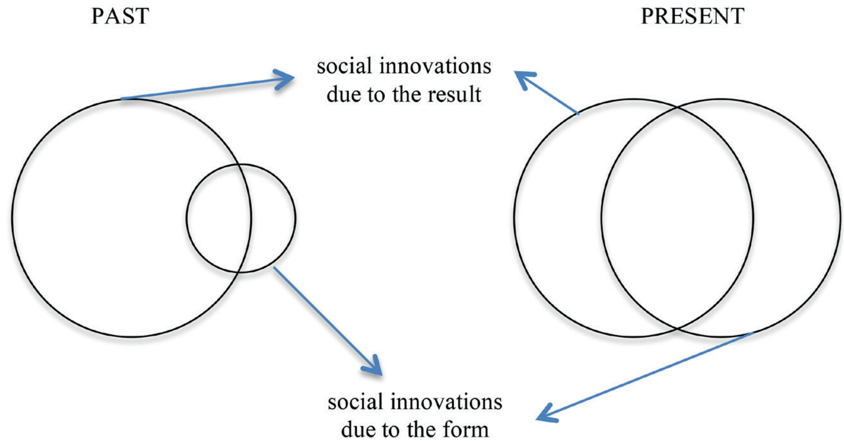


Figure 2. Scope of the term "social innovations" the past and present Distinguishing criteria

Source: own elaboration

Differences in understanding of technical, organisational, and especially market innovations in the past and today can be reduced to a different location, in relation to the enterprise, of such innovation process elements as: reasons for innovation, resources used in the activities, and products of these activities.

Technical and organisational innovations, hitherto usually qualified as "internal" (ideas, resources and products) in the relation to the analysed enterprise, today become more open due to the fact that they are implemented in networks to which this enterprise belongs.

Market innovations of supply type, partially open by definition (internal: ideas, resources, internal-external: products "pushed out of the enterprise", "pushed by technology"), today become even more open: ideas and resources are supported from the outside "by knowledge of social networks" or "social wisdom".

Market innovations of demand type, more open than supply ones (external: needs, expectations, internal: ideas, internal-external: products "pulled by the market", "pulled by demand", internal: resources), today are open even more: internal ideas co-exist with external ones, they are supported from the outside by "knowledge of social networks" and "social wisdom", the implementation of activities takes place inside and outside, as internal and external resources are used.

6. Conclusions

To sum up, the following general conclusions should be formulated:

1. Today, compared to the relatively recent past, greater openness of market innovations takes place, as well as technical and organisational innovations, traditionally considered as closed.
2. Today, much greater socialisation of social innovations is visible, innovations classified as social ones due to the result and due to the form. It is often a "double socialisation" – both because of the result and the form.
3. Differences between market and social innovations, as well as between market, social and technical innovations, are becoming increasingly blurred. This is facilitated by the occurrence of such processes as co-creation, co-dissemination and sharing which make up the new philosophy of management known as the collaborative economy.

Acknolewledgments

The views and opinions expressed in this paper are those of the authors and do not necessarily reflect the views and opinions of the National Bank of Poland. The project entitled "Discussion Forum – Measurement and Evaluation of Economic and Social Phenomena" (MASEP2017) is implemented in cooperation with the National Bank of Poland within the framework of economic education.

References

Baruk J. (2009), *Istota innowacji. Podatność społeczeństw na innowacje*, "Marketing i Rynek" no. 3, pp. 12–18.

Baruk J. (2013), *Innowacja jako czynnik sukcesu organizacji*, "Zarządzanie i Finanse", vol. 11, no. 4, part 1, pp. 7–16.

Benkler Y. (2008), *Bogactwo sieci. Jak produkcja społeczna zmienia rynki i wolność*, Wydawnictwa Akademickie i Profesjonalne, Warszawa.

Birkinshaw J., Hamel G., Mol M.J. (2008), *Management Innovation*, "Academy of Management Review", October 2008, vol. 33, no. 4, pp. 825–845.

Bukowski A., Rudnicki S. (2014), *Innowacje i komercjalizacja wiedzy w naukach społecznych*, [in:] A. Zorska, M. Molęda-Zdziech, B.M. Jung (eds.), *Kreatywność i innowacyjność w erze cyfrowej. Twórcza destrukcja 2*, Oficyna Wydawnicza SGH – Szkoła Główna Handlowa w Warszawie, Warszawa.

Compagni A., Mele V., Ravasi D. (2015), *How Early Implementations Influence Later Adoptions of Innovation: Social Positioning and skill Reproduction in the Diffusion of Robotic Surgery*, "Academy of Management Journal", vol. 58, issue 1, pp. 242–247.

Czerni S., Topulos A., Tabaczkiewicz E., Iwańska J., Pęzińska Z., Bryczkowski P. (eds.) (1984), *Leksykon naukowo-techniczny*, 3rd ed., Wydawnictwa Naukowo-Techniczne, Warszawa.

Damanpour F. (1991), *Organizational Innovation: A Meta-Analysis of Effects of Determinants and Moderators*, "Academy of Management Journal", vol. 34, no. 3, pp. 555–590.

Deschamps J.-P. (2011), *Liderzy innowacyjności. Jak rozwijać i utrzymywać innowacyjność w firmie*, Oficyna a Wolters Kluwer business, Warszawa.

Drucker P.F. (1992), *Innowacje i przedsiębiorczość. Praktyka i zasady*, Państwowe Wydawnictwo Ekonomiczne, Warszawa.

Heiskala R. (2007), *Social innovations: structural and power perspectives*, [in:] T.J. Hämäläinen, R. Heiskala, E. Elgar (eds.), *Social innovations, institutional change and economic performance*, Edward Elgar Publishing, Cheltenham–Northampton, pp. 52–79.

http://eprints.whiterose.ac.uk/114585/2/Schippers-West-Dawson-JOM-def.pdf [accessed: 9.09.2020]

Jia N., Huang K.G., Zhang C.M. (2019), *Public Governance, Corporate Governance, and Firm Innovation: An Examination of State-Owned Enterprises*, "Academy of Management Journal", vol. 62, no. 1, pp. 220–247.

Kieżun W. (2011), *Idea społeczeństwa innowacyjnego S. Kwiatkowskiego*, [in:] W. Kieżun (ed.), *Krytycznie i twórczo o zarządzaniu. Wybrane zagadnienia*, Oficyna a Wolters Kluwer business and Akademia Leona Koźmińskiego, Warszawa.

Koch R. (1997), *Słownik zarządzania i finansów. Narzędzia, terminy, techniki od A do Z*, Wydawnictwo Profesjonalnej Szkoły Biznesu, Kraków.

Kotler Ph. (1994), *Marketing. Analiza, planowanie, wdrażanie i kontrola*, Gebethner i S-ka, Warszawa.

Lampa naftowa (2017), https://pl.wikipedia.org/wiki/Lampa_naftowa [accessed: 7.11.2017].

Ługowik M. (2017), *Historia kołem się toczy*, http://www.dobresobie.pl/historia/311/wynalezienie-kola [accessed: 5.06.2017].

Matusiak K.B. (ed.) (2005), *Innowacje i transfer technologii. Słownik pojęć*, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa.

Mistrzostwa świata w kolarstwie szosowym (2017), https://pl.wikipedia.org/wiki/Mistrzostwa_świata_w_kolarstwie_szosowym [accessed: 17.09.2017].

Ornarowicz U. (2020), *Wspólnoty sieciowe i ich miejsce w ekonomii współpracy, współtorzenia i współużytkowania*, [in:] U. Ornarowicz, P. Płoszajski (eds.), *Społeczna odpowiedzialność biznesu. W poszukiwaniu nowego paradygmatu*, Oficyna Wydawnicza SGH – Szkoła Główna Handlowa w Warszawie, Warszawa [book in print].

Oslo manual. Guidelines for collecting and interpreting innovation data. (2005), 3rd ed., OECD, Eurostat, Paris, http://ec.europa.eu/eurostat/documents/3859598/5889925/OSLO-EN.PDF [accessed: 10.12.2017].

Pierwszy polski samochód wodorowy. Na taki projekt przemysł czekał od dawna (2016), http://www.gazetakrakowska.pl/strefa-biznesu/wiadomosci/a/pierwszy-polski-samochod-wodorowy-na-taki-projekt-przemysl-czekal-od-dawna,10159620/ [accessed: 23.05.2016].

Pietrasiński Z. (1971), *Ogólne i psychologiczne zagadnienia innowacji*, Państwowe Wydawnictwo Naukowe, Warszawa.

Pojazd elektryczny (2017), https://pl.wikipedia.org/wiki/Pojazd_elektryczny [accessed: 24.10.2017].

Popkiewicz M., Malinowski S.P. (2017), *Huragany – skąd się biorą, jakie sieją zniszczenie i jak wpływa na nie zmiana klimatu?*, http://naukaoklimacie.pl/aktualnosci/huragany-skad-sie-biora-jak-sieja-zniszczenie-i-jak-wplywa-na-nie-zmiana-klimatu–245 [accessed: 11.09.2017].

Prahalad C.K., Krishnan M.S. (2010), *Nowa era innowacji*, Wydawnictwa Profesjonalne PWN, Warszawa.

Prezydent Duda otworzył Arenę Lodową. „To jest nowoczesna Polska, to jest Polska ambitna" (2017), https://wpolityce.pl/polityka/371941-prezydent-duda-otworzyl-arene-lodowa-to-jest-nowoczesna-polska-to-jest-polska-ambitna-zdjecia [accessed: 14.12.2017].

Pszczołowski T. (1978), *Mała encyklopedia prakseologii i teorii organizacji*, Zakład Narodowy im. Ossolińskich Wydawnictwo, Wrocław–Warszawa–Kraków–Gdańsk.

Robbins S.P., DeCenzo D.A. (2002), *Podstawy zarządzania*, Polskie Wydawnictwo Ekonomiczne, Warszawa.

Rogers E.M. (2003), *Diffusion of innovations*, 5th ed., Free Press, New York.
Rogers E.M., Shoemaker F.F. (1971), *Communication of innovations: a cross-cultural approach*, Free Press, New York.
Ryfkin J. (2016), *Społeczeństwo zerowych kosztów krańcowych. Internet przedmiotów. Ekonomia współdzielenia. Zmierzch kapitalizmu*, Wydawnictwo Studio Emka, Warszawa.
Schippers M.C., West M.A., Dawson J.F. (2012), *Team reflexivity and innovation: the moderating role of team context*, repository, University of Sheffield, White Rose University Consortium Universities of Leeds, Sheffield–York, pp. 1–42,
Skorupka S., Auderska H., Łępicka Z. (eds.), (1968), *Mały słownik języka polskiego*, Państwowe Wydawnictwo Naukowe, Warszawa.
Słownik języka polskiego PWN (2017), https://sjp.pwn.pl/sjp/innowacja;2561627.html [accessed 10.12.2017].
Sobol E. (ed.) (2003), *Nowy słownik języka polskiego*, Wydawnictwo Naukowe PWN, Warszawa.
Stevens T. (1999), *A modern-day Ben Franklin*, "Industry Week", 1 March, pp. 20–25.
Szymczak M. (ed.) (1988), *Słownik języka polskiego*, vol. 1, Państwowe Wydawnictwo Naukowe, Warszawa.
Utterback J.M. (1971), *The Process of Technological Innovation within the Firm*, "Academy of Management Journal", March, vol. 14, issue 1, pp. 75–88.
West M.A., Farr J.L. (eds.) (1990), *Innovation and creativity at work: psychological and organizational strategies*, John Wiley & Sons, Chichester.
Wright G.H von (1963), *Norm and action: a logical enquiry*, Routledge & Kegan Paul, The Humanities Press, London–New York.
Wrzosek W. (1998), *Funkcjonowanie rynku*, Polskie Wydawnictwo Ekonomiczne, Warszawa.

Innowacje. Aspekt rynkowy i społeczny

Streszczenie: Obszary badań w ekonomii i naukach o zarządzaniu stają się sobie coraz bliższe, coraz mniej rozłączne – przenikają się i upodabniają. W niespotykanym dotąd tempie pojawiają się nowe zdarzenia, nowe wytwory działań ludzi, nowe wzory zachowań, których instytucjonalizacja, szeroko rozumiana jako ich utrwalanie, też przybiera nowe formy. Żyjemy w epoce wszechobecnej innowacyjności. W naturalny sposób rodzą się pytania: Czy innowacje należy dziś rozumieć tak samo jak dawniej?, Czy w ostatnim czasie pojawiły się klasy innowacji o cechach wcześniej niespotykanych?, Czy dotychczasowe definicje innowacji rynkowych i społecznych zachowują swą aktualność?.

Celem artykułu jest przedstawienie zmiany podejścia do innowacji w czasie, ze szczególnym uwzględnieniem ich aspektu rynkowego i społecznego. Autorka próbuje odpowiedzieć na pytania: Jak postęp technologiczny, wyrażający się w usieciowieniu gospodarki informacyjnej, wpłynął na zmianę rozumienia innowacji społecznych?, Jaki wpływ na definiowanie innowacji społecznych ma wzrastająca rola społecznej produkcji i wymiany kosztem wymiany rynkowej?, W jakim stopniu wspólnotowa forma współpracy w przestrzeni wirtualnej jest wyróżnikiem szczególnej klasy innowacji społecznych?.

Przyjęta i stosowana przez autorkę metoda badań sprowadza się do studiów literaturowych nad innowacjami i ekonomią współpracy (dostępu, współdzielenia, współużytkowania) – analizy różnych koncepcji innowacji, w szczególności różnych definicji tej nazwy, różnych podejść do ekonomii współpracy, zestawienia wyników owych analiz i sformułowania wniosków.

Podejście do innowacji zmienia się w czasie – od podejścia technicznego, przez społeczne i rynkowe, do społecznego, ale rozumianego dziś inaczej. Inne jest obecnie kryterium wyróżnienia „społeczności" innowacji. Na rozumienie innowacji wpływa wzrost roli społecznej produkcji i wymiany kosztem wymiany rynkowej. Usieciowienie gospodarki informacyjnej znacznie wzmacnia wymiar społeczny innowacji. Wspólnotowa forma współpracy, z uwzględnieniem współtworzenia dóbr, dostępu do nich, ich współużytkowania i dzielenia się nimi, jest skrajnym przykładem przewagi wymiaru społecznego innowacji nad ich wymiarem rynkowym.

Słowa kluczowe: innowacje społeczne, innowacje rynkowe, innowacje techniczne

JEL: O3, M1

OPEN ACCESS

© by the author, licensee Lodz University – Lodz University Press, Łódź, Poland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license CC-BY (https://creativecommons.org/licenses/by/4.0/)

Received: 2018-01-21; verified: 2019-12-13. Accepted: 2020-10-01

COPE Member since 2018 JM13703

This journal adheres to the COPE's Core Practices https://publicationethics.org/core-practices

Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663
www.czasopisma.uni.lodz.pl/foe/
4(349) 2020
DOI: http://dx.doi.org/10.18778/0208-6018.349.08

Małgorzata Graczyk
Poznań University of Life Sciences, Faculty of Agronomy and Bioengineering
Department of Mathematical and Statistical Methods, malgorzata.graczyk@up.poznan.pl

Bronisław Ceranka
Poznań University of Life Sciences, Faculty of Agronomy and Bioengineering
Department of Mathematical and Statistical Methods, bronislaw.ceranka@up.poznan.pl

New Results Regarding the Construction Method for D-optimal Chemical Balance Weighing Designs

Abstract: We study an experiment in which we determine unknown measurements of p objects in n weighing operations according to the model of the chemical balance weighing design. We determine a design which is D-optimal. For the construction of the D-optimal design, we use the incidence matrices of balance incomplete block designs, balanced bipartite weighing designs and ternary balanced block designs. We give some optimality conditions determining the relationships between the parameters of a D-optimal design and we present a series of parameters of such designs. Based on these parameters, we will be able to set down D-optimal designs in classes in which it was impossible so far.

Keywords: balanced bipartite weighing design, balanced incomplete block design, chemical balance weighing design, D-optimality, ternary balanced block design

JEL: C02, C18, C90

1. Introduction

In this paper, we consider the linear model

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \mathbf{e},$$

where:
$\mathbf{y}$ is an $n \times 1$ random vector of observations,
$\mathbf{X} \in \mathbf{\Phi}_{n\times p}(-1,0,1)$, the class of $n \times p$ matrices $\mathbf{X} = (x_{ij})$ of known elements where x_{ij} equals –1, 0 or 1,
$\mathbf{w}$ is a $p \times 1$ vvector of unknown measurements of objects,
$\mathbf{e}$ is an $n \times 1$ random vector of errors.

We assume that E($\mathbf{e}$) = $\mathbf{0}_n$ and Var($\mathbf{e}$) = $\sigma^2\mathbf{I}_n$, where $\mathbf{0}_n$ is the $n \times 1$ vector with zero elements everywhere, I_n denotes the identity matrix of rank n. Such form of the matrix Var($\mathbf{e}$) indicates that errors are uncorrelated and have the same variance.

In order to estimate w, we use the least squares method and the normal equations of the form $\mathbf{X}'\mathbf{X}\hat{\mathbf{w}} = \mathbf{X}'\mathbf{y}$. Any chemical balance weighing design is singular or non-singular, depending on whether the matrix $\mathbf{X}'\mathbf{X}$ is singular or non-singular, respectively. If $\mathbf{X}$ is of full column rank, the least squares estimator of w is equal to $\hat{\mathbf{w}} = (\mathbf{M})^{-1}\mathbf{X}'\mathbf{y}$ and the covariance matrix of $\hat{\mathbf{w}}$ is given by Var($\hat{\mathbf{w}}$) $= \sigma^2(\mathbf{M})^{-1}$, where $\mathbf{M} = \mathbf{X}'\mathbf{X}$ is called the information matrix for the design $\mathbf{X}$. In the literature, basic problems of weighing designs are discussed. Jacroux, Wong and Masaro (1983), Sathe and Shenoy (1990) gave the introduction to different optimality criteria.

Here, we consider chemical balance weighing designs under the basic assumption that the design is D-optimal. The weighing design is stated by entering its matrix. The design $\mathbf{X}_D$ is called D-optimal in the given class $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(-1,0,1)$ if $\det(\mathbf{X}'_D\mathbf{X}_D) = \max(\det(\mathbf{M}) : \mathbf{X} \in \mathbf{\Phi}_{n\times p}(-1,0,1))$. Moreover, if det($\mathbf{M}$) attains the upper bound, then the design is called regular D-optimal. For more theory, we refer the reader to the papers of Katulska and Smaga (2013), Ceranka and Graczyk (2016).

Based on the results given in Ceranka and Graczyk (2017), we have:
Theorem 1.1. Any chemical balance weighing design $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(-1,0,1)$ with the variance matrix of errors $\sigma^2\mathbf{I}_n$ is regular D-optimal if and only if $\mathbf{X}'\mathbf{X} = m\mathbf{I}_p$, where m is the maximal number of elements different from zero in the j-th column, where $j = 1, \ldots, p$.

The relations between the parameters of the D-optimal chemical balance weighing design imply that for any combination of numbers p and n, we are not able to determine a D-optimal design. In other words, in any class $\mathbf{\Phi}_{n\times p}(-1,0,1)$, a D-optimal chemical balance weighing design may not exist.

Therefore, the aim of this paper is an investigation of a new construction method of a D-optimal chemical balance weighing design. Based on this method, we will be able to set down D-optimal designs in classes in which it was impossible so far. Thus, we can determine estimators of unknown parameters having the smallest possible product of its variances.

We construct the design matrix of the D-optimal chemical balance weighing design by use of incidence matrices of known block designs. Here we take the incidence matrices of the balanced incomplete block design, the balanced bipartite weighing design and the ternary balanced block design. New matrix construction methods will allow us to determine the D-optimal chemical balance weighing design for new combinations of the number of objects and the number of measurements which are not known in the literature. The properties of mentioned designs are presented in Section 2, whereas Section 3 contains the methods of construction of the design matrix. Finally, some examples of experimental plans are given.

2. Balanced block design

In this section, we present the definition and properties of the balanced incomplete block design given in Raghavarao (1971), the balanced bipartite weighing design given in Huang (1976) and the ternary balanced block design given in Billington (1984).

A balanced incomplete block design (BIBD) with the parameters ν, b, r, k, λ is an arrangement of ν treatments into b blocks, each of size k. Each treatment occurs at most once in each block, occurs in exactly r blocks, and every pair of treatments occurs together in exactly λ blocks. Let N be the incidence matrix of a balanced incomplete block design. The parameters are related by the following identities $vr = bk$, $\lambda(v-1) = r(k-1)$, $\mathbf{NN'} = (r-\lambda)\mathbf{I}_v + \lambda\mathbf{1}_v\mathbf{1}_v'$, where $\mathbf{1}_v$ is $v \times 1$ vector of ones.

A balanced bipartite weighing design (BBWD) with the parameters ν, b, r, k_1, k_2, λ_1, λ_2 is an arrangement of ν treatments into b blocks. Each block containing k distinct treatments is divided into 2 subblocks containing k_1 and k_2 treatments, respectively, where $k = k_1 + k_2$. Each treatment appears in r blocks. Every pair of treatments from different subblocks appears together in λ_1 blocks and every pair of treatments from the same subblocks appers together in λ_2 blocks. Let $\mathbf{N}^*$ be the incidence matrix of such a design. The parameters are not independent and they are related by the following equalities

$$vr = bk, b = \frac{\lambda_1 v(v-1)}{2k_1k_2}, \lambda_2 = \frac{\lambda_1\left[k_1(k_1-1)+k_2(k_2-1)\right]}{2k_1k_2}, r = \frac{\lambda_1 k(v-1)}{2k_1k_2},$$

$$\mathbf{N}^*\mathbf{N}^{*\prime} = (r-\lambda_1-\lambda_2)\mathbf{I}_v + (\lambda_1+\lambda_2)\mathbf{1}_v\mathbf{1}_v'$$

A ternary balanced block design (TBBD) with the parameters ν, b, r, k, λ, ρ_1, ρ_2 is an arrangement of ν treatments in b blocks each of size k. Each treatment appears 0, 1, 2 times in a given block, repeated r times. Each of the distinct pairs of treatments occurs λ times. Each element appears once in ρ_1 block and twice in ρ_2 blocks, where ρ_1 and ρ_2 are a known constant for the design. Let **N** be the incidence matrix of a ternary balanced block design. The following relations are satisfied

$$\begin{gathered} vr = bk, r = \rho_1 + 2\rho_2, \lambda(v-1) = \rho_1(k-1) + 2\rho_2(k-2) = r(k-1) - 2\rho_2, \\ \mathbf{NN'} = (\rho_1 + 4\rho_2 - \lambda)\mathbf{I}_v + \lambda \mathbf{1}_v \mathbf{1}_v' = (r + 2\rho_2 - \lambda)\mathbf{I}_v + \lambda \mathbf{1}_v \mathbf{1}_v' \end{gathered}.$$

3. Construction

A large number of publications presenting construction methods of optimal chemical balance weighing designs can be found in the literature. Generally, the construction methods are based on the incidence matrices of known block designs, see Ceranka and Graczyk (2018), Graczyk and Janiszewska (2019). When we determine the design matrix of the D-optimal chemical balance weighing design, then we prepare a plan of an experiment in which we determine unknown measurements of p objects by using n measurement operations.

Let $\mathbf{N}_1$ be the incidence matrix of BIBD with the parameters ν, b_1, r_1, k_1, λ_1. Moreover, let $\mathbf{N}_2^*$ be the incidence matrix of BBWD with the parameters ν, b_2, r_2, k_{12}, k_{22}, λ_{12}, λ_{22}. Based on the matrix $\mathbf{N}_2^*$, we form the matrix $\mathbf{N}_2$ by replacing k_{12} elements equal to +1 in each column which corresponds to the elements belonging to the first subblock by –1. Consequently, each column of $\mathbf{N}_2$ will contain k_{12} elements equal to –1, k_{22} elements equal to 1 and $\nu - k_{12} - k_{12}$ elements equal to 0. Furthermore, let $\mathbf{N}_3$ be the incidence matrix of TBBD with the parameters ν, b_3, r_3, k_3, λ_3, ρ_{13}, ρ_{23}. Then, the design matrix $\mathbf{X} \in \mathbf{\Phi}_{n \times p}(-1,0,1)$ has the form

$$\mathbf{X} = \begin{bmatrix} 2\mathbf{N}_1' - \mathbf{1}_{b_1}\mathbf{1}_v' \\ \mathbf{N}_2' \\ \mathbf{N}_3' - \mathbf{1}_{b_3}\mathbf{1}_v' \end{bmatrix}. \tag{3.1}$$

Each of the $p = \nu$ objects is weighed $m = b_1 + r_2 + b_3 - \rho_{13}$ times in $n = b_1 + b_2 + b_3$ measuring operations.

From Graczyk and Janiszewska (2019), we have:
Lemma 3.1. Any chemical balanced weighing design $\mathbf{X} \in \mathbf{\Phi}_{n \times p}(-1,0,1)$ given in (3.1) is non-singular if and only if $2k_1 \neq v$ or $v \neq k_3$ or $2k_1 \neq k_3$ or $k_{12} \neq k_{22}$.

Theorem 3.1. Any non-singular chemical balance weighing design $\mathbf{X} \in \Phi_{n \times p}(-1,0,1)$ given in (3.1) is regular D-optimal if and only if

$$b_1 - 4(r_1 - \lambda_1) + \lambda_{22} - \lambda_{12} + b_3 - 2r_3 + \lambda_3 = 0. \quad (3.2)$$

In particular, the equality (3.2) is true, when any combination of these parameters which in total gives zero is true. Based on the series of parameters given by Raghavaro (1971), Huang (1976), Billington (1984), and Ceranka and Graczyk (2004a; 2004b) of the block designs presented in Section 2, we formulate the following corollaries.

Corollary 3.1. Let $v = 4s + 1$. The existence of the balanced incomplete block design with the parameters $b_1 = 2(4s+1),\ r_1 = 4s,\ k_1 = 2s,\ \lambda_1 = 2s - 1$ and the balanced bipartite weighing design with the parameters

$$(i)\ b_2 = s(4s+1), r_2 = 8s,\ k_{12} = 2, k_{22} = 6,\ \lambda_{12} = 6,\ \lambda_{22} = 8\ ,$$

$$(ii)\ b_2 = 2s(4s+1), r_2 = 10s,\ k_{12} = 1, k_{22} = 4,\ \lambda_{12} = 6,\ \lambda_{22} = 8$$

and the ternary balanced block design with the parameters

$$b_3 = u(4s+1),\ r_3 = u(4s-t),\ \ k_3 = 4s - t,\ \lambda_3 = u(4s - 2t - 1),$$
$$\rho_{13} = u(4s - t(t+2)),\ \rho_{23} = 0.5tu(t+1),$$
$$t = 1, 2, 3, u, s = 1, 2, \ldots., 4s > t(t+2),$$

$4s + 1$ is a prime or a prime power, implies the existence of the regular D-optimal chemical balance weighing design $\mathbf{X} \in \Phi_{n \times p}(-1,0,1)$ in (3.1) with the variance matrix of errors $\sigma^2 \mathbf{I}_n$.

Corollary 3.2. Let $v = 4(s+1)$. The existence of the balanced incomplete block design with the parameters $b_1 = 2(4s+3),\ r_1 = 4s+3,\ k_1 = 2(s+1),\ \lambda_1 = 2s+1$ and the balanced bipartite weighing design with the parameters

$$b_2 = 4(s+1)(4s+3), r_2 = 7(4s+3),\ k_{12} = 2, k_{22} = 5,\ \lambda_{12} = 20,\ \lambda_{22} = 22$$

and the ternary balanced block design with the parameters

$$b_3 = 4u(s+1),\ r_3 = u(4s - t + 3),\ \ k_3 = 4s - t + 3,\ \lambda_3 = 2u(2s - t + 1),$$
$$\rho_{13} = u(4s - (t-1)(t+3)),\ \rho_{23} = 0.5ut(t+1),$$
$$t = 1, 2, 3, u, s = 1, 2, \ldots., 4s > (t-1)(t+3),$$

$4s + 1$ is a prime or a prime power, implies the existence of the regular D-optimal chemical balance weighing design $\mathbf{X} \in \Phi_{n\times p}(-1,0,1)$ in (3.1) with the variance matrix of errors $\sigma^2\mathbf{I}_n$.

In the special case when $s = t = u = 1$, we obtain the Corollary 3.14 (Graczyk, Janiszewska, 2019).

Corollary 3.3. Let $v = 4s+3$. The existence of the balanced incomplete block design with the parameters $b_1 = 4s+3,\ r_1 = 2s+1,\ k_1 = 2s+1,\ \lambda_1 = s$ and the balanced bipartite weighing design with the parameters

$$b_2 = (2s+1)(4s+3),\ r_2 = 7(2s+1),\ k_{12} = 2,\ k_{22} = 5,\ \lambda_{12} = 10,\ \lambda_{22} = 11$$

and the ternary balanced block design with the parameters

$$(i)\ b_3 = u(4s+3),\ r_3 = u(4s-t+2),\ \ k_3 = 4s-t+2,\ \lambda_3 = u(4s-2t+1),$$
$$\rho_{13} = u\left(4s-t^2-2t+2\right),\ \rho_{23} = 0.5ut(t+1),$$
$$(ii)\ b_3 = 2u(4s+3),\ r_3 = 8u(s+1),\ \ k_3 = 4(s+1),$$
$$\lambda_3 = 2u(4s+5),\ \rho_{13} = 4u(2s+1),\ \rho_{23} = 2u$$

$t = 1,2,3, u, s = 1,2,\ldots,4s > t(t+2), 4s+3$ is a prime or a prime power, implies the existence of the regular D-optimal chemical balance weighing design $\mathbf{X} \in \Phi_{n\times p}(-1,0,1)$, in (3.1) with the variance matrix of errors $\sigma^2\mathbf{I}_n$.

In the special case when $s = t = u = 1$, we obtain the Corollary 3.7 (Graczyk, Janiszewska, 2019), when $s = 3$, $t = u = 1$, we obtain the Corollary 3.31 (ii) (Graczyk, Janiszewska, 2019).

Corollary 3.4. Let $v = 8s+7$. The existence of the balanced incomplete block design with the parameters

$$b_1 = 8s+7,\ r_1 = 4s+3,\ k_1 = 4s+3,\ \lambda_1 = 2s+1$$

and the balanced bipartite weighing design with the parameters

$$b_2 = (8s+7)(4s+3),\ r_2 = 7(4s+3),\ k_{12} = 2,\ k_{22} = 5,\ \lambda_{12} = 10,\ \lambda_{22} = 11$$

and the ternary balanced block design with the parameters

$$(i)\ b_3 = u(8s+7),\ r_3 = u(8s-t+6),\ \ k_3 = 8s-t+6,\ \lambda_3 = u(8s-2t+5),$$
$$\rho_{13} = u\left(8s-t^2-2t+6\right),\ \rho_{23} = 0.5ut(t+1)\quad ,$$

$$(ii)\ b_3 = 2u(8s+7),\ r_3 = 16u(s+1),\ \ k_3 = 8(s+1),\ \lambda_3 = 2u(8s+9),$$
$$\rho_{13} = 4u(4s+3),\ \rho_{23} = 2u$$

where $t = 1,2,3, u = 1,2,\ldots,s = 2,3,\ldots,$ implies the existence of the regular D-optimal chemical balance weighing design $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(-1,0,1)$,in (3.1) with the variance matrix of errors $\sigma^2\mathbf{I}_n$.

In the special case when $s = t = u = 1$, we obtain the Corollary 3.31 (ii) (Graczyk, Janiszewska, 2019).

Corollary 3.5. Let $v = (2s+1)^2$. The existence of the balanced incomplete block design with the parameters $b_1 = 4u(2s+1),\ r_1 = 4su,\ k_1 = s(2s+1),\ \lambda_1 = u(2s-1)$ and the balanced bipartite weighing design with the parameters

$$(i)\ b_2 = s(s+1)(2s+1)^2,\ r_2 = 8s(s+1),\ k_{12} = 2,\ k_{22} = 6,\ \lambda_{12} = 6,\ \lambda_{22} = 8\ ,$$

$$(ii)\ b_2 = 2s(s+1)(2s+1)^2,\ r_2 = 10s(s+1),\ k_{12} = 1,\ k_{22} = 4,\ \lambda_{12} = 4,\ \lambda_{22} = 6\ ,$$

and the ternary balanced block design with the parameters

$$b_3 = 8s(s+1)+t+1,\ r_3 = 8s(s+1)+t+1,\ \ k_3 = (2s+1)^2,\ \lambda_3 = 8s(s+1)+t-1,$$
$$\rho_{13} = t+1,\ \rho_{23} = 4s(s+1),\ 4u \geq 2s+1$$

$u,s,t = 1,2,\ldots,$ implies the existence of the regular D-optimal chemical balance weighing design $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(-1,0,1)$,in (3.1) with the variance matrix of errors $\sigma^2\mathbf{I}_n$.

Corollary 3.6. Let $v = 4s^2$. The existence of the balanced incomplete block design with the parameters

$$(i)\ b_1 = 4s^2,\ r_1 = s(2s-1),\ k_1 = s(2s-1),\ \lambda_1 = s(s-1)\ ,$$

$$(ii)\ b_1 = 4s^2,\ r_1 = s(2s+1),\ k_1 = s(2s+1),\ \lambda_1 = s(s+1),$$

$$(iii)\ b_1 = 4st,\ r_1 = t(2s-1),\ k_1 = s(2s-1),\ \lambda_1 = t(s-1),$$

and the balanced bipartite weighing design with the parameters

$$b_2 = 4s^2(4s^2-1),\ r_2 = 7(4s^2-1),\ k_{12} = 2,\ k_{22} = 5,\ \lambda_{12} = 20,\ \lambda_{22} = 22$$

and the ternary balanced block design with the parameters

$$(i)\ b_3 = 2\left(8s^2+1\right),\ r_3 = 2\left(8s^2+1\right),\ \ k_3 = 4s^2,\ \lambda_3 = 16s^2,$$
$$\rho_{13} = 4\left(2s^2+1\right),\ \rho_{23} = 4s^2-1,$$

$$(ii)\ b_3 = 2\left(8s^2-3\right),\ r_3 = 2\left(8s^2-3\right),\ \ k_3 = 4s^2,\ \lambda_3 = 8\left(2s^2-1\right),$$
$$\rho_{13} = 4\left(2s^2-1\right),\ \rho_{23} = 4s^2-1$$

$$(iii)\ b_3 = 16s^2,\ r_3 = 16s^2,\ \ k_3 = 4s^2,\ \lambda_3 = 2\left(8s^2-1\right),\ \rho_{13} = 2\left(4s^2+1\right),\ \rho_{23} = 4s^2-1,$$

$$(iv)\ b_3 = 4\left(4s^2+1\right),\ r_3 = 4\left(4s^2+1\right),\ \ k_3 = 4s^2,\ \lambda_3 = 2\left(8s^2+1\right),$$
$$\rho_{13} = 2\left(4s^2+3\right),\ \rho_{23} = 4s^2-1$$

$$(v)\ b_3 = 2\left(4s^2-1\right)+u,\ r_3 = 2\left(4s^2-1\right)+u,\ \ k_3 = 4s^2,$$
$$\lambda_3 = 4\left(2s^2-1\right)+u,\ \rho_{13} = u,\ \rho_{23} = 4s^2-1,$$

$$(vi)\ b_3 = 2\left(8s^2-1\right),\ r_3 = 2\left(8s^2-1\right),\ \ k_3 = 4s^2,$$
$$\lambda_3 = 4\left(4s^2-1\right),\ \rho_{13} = 8s^2,\ \rho_{23} = 4s^2-1,$$

where $s = 1, 2, \ldots$, implies the existence of the regular D-optimal chemical balance weighing design $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(-1,0,1)$,in (3.1) with the variance matrix of errors $\sigma^2\mathbf{I}_n$.

Corollary 3.7. Let $v = (2s+1)^2$. The existence of the balanced incomplete block design with the parameters $b_1 = 4u(2s+1),\ r_1 = 4su,\ k_1 = s(2s+1),\ \lambda_1 = u(2s-1)$ and the balanced bipartite weighing design with the parameters

$$(i)\ b_2 = u(s+1)(2s+1)^2, r_2 = 5s(s+1),\ k_{12} = 1, k_{22} = 4,\ \lambda_{12} = 2,\ \lambda_{22} = 3,$$

$$(ii)\ b_2 = 2s(s+1)(2s+1)^2, r_2 = 14s(s+1),\ k_{12} = 2, k_{22} = 5,\ \lambda_{12} = 10,\ \lambda_{22} = 11,$$

and the ternary balanced block design with the parameters

$$(i)\ b_3 = 2(2s+1)^2,\ r_3 = 8s(s+1),\ \ k_3 = 4s(s+1),$$
$$\lambda_3 = 8s^2+8s-3,\ \rho_{13} = 4s(s+1),\ \rho_{23} = 2s(s+1),$$

$$(ii)\ b_3 = (2s+1)^2,\ r_3 = (2s+1)^2,\ \ k_3 = (2s+1)^2,$$
$$\lambda_3 = 2\left(2s^2+2s+1\right),\ \rho_{13} = 1,\ \rho_{23} = 2s^2+2s+1$$

$$(iii)\ b_3 = (2s+1)^2,\ r_3 = 2s^2+2s+3,\ \ k_3 = 2s^2+2s+3,$$
$$\lambda_3 = s^2+s+2,\ \rho_{13} = 2s^2+2s+1,\ \rho_{23} = 2,$$

$4u \ge 2s+1$, $u,s,t = 1,2,\ldots$, implies the existence of the regular D-optimal chemical balance weighing design $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(-1,0,1)$, in (3.1) with the variance matrix of errors $\sigma^2\mathbf{I}_n$.

Corollary 3.8. Let $v = 4s^2$. The existence of the balanced incomplete block design with the parameters

$$(i)\ b_1 = 4s^2,\ r_1 = s(2s-1),\ k_1 = s(2s-1),\ \lambda_1 = s(s-1),$$
$$(ii)\ b_1 = 4s^2,\ r_1 = s(2s+1),\ k_1 = s(2s+1),\ \lambda_1 = s(s+1),$$
$$(iii)\ b_1 = 4st,\ r_1 = t(2s-1),\ k_1 = s(2s-1),\ \lambda_1 = t(s-1),$$

and the balanced bipartite weighing design with the parameters

$$b_2 = 2s^2w(4s^2-1),\ r_2 = 2w(4s^2-1),\ k_{12} = 1,\ k_{22} = 3,\ \lambda_{12} = 3w,\ \lambda_{22} = 3w$$

and the ternary balanced block design with the parameters

$$(i)\ b_3 = 4us^2,\ r_3 = u(4s^2-t-1),\ \ k_3 = 4s^2-t-1,$$
$$\lambda_3 = u(4s^2-2(t+1)),\ \rho_{13} = u(4s^2-(t+1)^2),\ \rho_{23} = 0.5ut(t+1),$$

if $t = 1,2$ then $s = 2,3,\ldots$, if $t = 3$ then $s = 3,4,\ldots$, $u,w = 1,2,\ldots$,

$$(ii)\ b_3 = 8su,\ r_3 = 4u(2s-1),\ \ k_3 = 2s(2s-1),$$
$$\lambda_3 = 8u(s-1),\ \rho_{13} = 2u(s-1),\ \rho_{23} = 2s^2+2s+1,$$
$$(iii)\ b_3 = 16s^2,\ r_3 = 4(4s^2+1),\ \ k_3 = 4s^2+1,$$
$$\lambda_3 = 4(4s^2+3),\ \rho_{13} = 4(4s^2-1),\ \rho_{23} = 4,$$
$$(iv)\ b_3 = 8s^2,\ r_3 = 2(4s^2+1),\ \ k_3 = 4s^2+1,$$
$$\lambda_3 = 2(4s^2+2),\ \rho_{13} = 2(4s^2-1),\ \rho_{23} = 2,$$
$$(v)\ b_3 = 8s^2+u-2,\ r_3 = 8s^2+u-2,\ \ k_3 = 4s^2,$$
$$\lambda_3 = 8s^2+u-4,\ \rho_{13} = u,\ \rho_{23} = 4s^2-1,$$

$s,u,w = 1,2,\ldots$ implies the existence of the regular D-optimal chemical balance weighing design $\mathbf{X} \in \mathbf{\Phi}_{n\times p}(-1,0,1)$ in (3.1) with the variance matrix of errors $\sigma^2\mathbf{I}_n$.

4. Examples

Let us consider an experiment in which we determine unknown measurements of $p = 5$ objects and $n = 30$ measurements. According to the Theorem 3.3, we consider the balanced incomplete block design with the parameters $v = 5$, $b_1 = 10$, $r_1 = 4$, $k_1 = 2$, $\lambda_1 = 1$ and the incidence matrix $\mathbf{N}_1$, the balanced bipartite weighing design with the parameters $v = 5$, $b_2 = r_2 = 5$, $k_{12} = 1$, $k_{22} = 4$, $\lambda_{12} = 2$, $\lambda_{22} = 3$ and the incidence matrix $\mathbf{N}_2^*$, and also the ternary balanced block design with the parameters $v = 5$, $b_3 = 15$, $r_3 = 9$, $k_3 = 3$, $\lambda_3 = 4$, $\rho_{13} = 7$, $\rho_{23} = 1$ and the incidence matrix N_3, where

$$\mathbf{N}_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{N}_2^* = \begin{bmatrix} 1_1 & 1_2 & 1_2 & 1_2 & 1_2 \\ 1_2 & 1_1 & 1_2 & 1_2 & 1_2 \\ 1_2 & 1_2 & 1_1 & 1_2 & 1_2 \\ 1_2 & 1_2 & 1_2 & 1_1 & 1_2 \\ 1_2 & 1_2 & 1_2 & 1_2 & 1_1 \end{bmatrix},$$

$$\mathbf{N}_3 = \begin{bmatrix} 2 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 2 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 2 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 2 & 1 & 1 \\ 0 & 0 & 0 & 2 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}.$$

Here, 1_h denotes the element belonging to the h-th subblock, $h = 1, 2$. Thus, the design matrix of the regular D-optimal chemical balance weighing design $\mathbf{X} \in \Phi_{n \times p}(-1,0,1)$ is given in the form

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 \\ -1 & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & -1 & 1 \\ -1 & -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 \\ 1 & 0 & -1 & -1 & -1 \\ 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & -1 & -1 \\ -1 & 0 & -1 & -1 & 1 \\ 0 & -1 & 0 & -1 & 0 \\ -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 \\ -1 & 1 & 0 & -1 & -1 \\ -1 & 0 & 0 & 0 & -1 \\ 0 & -1 & 0 & -1 & 0 \\ -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 & 0 \\ 0 & -1 & -1 & 1 & -1 \\ -1 & -1 & 1 & 0 & -1 \\ -1 & -1 & 0 & 0 & 0 \end{bmatrix}.$$

5. Discussion

Chemical balanced weighing designs are considered in the literature as experimental plans in the studies in which we determine unknown measurements of p objects in n measurement operations. Determining D-optimal designs, we set down the estimators of parameters with the smallest possible product of variances of the estimator. The design matrix is interpreted as a plan of an experiment and it sets the allocation of objects to particular weighing. From this point of view, the parameters presented in corollaries 3.1–3.8 allow us to construct the incidence matrices of block designs and simultaneously experimental plans with the required properties. Given this interpretation and for different optimality criteria, the application of chemical balance weighing designs in economic research is presented in Banerjee (1975) and Ceranka and Graczyk (2014). The applications of such designs are not limited to only one field of science. In addition, these types of experiments are used in agricultural experimental practice. A detailed description of the applications was given in Ceranka and Katulska (1987) and Graczyk (2013).

It is worth emphasising that other optimality criteria are also considered in the literature. For example, detailed research on A-optimal chemical balance weighing designs is given in Ceranka and Graczyk (2015).

References

Banerjee K.S. (1975), *Weighing Designs for Chemistry, Medicine, Economics, Operations Research, Statistics*, Marcel Dekker Inc., New York.

Billington E.J. (1984), *Balanced n-array designs: a combinatorial survey and some new results*, "Ars Combinatoria", no. 17A, pp. 37–72.

Ceranka B., Graczyk M. (2004a), *Balanced ternary block designs under the certain condition*, "Colloquium Biometricum", no. 34, pp. 63–75.

Ceranka B., Graczyk M. (2004b), *Balanced bipartite weighing designs under the certain condition*, "Colloquium Biometricum", no. 34a, pp. 17–28.

Ceranka B., Graczyk M. (2014), *On certain A-optimal biased spring balance weighing designs*, "Statistics in Transition new series", Spring, vol. 15(2), pp. 317–326.

Ceranka B., Graczyk M. (2015), *Construction method of A-optimal chemical balance weighing designs*, "Acta Universitatis Lodziensis. Folia Oeconomica", vol. 3(314), pp. 43–51, http://doi.org/10.18778/0208-6018.314.06

Ceranka B., Graczyk M. (2016), *New construction of D-optimal weighing design with non-negative correlations of errors*, "Colloquium Biometricum", no. 46, pp. 31–45.

Ceranka B., Graczyk M. (2017), *Some D-optimal chemical balance weighing designs: theory and examples*, "Biometrical Letters", no. 54, pp. 137–154.

Ceranka B., Graczyk M. (2018), *Regular D-optimal weighing design with non-negative correlations of errors constructed from same block designs*, "Colloquium Biometricum", no. 48, pp. 1–17.

Ceranka B., Katulska K. (1987), *Zastosowanie optymalnych sprężynowych układów wagowych*, [w:] *Siedemnaste Colloquium Metodologiczne z Agro-Biometrii*, PAN, Warszawa, pp. 98–108.

Graczyk M. (2013), *Some applications on weighing designs*, "Biometrical Letters", vol. 50(1), pp. 15–26.

Graczyk M., Janiszewska M. (2019), *Remarks about construction methods about D-optimal chemical balance weighing designs*, "Biometrical Letters" [to appear].

Huang Ch. (1976), *Balanced bipartite block design*, "Journal of Combinatorial Theory (A)", no. 21, pp. 20–34.

Jacroux M., Wong C.S., Masaro J.C. (1983), *On the optimality of chemical balance weighing design*, "Journal of Statistical Planning and Inference", no. 8, pp. 213–240.

Katulska K., Smaga Ł. (2013), *A note on D-optimal chemical balance weighing designs and their applications*, "Colloquium Biometricum", no. 43, pp. 37–45.

Raghavarao D. (1971), *Constructions and Combinatorial Problems in designs of Experiments*, John Willey Inc., New York.

Sathe Y.S., Shenoy R.G. (1990), *Construction method for some A- and D-optimal weighing designs when* $N \equiv 3 (mod\ 4)$, "Journal of Statistical Planning and Inference", no. 24, pp. 369–375.

Nowe wyniki dotyczące metody konstrukcji D-optymalnych chemicznych układów wagowych

Streszczenie: W artykule rozważamy doświadczenie, w którym wyznaczamy nieznane miary *p* obiektów przy użyciu *n* operacji pomiarowych zgodnie z modelem chemicznego układu wagowego. Wyznaczamy układ, który spełnia kryterium D-optymalności. Do konstrukcji D-optymalnego układu wykorzystujemy macierze incydencji układów zrównoważonych o blokach niekompletnych, dwudzielne układy blokowe oraz trójkowe zrównoważone układy blokowe. Podajemy pewne warunki optymalności, określające zależności między parametrami D-optymalnego układu i prezentujemy serie parametrów takich układów. Na podstawie tych parametrów będziemy mogli wyznaczyć D-optymalne układy w klasach, w których do tej pory nie było to możliwe.

Słowa kluczowe: dwudzielny układ bloków, układ zrównoważony o blokach niekompletnych, chemiczny układ wagowy, układ D-optymalny, trójkowy zrównoważony układ bloków

JEL: C02, C18, C90

OPEN ACCESS

© by the author, licensee Łódź University – Łódź University Press, Łódź, Poland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license CC-BY (https://creativecommons.org/licenses/by/4.0/)

Received: 2020-04-06; verified: 2020-06-05. Accepted: 2020-11-03

COPE Member since 2018 JM13703

This journal adheres to the COPE's Core Practices https://publicationethics.org/core-practices

REDAKTOR INICJUJĄCY
Agnieszka Kałowska

SKŁAD I ŁAMANIE
Mateusz Poradecki

KOREKTA TECHNICZNA
Elżbieta Rzymkowska

PROJEKT OKŁADKI
Katarzyna Turkowska

Wydrukowano z gotowych materiałów dostarczonych do Wydawnictwa UŁ

Wydane przez Wydawnictwo Uniwersytetu Łódzkiego
Wydanie I. W.09854.20.0.Z

Ark. druk. 8,875

Wydawnictwo Uniwersytetu Łódzkiego
90-131 Łódź, ul. Lindleya 8
www.wydawnictwo.uni.lodz.pl
e-mail: ksiegarnia@uni.lodz.pl
tel. (42) 665 58 63