STATISTICS IN TRANSITION new series, September 2023
Vol. 24, No. 4, pp. 139–151, https://doi.org/10.59170/stattrans-2023-056
Received – 28.11.2022; accepted – 12.05.2023

Analysis of social and economic conditions of microenterprises based on taxonomy methods

Aneta Ptak-Chmielewska[1], **Agnieszka Chłoń-Domińczak**[2]

Abstract

The situation of microenterprises on the market is difficult as they are faced with barriers and competitiveness imposed by larger units. We used data from the REGIOGMINA project to assess the situation of microenterprises. The REGIOGMINA project was implemented by a consortium of Kujawsko–Pomorskie Voivodship, the SGH Warsaw School of Economics and the Nicolaus Copernicus University in Toruń in the scope of the GOSPOSTRATEG initiative. This data set was complemented by data provided by the Local Data Bank of Statistics Poland. The analysis describes the situation in Kujawsko-Pomorskie Voivodship. The information from both sources illustrates the situation of microenterprises at a local level (gmina) in the years 2019–2020. A cluster analysis based on taxonomy methods was performed. The aim of the research is to expand the knowledge and contribute to a better understanding of the social and economic problems that microenterprises are confronted with at a local level. The study covers the period before the COVID-19 pandemic and the first year following its outbreak, which makes it possible to assess the effects that the measures taken against the pandemic had on the situation of microenterprises at a local level.

Key words: microenterprises, taxonomy methods, COVID-19.

1. Introduction

The COVID-19 pandemic brought unprecedented challenges for SMEs around the world, including disruption of their activities and financial situation, as well as other difficulties. SMEs faced demand disruptions related to lockdowns, but also logistical challenges related to the break of global value chains. The severity of the challenges depended also on the type of industry in which SMEs operated. The support provided to SMEs under the COVID Anti-Crisis shield provided some support to these

[1] Warsaw School of Economics, Poland. E-mail: aptak@sgh.waw.pl. ORCID: https://orcid.org/0000-0002-9896-4240.

[2] Warsaw School of Economics, Poland. E-mail: achlon@sgh.waw.pl. ORCID: https://orcid.org/0000-0001-5068-0276.

companies. However, their capacity to weather the crisis, frequently with limited resources, differed from large companies.

During the COVID-19 pandemic, Polish enterprises could benefit from various solutions under the "Anti-crisis Shield", comprising six stages of support, including one directed towards the tourist sector and one focused on selected sectors that were more affected by the lockdown measures. The majority of Polish enterprises applied for this support.

The COVID-19 pandemic was followed by the war in Ukraine and the energy crisis that has further affected the situation of SMEs. Therefore, it is important to provide continuous monitoring of the situation of SMEs not only at the national or regional level, but also at the local level, given the very diverse socioeconomic conditions depending on the locality of enterprises.

In our paper, we analyse the performance of micro-enterprises at the local level in the Kujawsko-Pomorskie region, using the unique set of administrative data collected from the Social Insurance Institution and tax authorities. We assess to what extent characteristics of microenterprises at the local level changed and which factors are the main drivers of the observed heterogeneity in the first two years of the COVID-19 pandemic and resulting lockdown measures. Furthermore, we provide evidence on the potential of the administrative data in monitoring the situation of enterprises. The use of administrative data enables timely and cost-effective collection of the information related to the situation of enterprises and how the enterprises changed over time. The main goal of this paper is to provide an example of the potential design of the monitoring tools that support analysis of the situation of SMEs in the medium and long term that can support evidence-based policy design focused on the development of this sector during and after the observed shocks and crises.

2. Literature review

Small and medium-sized enterprises are the foundation of the European economy. In 2020, slightly more than 21 million micro, small and medium-sized SMEs were active in the EU- 27, accounting for 99.8% of all enterprises in the EU-27 non-financial business sector (NFBS). Of this total, 93% were micro-SMEs. SMEs were generating 53% of the total value added produced by the EU- 27 NFBS, and employed 65% of workers in the NFBS (European Commission, 2021: 1). The pandemic had a major impact on EU-27 SMEs in 2020. Many SMEs faced financial difficulties, mainly resulting from large declines in sales. Other key challenges faced by many SMEs in 2020 included supply disruptions, an upsurge in late payments and operating at a loss. SMEs implemented a wide range of mitigation measures. In particular, SMEs used the different support programmes implemented by national governments, especially to pay

their wages, overcome cash flow issues, and reduce working hours and/or staffing (European Commission, 2021).

The pandemic shock for SMEs was of unprecedented magnitude. As summarised by Juergensen et al. (2020), during the peak of the COVID-19 pandemic, 41% of UK SMEs ceased operations, half of SMEs in Germany expected decline of revenues exceeding 10%, and 70% of SMEs in Italy were directly affected by the crisis. One of the major challenges faced by the sector during the pandemic was financial challenge (Cepel et al., 2020; Juergensen et al., 2020; Zutshi et al., 2021). The liquidity issues and financial constraints observed in the short term translated also into medium- and long-term responses, related, for instance, to upgrading digital infrastructure to enable online sales, and reorganisation of global value chains, which will also require adjustments of SMEs involved in the GVCs (Juergensen et al., 2020). The COVID pandemic also shifted the perception of business risk. Evidence from Czechia and Slovakia shows that while before COVID entrepreneurs perceived personnel risk as the most important, after the pandemic market risk and financial risk was faced by more than a half of SMEs (Cepel et al., 2020).

The role of the SME sector is also important in Poland. Despite the COVID pandemic, the number of microenterprises in 2020 increased by 2.1%, while this was the lowest rate of growth noted in recent years (Statistics Poland, 2021). SMEs benefited from the government Anti-Crisis Shield support. According to the evaluation of the Polish Economy Institute (Dębkowska et al., 2021), one of the main instruments used by SMEs was Anti-Crisis Shield 1.0. In connection with this support, microenterprises received PLN 18.9 billion in support. Regarding the support coordinated by Bank Gospodarstwa Krajowego (BGK), including the de-minimis guarantee, 60% of all beneficiaries were microenterprises. They also received support from the Labour Fund, in the form of low-interest loans (1.9 million claims for PLN 9.3 billion). Last but not least, microentrepreneurs were able to use the levy from social security contributions. In total, 70% of microenterprises used this levy, while 61% claimed stand-by benefit, and a similar percentage used support under the Anti-Crisis Shield.

The broad support received during the COVID-19 pandemic helped preserve the SME sector in 2020. This also applies to the Kujawsko-Pomorskie region. In 2020, there were more than 113 thousand microenterprises in this region, representing 55 companies per 1000 population (which is below the average in Poland of 59). The number of people employed in this sector per 1000 population was 103, compared to 113 in Poland, which is also below the average. Microenterprises in this region generated 3.6% of total revenues in 2020 and 4.0% of wages. The Kujawsko-Pomorskie region also noted one of the highest gross turnover profitability indicators of microenterprises (20.1% compared to 14% in Poland) (Statistics Poland, 2021).

3. Data and research methods

To assess the situation of microenterprises, we used two types of data sources using administrative data. The first one is data from the Social Insurance Institution administrative register. It includes information on enterprises – payers of social security contributions collected in the Kujawsko-Pomorskie region. The second is data from the tax authorities on the revenues and taxes of enterprises that were paid in the same region. This unique database was collected as part of the project REGIOGMINA. The project is implemented by a consortium led by the regional government of Kujawsko-Pomorskie Voivodeship with the SGH Warsaw School of Economics and Nicolaus Copernicus University in Toruń, financed by the National Centre for Research and Development. We also used context data from the Statistics Poland Local Data Bank.

The database compiled from both sources provides information about the situation of microenterprises on the gmina (municipality) level. The analysis covers the Kujawsko-Pomorskie region, which was involved in this innovative project, which, for instance, developed the proposal of the Regional Entrepreneurship Observatory, to monitor the situation of SMEs using administrative data. We focused on data from the years 2019 and 2020. 2019 was the reference year before the Covid-19 pandemic with limited observed turbulence. 2020 was the first year of full lock-down caused by the Covid-19 pandemic. A comparison of changes between those two years in the situation of microenterprises defines the impact of the pandemic on their situation and general economic situation on a local (gmina) level.

In the dataset, we identify the size of the companies based on the number of workers covered by social security contributions. Thus, we treat companies that employ nine or less workers, for whom they paid social security contributions, as microenterprises.

Variables on gmina level that were used in the analysis include average revenues from microenterprises based on tax information, the number of microenterprises per 10 thousand population in the gmina, type of gmina (urban, rural, rural-urban), gmina revenues per capita, gmina debt per capita (information from Ministry of Finance), share of population of working age, and the unemployment rate.

Table 1: Variables used in the analysis

Variable (label)	Description	Source
rev_pc_micro	average revenues - microenterprises (taxes) per capita in million PLN	Tax Office
micro_per10k	number of microenterprises per 10 000 population	Social Insurance Institution

Table 1: Variables used in the analysis (cont.)

Variable (label)	Description	Source
type_region	type of local region (gmina)	Statistics Poland
pop_prod_share	share of population of working age in percent	Statistics Poland
unemp_prod_share	unemployment rate in percent	Statistics Poland
rev_pc	local region revenues per capita in PLN	Statistics Poland
debt_pc	Debt per capita in ths PLN	Ministry of Finance

Source: own work.

For grouping of variables and revealing hidden factors the Factor Analysis method was applied. In this method principal component calculation method was used and varimax rotation. According to the principal component method, it is correlation matrix used its eigenvalues with eigen-vectors to calculate the coefficients for linear combination of variables. Linear combination of the variables and the coefficients values provides the information on hidden factors. The coefficients provide the input from each variable into the factor with a sign informing on the correlation between variable and the factor itself. This information provides the interpretation of hidden factors (Panek and Zwierzchowski, 2013).

For clustering, the k-means method was used. The k-means method is the most frequently used method. Euclidean distance is used as the default distance measure. The number of clusters is determined for the start, and next cluster seeds are chosen at random. Each observation (i=1,..., n) is classified in the group with the nearest cluster seed measured by Euclidean distance. For all clusters (j=1,..., k) the new cluster centres are calculated as the arithmetic mean of all observations belonging to the group. Those steps are repeated until there are no other moves between groups. The error function is calculated at each step – the sum of the quadratic distance intergroup calculated from group centres (Frątczak ed., 2009):

$$\mathrm{F} = \sum_{\mathrm{j}=1}^{\mathrm{k}} \mathrm{d}\left(O_i, M_j\right) \tag{1}$$

where d is Euclidean distance, O_i – is centre for i-th group, M_j – observation j=1,…k.

In practice, this process is convergent after a few iterations, but in general, as this algorithm does not have to be convergent, the maximum number of iterations is pre-defined.

4. Empirical results

An empirical analysis was conducted in three steps. In the first step, the factor analysis made it possible to identify the hidden factors and to verify their impact on

gminas. Next, the gminas were grouped into clusters, and finally in the last step the profile of each cluster was specified and described. Profiles were compared between clusters and between two periods, 2019 and 2020, before and the pandemic periods. Descriptive statistics for variables are presented in Table 2.

Table 2: Descriptive statistics for variables

a) 2019

Variable	Mean	Std. dev.	Min	Max	Q1	Median	Q3
expenses_pc	5358.29	676.89	4054.68	7765.99	4872.14	5254.40	5721.99
pop_prod_share	61.70	1.43	53.87	65.39	61.15	61.77	62.51
unemp_prod_share	3.89	1.39	1.19	8.32	2.93	3.88	4.68
rev_pc	5331.42	534.78	4183.56	7701.28	4961.73	5245.48	5628.84
debt_pc	1338.57	1190.64	0.00	7006.69	566.94	1140.48	1657.17
micro_per10ths	149.08	215.72	0.00	2232.48	95.28	124.33	159.02
rev_pc_micro (mln)	181.40	441.64	6.37	3899.69	48.71	80.74	145.65

b) 2020

Variable	Mean	Std. dev.	Min	Max	Q1	Median	Q3
expenses_pc	5643.41	659.03	4650.36	8568.85	5157.25	5528.00	6064.12
pop_prod_share	61.30	1.49	53.10	64.99	60.67	61.31	62.21
unemp_prod_share	4.34	1.43	1.44	8.46	3.45	4.34	5.26
rev_pc	5881.24	570.61	4764.34	7794.14	5416.90	5780.42	6260.01
debt_pc	5643.41	659.03	4650.36	8568.85	5157.25	5528.00	6064.12
micro_per10ths	96.05	126.70	0.00	1307.86	0.00	107.17	136.15
rev_pc_micro (mln)	172.68	346.22	10.94	3046.78	53.51	82.77	147.75

Source: own calculations in SAS 9.4.

4.1. Factor analysis

Factor analysis is performed using principal components with varimax orthogonal rotation. Factor analysis can be based on the principal component method to find factor weights. Using eigenvalues and eigenvectors, linear combinations of variables are calculated with coefficients driven by eigenvector components. Those linear combinations give the highest possible proportion of variance explained. The first few combinations are used for factors with the highest eigenvalue (highest proportion of explained variance). Orthogonal rotation (for example varimax) is applied to give better understanding and interpretability of results. The high coefficient (factor weight) is

high correlation between the variable and factor. The final combination can be interpreted as a hidden factor based on factor weights.

Factor Analysis was done using the principal component method and varimax rotation. Factors were selected based on minimum eigenvalue criteria above 1, which means variance above average.

Three factors were selected as satisfying the minimum eigenvalue criteria (see Table 2). Based on weights, the resultant factors can be interpreted as:

Factor 1. the factor most correlated with regions' revenues and region debt per capita but debt is significant only for 2020.

Factor 2. the factor strongly determined by two variables: share of population of working age and average revenues of microenterprises based on tax information, but the correlation in two periods, 2019 and 2020, is the opposite.

Factor 3. the factor defined mainly by the number of microenterprises per ten thousand population. Additionally, in 2020, there was a strong correlation with the unemployment rate.

The unemployment rate was significant in defining factors in 2020, the first pandemic year.

Table 3: Factors after Varimax rotation

a) 2019

Specification	Factor 1	Factor 2	Factor 3
pop_prod_share	0.18670	-0.77530	-0.10166
unemp_prod_share	-0.27000	-0.20258	-0.39967
rev_pc	0.93080	0.07174	0.10684
debt_pc	0.53898	0.65328	-0.01519
micro_per10ths	-0.05261	-0.01930	0.93633
rev_pc_micro	0.36139	0.79679	0.07793

b) 2020

Specification	Factor 1	Factor 2	Factor 3
pop_prod_share	0.10220	0.87344	0.12819
unemp_prod_share	-0.16087	0.20864	-0.72188
rev_pc	0.86714	0.00304	0.20070
debt_pc	0.97001	-0.08564	0.10297
micro_per10ths	0.09385	0.12731	0.80837
rev_pc_micro	0.28583	-0.75014	0.26444

Source: own calculations in SAS 9.4.

4.2. Cluster analysis

Clustering was done using the k-means method to group gminas into four clusters. Before clustering, the data were standardised.

In 2019 (Table 3), the last year before the pandemic, we received two clusters counting for 92 and 46 regions (gmina) respectively and two clusters outliers counting for only 4 and 2 regions. The most frequent cluster counting for 92 municipalities is the cluster with the majority of rural gminas. In the subsequent cluster, counting for 46 regions, there was a quite equal proportion of urban, urban-rural and rural municipalities. The cluster with only four regions is the cluster of cities: Bydgoszcz, Grudziądz, Toruń and Włocławek. The outlying cluster with only two regions is the cluster containing the Tuchola urban-rural gmina and Grudziądz urban gminas. The profile of those outlying regions was very different to other clusters.

In 2020 (table 3), the first year of the pandemic with high restrictions on economic and business activity resulted in a very frequent cluster of 121 regions with diversified structure and cluster counting for 18 gminas with a high proportion of rural gminas. Additionally, two outlying clusters were created, counting for four gminas and one gmina. The cluster with only four gminas is the cluster of cities: Bydgoszcz, Grudziądz, Toruń and Włocławek, and this is exactly the same situation as in the previous year, 2019. The outlying cluster with only one gmina is the urban gmina Grudziądz. The profile of those outlying regions was very different to other clusters.

Table 4: Types of regions in k-means clusters

a) 2019

Type of region	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
urban	4	2	11	0	17
urban-rural	0	17	17	1	35
rural	0	73	18	1	92

b) 2020

Type of region	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
urban	13	0	0	4	17
urban-rural	32	0	3	0	35
rural	76	1	15	0	92

Source: own calculations in SAS 9.4.

For comparison of profiles between clusters and between the years 2019 and 2020, the mean values of variables in clusters were calculated and used (Table 4, Figures 1 and 2).

In 2019, the year with the biggest cluster (92 municipalities) with a profile of rural and mixed gminas, the revenues per capita were medium, with low gmina debt per capita. With a medium level of the average number of microenterprises per ten thousand population, those gminas have a very low level of average tax income from microenterprises.

The next cluster is the group of 46 municipalities with a mixed profile. In this group, like the biggest cluster, the revenues per capita were at the average level, with a low level of gmina debt per capita. Those levels were generally lower comparing to the biggest cluster. Having a low number of microenterprises per ten thousand population, those regions also have a low level of average revenues from taxes from microenterprises.

The cluster with four cities only: Bydgoszcz, Grudziądz, Toruń and Włocławek, is the group with lowest level of unemployment. At the same time, those cities have the highest level of revenues per capita but a very high level of debt per capita, almost equal to revenues . In those cities, there are on average only 108.5 microenterprises per ten thousand population, but average revenues from microenterprises are more than ten times higher comparing to other gminas.

The last cluster with only two regions, the Tuchola urban-rural gmina and Grudziądz rural gmina, is outlying, with their profile regions based on ten times the average number of microenterprises per ten thousand population. On the other hand, the revenues from those microenterprises are only two times higher comparing to other dominant gminas (with the exception of the cluster with four cities).

In 2020, a majority of gminas were grouped into one cluster counting for 121 municipalities. This is the cluster with the lowest level of average revenues per capita with almost the same average level of debt per capita. The average number of microenterprises per ten thousand population was only 81 enterprises, which is much lower comparing to 2019 in comparable clusters. At the same time, those regions have the highest average unemployment rate.

The cluster with 18 gminas has a high level of average revenues per capita with almost an equal level of average debt per capita.

The cluster of four cities had the same composition as in 2019: Bydgoszcz, Grudziądz, Toruń and Włocławek. The profile of those cities is almost the same as in the year before the pandemic. Those cities have the lowest unemployment rate, but slightly higher comparing to 2019. Those cities have the highest revenues but also the highest debt per capita, higher even than revenues. In this cluster there are on average 157 microenterprises per ten thousand population, which is 50% more comparing to 2019, but average revenues from those enterprises are lower comparing to the year before the pandemic.

In the last outlying cluster, there is only one rural gmina, Grudziądz, because of a very high average number of enterprises per 10 thousand population.

Changes are visible between the years 2019 and 2020. In 2020, there was higher unemployment. In the biggest cluster, the unemployment rate is 4.46%. For comparison, unemployment for the two biggest clusters in 2019 was on average 3.79% and 4.21%. In 2020, there was a significant drop in the number of microenterprises.

Table 5: Mean values for variables in clusters

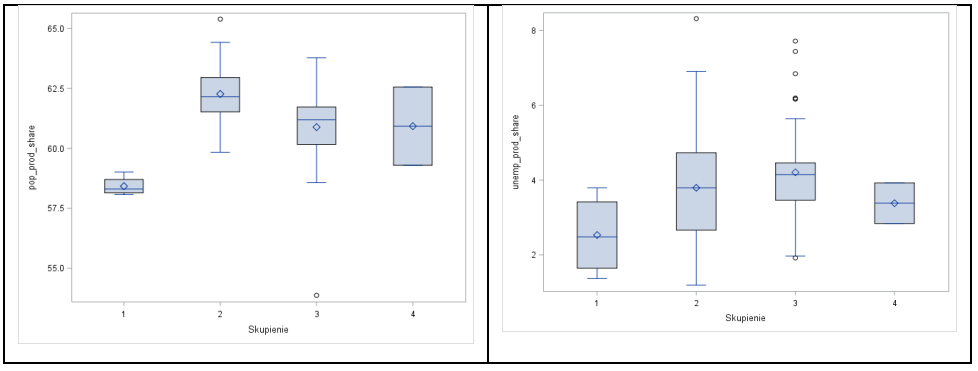
a) 2019

Variable	Cluster 1 (n=4)	Cluster 2 (n=92)	Cluster 3 (n=46)	Cluster 4 (n=2)
pop_prod_share	58.42	62.27	60.89	60.93
unemp_prod_share	2.53	3.79	4.21	3.38
rev_pc	6643.46	5503.33	4877.09	5248.83
debt_pc	6812.92	1282.18	972.23	1409.68
micro_per10ths	108.55	136.39	106.76	1787.08
rev_pc_micro	2397.92	106.29	136.41	238.77

b) 2020

Variable	Cluster 1 (n=121)	Cluster 2 (n=1)	Cluster 3 (n=18)	Cluster 4 (n=4)
pop_prod_share	61.26	62.30	62.28	57.81
unemp_prod_share	4.46	3.55	3.83	3.16
rev_pc	5729.34	6147.93	6634.23	7021.07
debt_pc	5438.61	6109.79	6671.68	7094.96
micro_per10ths	81.14	1307.86	115.41	157.04
rev_pc_micro	118.08	342.95	139.98	1929.02

Source: own calculations in SAS 9.4.



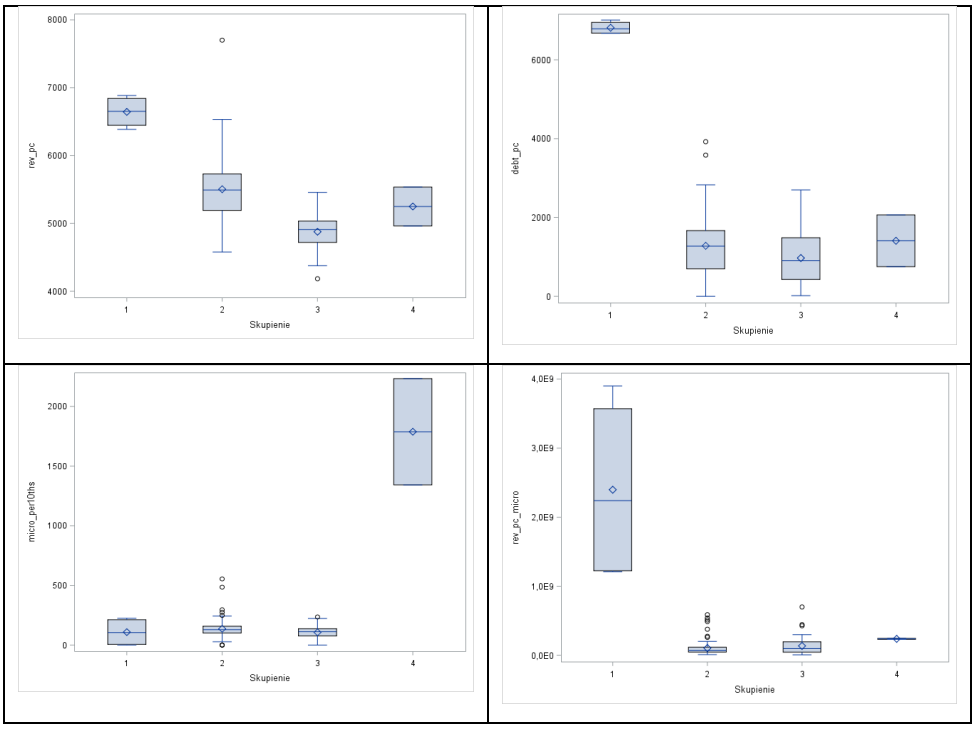
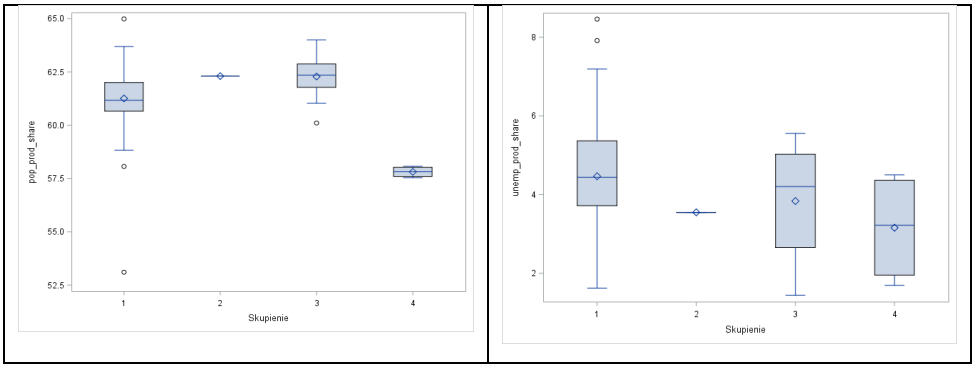


Figure 1: Box-and-whisker plots for variables in clusters – 2019

Source: own calculations in SAS 9.4.



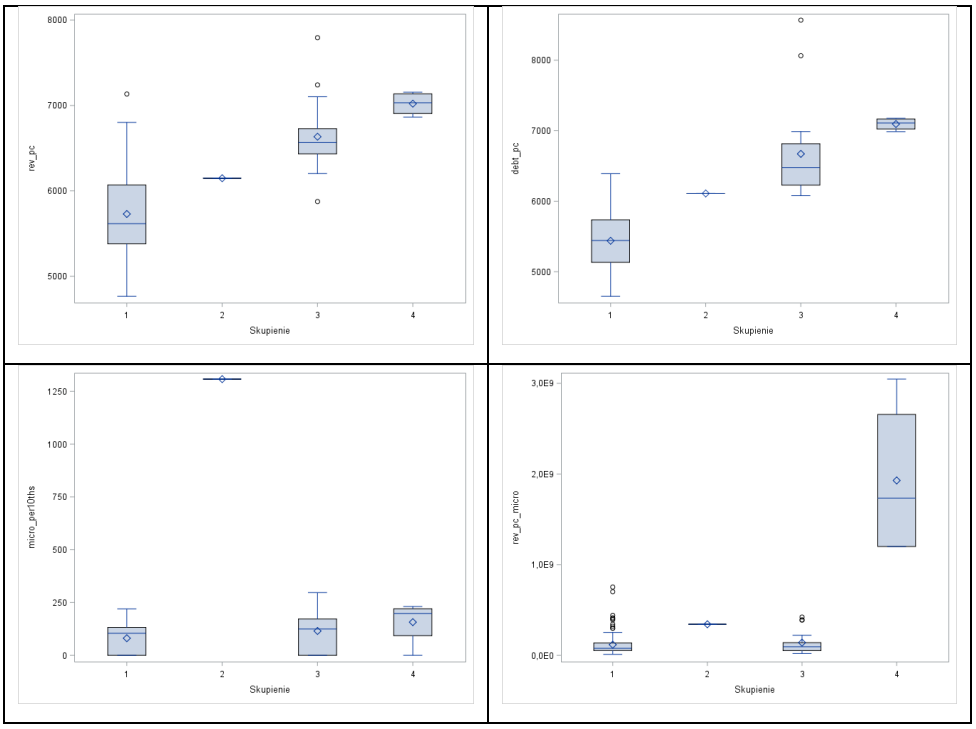


Figure 2: Box-and-whisker plots for variables in clusters – 2020

Source: own calculations in SAS 9.4.

5. Summary and conclusions

In this article, we used a unique data set of the situation of SMEs in the Kujawsko-Pomorskie region to assess the changes of the characteristics of the microenterprise sector at the local level in Poland between 2019 and 2020, that is during the first years of the COVID pandemic. The SME sector in Poland received various types of support from public funds, mainly in the form of stand-by benefits, social security contribution exemptions, and support and guarantees from the Anti-Crisis Shields.

Our analysis shows that there are visible changes in the microenterprise sector and the economic conditions under which they operated. In the largest clusters of gminas, there is a drop in the number of microenterprises per 10 000 population. There is also a significant decline in average revenues reported to tax authorities. This data is consistent with other national statistics, but also observations at the European level, which show that the drop in revenues and financial situation became one of the most important risks faced by the SME sector. We also conclude that the extent to which microenterprises were affected by the pandemic crisis depends on the type of gmina.

Our analysis confirmed that administrative data are a valuable source of information on the situation of SMEs at the local level and potentially a very good source for monitoring the situation of SMEs not only in the Kujawsko-Pomorskie region, but also in other Polish regions.

References

Cepel, M., Gavurova, B., Dvorsky, J. and Belas, J., (2020). *The Impact of the COVID-19 Crisis on the Perception of Business Risk in the SME Segment*, 13, pp. 248–263. https://doi.org/10.14254/2071-8330.2020/13-3/16

Dębkowska, K., Kłosiewicz-Górecka, U., Szymańska, A., Ważniewski, P., Zybertowicz, K., (2021). *Tarcza Antykryzysowa. Koło ratunkowe dla firm i gospodarki*?

European Commission, (2021). Annual Report on European SMEs. *Digitalisation of SMEs* (Issue July).

Frątczak, E. red., (2009). Wielowymiarowa analiza statystyczna. Teoria i przykłady zastosowań z systemem SAS, Warszawa, *Szkoła Główna Handlowa w Warszawie*.

Juergensen, J., Guimón, J. and Narula, R., (2020). European SMEs Amidst the COVID-19 Crisis: Assessing Impact and Policy Responses. *Journal of Industrial and Business Economics*, 47(3), pp. 499–510. https://doi.org/10.1007/s40812-020-00169-4

Panek, T., Zwierzchowski J., (2013). Statystyczne metody wielowymiarowej analizy porównawczej. Teoria i zastosowania. *Oficyna wydawnicza SGH*, Warszawa.

Statistics Poland, (2021). *Activity of Enterprises with up to 9 Persons Employed in 2020*, pp. 23–70.

Zutshi, A., Mendy, J., Sharma, G. D., Thomas, A. and Sarker, T., (2021). From Challenges to Creativity : *Enhancing SMEs' Resilience in the Context of COVID-19*, pp. 1–16.