Czapkiewicz, A., Sobieraj, M., & Brzozowska-Rup, K. (2025). Estimating the shadow economy in enterprises: A new approach. *Journal of International Studies, 18*(2), 270-282. doi:10.14254/2071-8330.2025/18-2/15

Journal
of International
Studies

Centre of
Sociological
Research

Scientific Papers

Estimating the shadow economy in enterprises: A new approach

Anna Czapkiewicz
Faculty of Management, AGH University of Science and Technology, Krakow, Poland;
Statistical Office in Kielce: Centre for Non-observed Economy Studies, Poland
czapkiewicz@agh.edu.pl
ORCID 0000-0002-6144-8381

Małgorzata Sobieraj
Statistical Office in Kielce, Centre for Non-observed Economy Studies, Poland
M.Sobieraj@stat.gov.pl

Katarzyna Brzozowska-Rup
Faculty of Management and Computer Modelling, Kielce University of Technology, Poland;
Statistical Office in Kielce, Centre for Non-observed Economy Studies, Poland
K.Brzozowska-Rup@stat.gov.pl
ORCID 0000-0003-1231-8027

Abstract. This article presents an innovative method for estimating the size of the shadow economy in microenterprises that combines the use of survey data and an econometric model specifically designed for this purpose. The main premise of the proposed approach is that expected gross income is calculated as revenue minus the costs of generating income. The analysis focuses on survey responses regarding the perceptions of satisfactory income and satisfactory revenue, and estimated costs. An econometric model (MUM) was constructed to capture respondents declarations and identify hidden components of income. The method was empirically tested using data from the Statistics Poland (GUS), focusing on microenterprises with 1 to 9 employees. The analysis covered six sectors based on the Polish Classification of Economic Activities (PKD). The results, which are consistent with official shadow economy estimates published by GUS, suggest that the proposed methodology can serve as a valuable tool for the early-stage analysis of the shadow economy in microenterprises.

Received: July, 2024
1st Revision: May, 2025
Accepted: June, 2025

DOI: 10.14254/2071-8330.2025/18-2/15

Keywords: shadow economy, Multiple Ultrastructural Model, survey

JEL Classification: G15, G17

1. INTRODUCTION

According to Eurostat guidelines (ESA 2010), shadow economy refers to economic activities that are legal in nature but not declared to the relevant authorities to evade taxes, regulations, or other associated costs. This article proposes a methodology for estimating the size of the shadow economy in officially registered microenterprises using survey data combined with an econometric model. The proposed approach complements existing indirect estimates produced by Statistics Poland (GUS) by offering a preliminary assessment tool that can be applied already at the stage of survey data collection by the statistical office.

The shadow economy (SE) attracts considerable interest from policymakers and economists due to its various implications. The negative consequences of SE include a reduced tax base, market competition distortions, weakened economic and social institutions, and ultimately, slower economic growth. However, as suggested by Choi and Thum (2003), it can also mitigate disruptions caused by government policies.

Individuals typically engage in the shadow economy when the perceived benefits outweigh the risks and costs of potential sanctions. A seminal study by Allingham and Sandmo (1972) explored the decision-making process behind tax evasion. Their mathematical model, based on expected utility theory, analyzed the conditions under which individuals chose to evade taxes. Schneider and Enste (2000) suggested that the growth of the shadow economy is driven by factors such as weak public institutions, low social trust, and high regulatory burdens as these conditions create an environment where individuals and firms are more likely to engage in informal or undeclared economic activities. High levels of taxation, excessive regulation, and a lack of trust in government institutions often push people to seek alternative ways of conducting business outside the formal economy (Yarovenko et al., 2024; Lyeonov et al. 2025).

Policymakers generally aim to reduce the size of the informal economy, but accurate estimation of its scale remains a significant challenge for both practitioners and statisticians. Methodological constraints and data availability pose substantial difficulties when estimating the size of the shadow economy. National statistical offices, which have access to the most detailed and comprehensive microdata, are responsible for providing estimates of the non-observed economy. However, these estimates are published with significant delays and are costly. As a result, econometric estimates have been developed as alternative means of assessing the shadow economy. No single method provides a definitive estimate of this sector. The literature highlights three main approaches used to measure the shadow economy: direct methods (e.g., surveys and interviews), indirect methods (e.g., macroeconomic analyses), and econometric models that identify the causes and consequences of the shadow economy based on statistical data.

In Poland, Statistics Poland (GUS) plays a crucial role in estimating the size of the shadow economy by using various methodologies. GUS combines data from multiple sources, including macroeconomic indicators, survey responses, and labor market analyses, to calculate estimates of the shadow economy. Despite these efforts, accurately estimating the shadow economy remains challenging due to the incomplete and unregistered nature of activities in this sector. Surveys conducted by GUS help gather information about undeclared employment and informal business activities, but they are subject to errors, non-responses, and biases in reporting. Consequently, there is a need for more advanced estimation methods that can complement the existing approaches used by GUS and provide deeper insights into the shadow economy in microenterprises.

Econometric models serve as essential tools in analysing the shadow economy, providing valuable insights into its structure and determinants. However, each model has its limitations, and the choice of an appropriate approach depends on factors such as data availability, research objectives, and the specific characteristics of the region or sector under study. One of the most widely used econometric models for analysing the shadow economy is the MIMIC (Multiple Indicators, Multiple Causes) model. This model operates under the assumption that the shadow economy is a latent variable, meaning it cannot be directly observed but can be inferred through measurable proxies. These proxies include macroeconomic indicators such as cash demand, labor force participation, and discrepancies between income and expenditure data. Simultaneously, the model incorporates multiple causal factors that influence the size and dynamics of the shadow economy. Key drivers identified in the literature include tax burden, labor market rigidity, excessive regulation, weak institutional quality, and low levels of trust in government institutions.

The MIMIC model has been widely used in empirical research, providing both cross-country comparisons and country-specific analyses of the shadow economy. Its application dates back to the late 1970s and early 1980s, with key contributions from Frey & Weck-Hanneman (1984) and later Giles (1999). Schneider & Enste (2000) conducted foundational research on the shadow economy, exploring various measurement approaches, including the MIMIC model. The study identifies key causes (e.g., tax burden, regulation, corruption) and indicators (e.g., labor force participation, cash usage, GDP growth) commonly used in shadow economy estimation. In subsequent research, Schneider (2005, 2007, 2011) extensively applied and refined the MIMIC model to estimate the size of the shadow economy across different countries. Torgler and Schneider (2007) investigated the relationship between tax morale, institutional quality, and the size of the shadow economy, aiming to understand the willingness of individuals and businesses to comply with tax laws. These studies played a crucial role in establishing the MIMIC model as a benchmark methodology in empirical research.

Despite its advantages, the MIMIC model has limitations, including sensitivity to model specification, potential biases in data sources, and challenges in identifying causal relationships. One of its main limitations is the subjectivity in the selection of causal and indicator variables. Different researchers may choose different sets of variables, leading to varying estimates of the shadow economy. The lack of a standardized methodology for variable selection reduces the comparability of results across studies and poses a challenge for drawing consistent conclusions (Breusch, 2005, 2016; Dell'Anno & Schneider, 2006; Dybka et al., 2019). In addition to the MIMIC model, alternative approaches to estimating the shadow economy include analyzing cash demand. Some studies propose a hybrid methodology that combines these techniques, as seen in the works of Dybka et al. (2019), Schneider (2022), and Torój & Cichocki (2023).

This paper proposes a novel method for supporting the estimation of the shadow economy in microenterprises. Despite access to microdata, such as survey data, several challenges persist. These surveys often suffer from low response rates, inconsistencies, errors, and other issues. Therefore, simply having access to microdata does not necessarily simplify the process of estimating the shadow economy.

The proposed method combines survey results, obtained from specifically designed questions, with an econometric model built for this purpose. This approach aims to capture the actual values reported by respondents, which may differ from those reported to tax authorities. The main premise of the proposed approach is that expected gross income is calculated as revenue minus the costs of generating income. We define the Multiple Ultrastructural Model (MUM), which incorporates variables subject to measurement errors due to the shadow economy (model properties are discussed by Czapkiewicz & Brzozowska-Rup, 2024).

The methodology was empirically tested using survey data collected by the Statistical Office in Kielce, covering microenterprises (with 1 to 9 employees) from across the entire country. The surveys were conducted in accordance with all applicable standards of representativeness and quality control, under the supervision of GUS. As such, the data can be considered nationally representative for this group of enterprises. The sample structure reflects the actual composition of Poland's microenterprise sector: around 74% of the surveyed entities were sole proprietorships (single-person firms), and the remaining 26% were businesses with 2–9 employees. These proportions were maintained in our analysis to ensure structural consistency. The analysis was conducted for six PKD sections: Industry (Section 1), Construction (Section 2), Trade (Section 3), Transport (Section 4), Real Estate and Business Services (Section 5) and other PKD sections (Section 6). It is important to note that when registering a business in Poland, entrepreneurs may indicate up to nine PKD (Polish Classification of Activities) codes. However, for statistical and analytical purposes, classification is based on the predominant activity, as defined by GUS criteria, to which each enterprise is assigned a specific PKD section.

The results obtained using the proposed methodology closely align with the official estimates published by Statistics Poland (GUS), which assesses the shadow economy indirectly using macroeconomic balancing techniques, discrepancies in national accounts, and cross-sectional data from multiple sources. In contrast, our approach enables a preliminary assessment of the shadow economy in microenterprises already at the stage of survey data collection conducted by the statistical office. Furthermore, the findings indicate that the estimated share of the shadow economy in microenterprises (expressed as a percentage of their revenues) is comparable to its overall level in the economy (expressed as a percentage of GDP). This suggests that microenterprises may serve as a meaningful proxy for estimating the scale of the shadow economy in Poland.

2. METHODOLOGY

2.1. The multiple ultrastructural model (MUM)

The size of the shadow economy is difficult to estimate due to the lack of direct observations and respondents' tendency to underreport or overreport sensitive information. As a result, statistical methods that explicitly account for measurement errors and latent variables are essential for producing reliable estimates. In this context, we propose the Multiple Ultrastructural Model (MUM), which assumes that observed variables are subject to measurement errors. This framework allows for the isolation and estimation of latent components—true but unobserved values—including the portion attributable to the shadow economy.

The MUM model assumes that observed variables contain measurement errors and that these errors can be accounted for using a statistical approach.

Let X_i^w , Y_i be observations of variables, whose true values are unknown. We define random variables as follows:

$$X_i^w = s_i^w + \varepsilon_i^w, \quad Y_i = \tilde{Y}_i + \eta_i,$$

where ε_i^w and η_i are measurement errors, and s_i^w and $\tilde{Y}_i$ denote the true values. The indices are defined as $w = 1, \dots, W$ and $i = 1, \dots, N$.

In this model, we assume that the residuals, the explanatory variables, and the dependent variables are normally distributed:

$$X_i^w \sim N(s_i^w,\, \sigma_\varepsilon^w), \quad Y_i \sim N(\tilde{Y}_i,\, \sigma_\eta).$$

Additionally, we assume a linear relationship between the true, unknown variables:

$$\tilde{Y}_i = \gamma_1 s_i^1 + \cdots . + \gamma_W s_i^W .$$

Such a model leads to non-identifiability, i.e. the unknown parameters cannot be uniquely estimated because different parameter values produce the same model output. To address the problem of non-identifiability, we propose a solution inspired by Dolby (1976), which involves replicating measurements of both the explanatory and dependent variables. This approach strengthens the structure of the model by providing additional data points that allow unambiguous estimation of unknown parameters. This makes it easier to estimate parameters that have the desired statistical properties, leading to more reliable and interpretable results.

Let X_{ij}^{w} and Y_{ij} represent observed random variables for $j = 1, \dots, T$, where:

$$X_{ij}^{w} = s_i^{w} + \varepsilon_{ij}^{w}, \quad Y_{ij} = \tilde{Y}_i + \eta_{ij}, \quad w = 1, \dots, W; \; i = 1, \dots, N,$$

where s_i^{w} and $\tilde{Y}_i$ are latent (unobserved) components, and ε_{ij}^{w}, η_{ij} are measurement disturbances. These variables are assumed to follow normal distributions[1], i.e.:

$$X_{ij}^{w} \sim N(s_i^{w}, \sigma_\varepsilon^{w}), \quad Y_{ij} \sim N(\tilde{Y}_i, \sigma_\eta).$$

It is important to note, however, that the true explanatory variables s_i^{w} and the latent dependent variable $\tilde{Y}$ are treated as deterministic but unknown. That is, we do not assume any particular probability distribution for these latent components. Only the observed variables are random due to measurement errors, which are assumed to follow a normal distribution. Therefore, the assumption of normality applies exclusively to the measurement disturbances.

The expected value of Y_{ij} is expressed as a linear combination of the latent variables s_i^{w}:

$$E(Y_{ij}) = \gamma_1 s_i^1 + \dots . + \gamma_W s_i^W . \; . \tag{1}$$

This formulation provides a structured representation of the relationship between the explanatory and dependent variables, accounting for measurement errors.

The model parameters can be estimated using the maximum likelihood method, as outlined by Dolby (1976) and further discussed in Czapkiewicz & Brzozowska-Rup (2024).

Let $X_{.i}^{w}$ and $Y_{.i}$ represent the arithmetic means of the replicated observations. Introducing the auxiliary variable z_i, we define:

$$z_i = Y_{.i} - \gamma_1 X_{.i}^1 - \dots - \gamma_W X_{.i}^W .$$

It can be shown that

$$X_{.i}^{w} - s_i^{w} = \frac{-z_i \gamma_w (\sigma_\varepsilon^{w})^2}{(\sigma_\eta)^2 + \gamma_1 (\sigma_\varepsilon^1)^2 + \dots + \gamma_W (\sigma_\varepsilon^W)^2} = R_i^{w}, \quad w = 1, \dots, W.$$

The parameters $\gamma_1, \dots, \gamma_W$ satisfy the following nonlinear system of equations:

$$\sum_{i=1}^{N} \left(X_{.i}^1 - R_i^1\right)\left(\gamma_1 R_{.i}^1 + \dots + \gamma_W R_{.i}^W + z_i\right) = 0$$

$$\dots$$

$$\sum_{i=1}^{N} \left(X_{.i}^W - R_i^W\right)\left(\gamma_1 R_{.i}^1 + \dots + \gamma_W R_{.i}^W + z_i\right) = 0.$$

We can observe that equation (1) can be rewritten as:

$$E(Y_{ij}) = \gamma_1 s_i^1 + \dots . + \gamma_W s_i^W = \gamma_1 (X_{.i}^1 - R_i^1) + \dots + \gamma_W (X_{.i}^W - R_i^W) =$$ (

$$= (\gamma_1 X_{.i}^1 + \dots + \gamma_W X_{.i}^W) - (\gamma_1 R_i^1 + \dots + \gamma_W R_i^W) = O_i - R_i . \tag{2}$$

[1] One important assumption of the MUM model is the normality of measurement error distributions. While this assumption may be seen as restrictive, it provides analytical clarity and tractability. Since respondents may both under- and over-report values, the normal distribution serves as a neutral and practical approximation. Future research may test the robustness of results under alternative distributional assumptions.

This result implies that the true (unknown) expected value of the random variables $E(Y_{ij})$ can be expressed as the difference between two components O_i and R_i. The term O_i represents part of the expected value related to the observed variables:

$O_i = \gamma_1 X_{.i}^1 + \cdots + \gamma_W X_{.i}^W$,

while

$$R_i = \gamma_1 R_i^1 + \cdots + \gamma_W R_i^W$$

corresponds to the **unobservable** part of the expected value. This unobservable part can be interpreted as the "shadow economy". The ratio $\frac{|R_i|}{\gamma_1 s_i^1 + \cdots + \gamma_W s_i^W}$ represents the percentage of hidden values in the system.

2.2. New approach for estimating the size of the shadow economy in enterprises

The proposed method combines survey results, obtained from specifically designed questions, with an econometric model built for this purpose. This approach aims to capture the actual values reported by respondents, which may differ from those reported to tax authorities. We define the Multiple Ultrastructural Model (MUM) incorporating variables subject to errors due to the hidden economy.

Survey-based studies often encounter challenges such as inconsistent responses and low response rates. To address these issues, the econometric model plays a crucial role in structuring and organizing the responses provided by participants. The discussed method integrates survey data with the MUM model to create a robust estimation framework. The dataset was derived from structured surveys conducted among microenterprises in Poland. Respondents were asked to provide information regarding their revenue, costs, and the level of income and revenue they considered satisfactory. To enhance the reliability of responses, a filtering process was applied to exclude incomplete or inconsistent entries.

The analysis starts with the assumption that, **within a given industry, companies with a similar number of employees have comparable average levels of income, costs, and revenue**.

To summarize, in the proposed approach, the MUM model is constructed for the survey data in accordance with the assumptions outlined in the previous chapter as follows:

$$\begin{aligned} X_{ij}^1 &= s_i^1 + \varepsilon_{ij}^1 \\ X_{ij}^2 &= s_i^2 + \varepsilon_{ij}^2 \\ Y_{ij} &= \gamma_1 s_i^1 + \gamma_2 s_i^2 + \eta_{ij} \end{aligned} \tag{3}$$

where

X_{ij}^1 – represents the owner's income (survey data);

X_{ij}^2 – represents the cost of generating income (including wages for employees)

Y_{ij} – represents the company's revenue,

$i -$ is the number of employees in the company,

j – is the survey number, $j = 1, \ldots, T$.

The algorithm for estimating the size of the shadow economy is as follows. The analysis is conducted for a specific industry (s), differentiating based on the number of employees. Let k_i^s represent the number of survey responses from industry s for companies with i employees. Initially, T is set to determine the sample size for the replications. T surveys are drawn from entities within the same industry and with the same number of employees. Since the number of returned surveys varies across entities with different characteristics, the simulation is repeated K times. For iteration k, $k = 1, \ldots, K$, and for company with i employees the following steps are carried out:

1. **Random Selection of Surveys**: T surveys were randomly selected, providing data on $X_{ij}^{1,k}$, $X_{ij}^{2,k}$, Y_{ij}^{k}.
2. **Estimation of MUM Model Parameters:** The parameters of the MUM model, γ_1^k and γ_2^k are estimated.
3. **Calculation of the Unobservable Component:** The unobservable component of the expected value is computed as:

$R_i^k = |\gamma_1^k R_{1,\mathrm{i}}^k + \gamma_2^k R_2^k|$.

4. **Computation of the Shadow Economy Share:** The ratio representing the percentage share of unobservable revenue in the total expected revenue is calculated as:

$$sz_i^k = \frac{|R_i^k|}{\gamma_1^k s_{1,\mathrm{i}}^k + \gamma_2^k s_2^k} * 100\%$$

This value represents the share of unobservable revenue in relation to the total expected revenue for the company.

Finally, with K bootstrap samples, we can calculate not only the average value of the shadow economy but also construct confidence intervals for this parameter. The level of the shadow economy for companies with i employees is calculated using the following formula: formula:

$$SE_{s,i} = \frac{1}{K}\sum_{k=1}^{K} sz_i^k. \tag{4}$$

Additionally, confidence intervals for the estimated parameter can be determined based on the bootstrap samples.

The level of the shadow economy in a given industry s is determined as a weighted average:

$$SE_s = \left(w_1^s SE_{s,1} + \cdots w_9^s SE_{s,N}\right) * 100\% \tag{5}$$

The weights w_i^s correspond to the share of companies with specific characteristics within the total number of enterprises in the industry. The weights w_i^s can be assumed to be $k_i^s / \sum_{i=1}^{n} k_i^s$, as they ensure the representativeness of the study results.

3. EMPIRICAL RESULTS AND DISCUSSION

3.1. Data

This section presents the empirical findings derived from the application of the proposed methodology. Research suggests that microenterprises with fewer than 10 employees exhibit the highest levels of shadow economy activity (Bednarski, 2019). Consequently, the validation of the proposed methodology was conducted using data from this category of enterprises.

The dataset, sourced from the Statistical Office in Kielce, includes survey responses assessing revenue, income, estimated costs, and perceptions of satisfactory income and revenue.

A total of 13,000 surveys were analyzed, focusing on microenterprises with 1 to 9 employees. The surveys covered six sections of the Polish Classification of Activities (PKD). The analysis specifically targeted microenterprises within six PKD sections, offering valuable insights into the scale and scope of the hidden economy across different sectors. To estimate the shadow economy in microenterprises, it was assumed that respondents (business owners) provided answers reflecting their actual financial situation rather than figures reported for tax purposes. However, the reliability of these responses was uncertain, as some answers appeared to be random or influenced by misunderstandings of the survey questions, leading to significant bias in the results.

Figure 1 illustrates the variation in income-related responses across four sectors for microenterprises employing between 1 and 9 workers. The box plot analysis shows that, within each sector, the median income level remains relatively stable regardless of the number of employees. A considerable number of outliers suggest random or inconsistent responses. Additionally, a substantial portion of the surveys came from single-owner microenterprises, with their share ranging from 40% in Section 2 to 66% in Section 3. This uneven data distribution poses challenges for analysis.

Therefore, given the inconsistencies and gaps in the data, the procedure described in the previous section was applied. The goal was to structure the data into a coherent framework using the MUM model. For comparison, a direct calculation method was also employed, where the difference between reported revenue and costs was directly compared to declared income

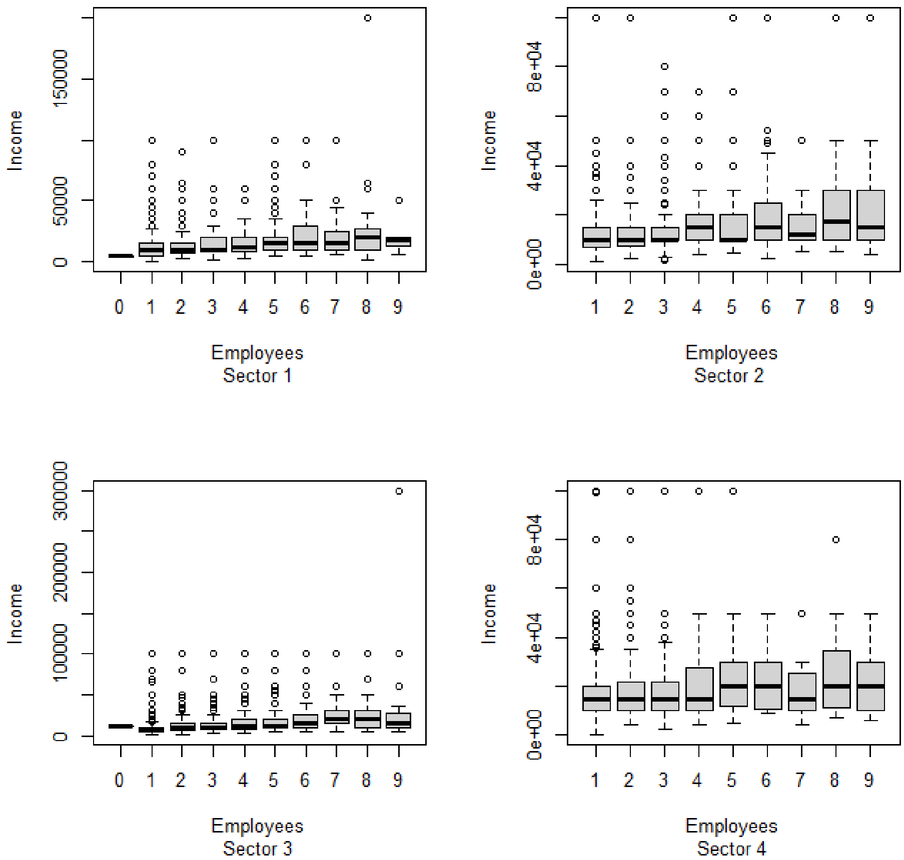

Figure 1. Income distributions (box plots) for all microenterprises, grouped by number of employees, in Sectors 1 through 4

3.2. Empirical study

The direct method, which calculates the hidden economy as the ratio of the difference between reported revenue and income (including operating costs) to total revenue, can be used as a preliminary estimate. For each sector, we computed: $L_{ij} = Y_{ij} - X_{ij}^1 - X_{ij}^2$ and define $SE_s = \frac{L_{ij}}{Y_{ij}}$.

The results are presented in Table1.

Table 1

The level of the hidden economy in microenterprises with up to 9 employees estimated by the 'direct' method

	PKD Section					
	Section 1	Section 2	Section 3	Section 4	Section 5	Section 6
	Percentage %					
SE_s	26.22	29.96	31.24	27.10	27.76	22.59

The hidden economy, estimated directly, exhibits significant variation across different sectors. The estimated levels range from 22.59% in Section 6 to 31.24% in Section 3. Although the direct method offers a useful initial estimate of the size of the shadow economy in each sector, it relies on self-reported data, which is susceptible to underreporting or overreporting. While Statistics Poland (GUS) implements rigorous verification and consistency checks to improve data quality, these limitations necessitate the development of additional approaches. To address these challenges, we propose using an econometric framework designed to explicitly account for measurement errors and latent variables. This offers a more robust and nuanced estimation of the shadow economy in microenterprises.

The size of the hidden economy was then analyzed using the MUM model and the procedure outlined in the previous section. The estimation was performed using bootstrap sampling with a fixed sample size of T = 100. In cases where the number of available surveys in certain microenterprise groups was less than T, resampling with replacement was applied. After K = 1000 iterations, the average estimated size of the shadow economy for each of the six PKD sections was obtained, along with corresponding confidence intervals. The bootstrap procedure used in this study is consistent with the assumptions of the Multiple Ultrastructural Model (MUM). It relies on empirical resampling from the observed data and does not alter the underlying distributional properties or the model's assumed error structure. By generating replicates that preserve these characteristics, the procedure strengthens the validity of statistical inference and supports the reliability of parameter estimation. To evaluate the robustness of the results, the procedure was also carried out using alternative sample sizes, which yielded comparable outcomes. This confirms the stability of the estimates and the soundness of the adopted sampling approach.

Considering that one-person microenterprises accounted for 74% of all microenterprises, while those with 2–9 employees made up only 26%, the contribution of single-person microenterprises to the shadow economy is particularly significant. Therefore, results for microenterprises are presented separately in Table 2, where SE denotes the average of all bootstrap samples according to equation (4), while L_{s1} and P_{s1} represent the 95% confidence intervals for $SE_{s,1}$.

Table 2

The level of the hidden economy in microenterprises with one employee estimated by the novel procedure

	PKD Section					
	Section 1	Section 2	Section 3	Section 4	Section 5	Section 6
	Percentage %					
$SE_{s,1}$	10.39	10.05	10.65	8.96	12.45	4.74
L_{s1}	1.53	1.60	1.17	0.90	4.31	0.06
P_{s1}	23.92	22.98	29.95	22.07	23.31	11.88

Examining the results presented in Table 2, we observe a consistent level of shadow economy activity among one-person enterprises, about 10.05-10.65 percent, suggesting that they constitute a significant share of the shadow economy. However, it is crucial to acknowledge the wide confidence intervals associated with these estimates[2]. These variations arise from the characteristics of the survey data and the presence of outliers, as depicted in Figure 1. Furthermore, the analysis encompasses all microenterprises across Poland, a group that is not homogeneous in terms of income levels. Narrowing these confidence intervals may be possible through a more region-specific analysis.

To estimate the shadow economy within i-employees entities in a given sector, we utilized survey data and the properties of the MUM model. Extending this estimate to the entire microenterprise sector requires appropriate weighting. The most natural approach is to use the proportion of specific units within each section as weights. The use of weights that ensure representativeness also allows the shadow economy in a given section to be estimated.

To evaluate the effectiveness of the proposed methodology, we used formula (5), where the proportion of entities with i employees in a given section served as the weight. The results, presented in Table 3, differ significantly from those in Table 1, where income was directly calculated as the difference between revenues and costs. This discrepancy arises because the direct method is highly sensitive to outliers, whereas the proposed weighted approach provides a more robust estimation.

Table 3

The level of the hidden economy in each Section of microenterprises with 1-9 employees estimated by the novel procedure

	PKD Section					
	Section 1	Section 2	Section 3	Section 4	Section 5	Section 6
	Percentage %					
SE_s	8.08	8.90	8.11	7.57	10.88	4.39
L	2.49	2.70	2.26	2.22	4.71	0.09
P	17.29	17.21	17.81	15.96	20.52	10.35

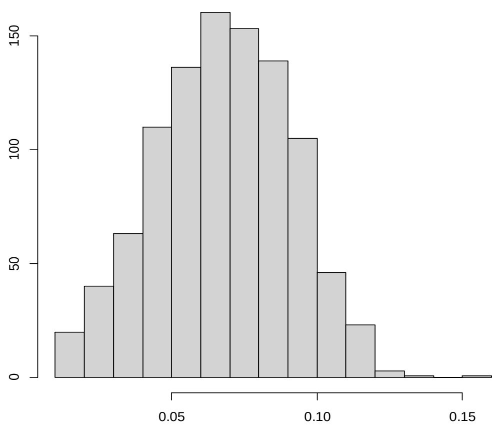

Figure 2. Histogram of the SE estimates calculated for Section 1 using the bootstrap method. The bootstrap samples were derived from the survey data

[2] Similar confidence intervals were observed in the study by Cichocki & Torój (2023) using different methodology.

A comparison of the results in Table 3 with those in Table 2 shows that estimates of the shadow economy are lower for microenterprises (1-9 persons employed) than for one-person enterprises. A possible explanation is that in microenterprises where an accountant is employed, it is often the accountant who completes the survey rather than the entrepreneur. Since the survey assumes the employer as the respondent, this discrepancy may influence the results.

To validate the model's performance, our estimates were compared with data published by Statistics Poland for 2022. For this purpose, the sectoral results from Table 3 were multiplied by the total revenue of microenterprises in each sector, and the resulting values were then normalized by the gross domestic product (GDP) for 2022. The overall contribution of the shadow economy generated by microenterprises was calculated using the following formula:

$$SE = \sum_{s=1}^{6} A_s SE_s \qquad (6)$$

where SE_s denotes the estimated share of the shadow economy in sector s and A_s represents the share of sector s's microenterprise revenue in total microenterprise revenue. The final result was then expressed as a share of GDP. The results are presented in Table 4, with the last row showing the corresponding shadow economy figures reported by Statistics Poland for 2022[3].

Table 4

The level of the hidden economy in each Section of microenterprises with 1-9 employees expressed as PKB

	PKD Section						SE
	Section 1	Section 2	Section 3	Section 4	Section 5	Section 6	
	Percentage %						
SE_s	1.03	1.35	3.94	0.59	0.90	0.33	8.13
L_s	0.32	0.41	1.10	0.17	0.39	0.01	2.40
P_s	2.20	2.92	7.12	1.31	1.70	0.84	17.61
SE GUS	1.10	1.70	3.00	0.80	1.00	1.00	8.90

Taking into account all stages of the analysis, we observe that the estimated share of the shadow economy in the microenterprise sector, as calculated using the proposed method, is generally consistent with the results published by Statistics Poland (GUS), which relies on its own comprehensive and centrally verified methodology. The fact that our estimates for microenterprises are generally lower than the official figures reported by GUS for all enterprises suggests that the proposed method produces realistic and coherent results. The exception is Sector 3, where our estimate slightly exceeds the official figure. This similarity of results demonstrates that the new method provides a reliable estimate of the shadow economy in microenterprises.. Moreover, the proposed method, based on preliminary analysis of survey data, provides an early indication of the shadow economy level in microenterprises—much earlier than official estimates produced by Statistics Poland (GUS) using more stringent methodologies. This approach offers valuable preliminary insights that can complement and support further, more comprehensive assessments. It is also worth noting that this method focuses specifically on microenterprises with 1–9 employees, a segment less frequently reported separately by GUS.

[3] GUS figures refer to the entire enterprise sector, not specifically to microenterprises.

4. CONCLUSION

This article presents a methodology for estimating the shadow economy in microenterprises by combining survey data with an econometric model specifically designed for this purpose. Given the challenges associated with survey data, such as outliers, incompleteness, and random responses, a survey sampling procedure was introduced. The results obtained using the proposed method are comparable with those published by Statistics Poland (GUS), demonstrating the methodology's potential as a valuable tool for statistical offices responsible for shadow economy estimation. While official GUS estimates are based on a comprehensive and methodologically rigorous framework, they are usually only available after a considerable delay. In contrast, the proposed approach enables a preliminary estimation of the shadow economy at an earlier stage, during the survey data collection process, providing timely insights that can complement later official assessments.

During the development of this estimation method, several challenges were encountered from both survey and modeling perspectives. Random responses, misinterpretations of survey questions, and non-responses were likely due to shortcomings in survey design. To address these issues, the questionnaire will be revised and enhanced with additional items aimed at assessing the reliability of respondents' answers. On the modeling side, refinements are also planned to better align with real-world conditions. For example, one key challenge involved the assumption of equal sample sizes across subgroups with differing numbers of observations - a simplification necessitated by the statistical characteristics of the data. Future versions of the model will incorporate adjustments to account for such disparities.

Overall, the proposed methodology shows promising potential and provides a practical framework for the early-stage estimation of the shadow economy in microenterprises.

REFERENCES

Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1(3-4), 323-338.

Adair, P. (2017). Non-Observed Economy vs. the Shadow Economy in the EU: The Accuracy of Measurement Methods and Estimates Revisited. 4th OBEGEF Interdisciplinary Insights on Fraud and Corruption, Porto, Portugal.

Adair, P. (2021). Non-Observed Economy vs. Shadow Economy and Informal Employment in Poland: A Range of Mismatching Estimates. In W. Andreff (Ed.), Comparative Economic Studies in Europe. *Studies in Economic Transition* (pp. 101-115). Palgrave Macmillan, Cham.

Bednarski, M. (2019). Społeczne uwarunkowania funkcjonowania szarej strefy w sektorze mikroprzedsiębiorstw i małych przedsiębiorstw z perspektywy lokalnej. Wnioski z badań empirycznych. *Problemy Polityki Społecznej*, 44, 87-100. https://doi.org/10.31971/16401808.44.1.2019.87100

Buehn, A., & Schneider, F. (2012). Shadow economies around the world: Novel insights, accepted knowledge, and new estimates. *International Tax and Public Finance*, 19(1), 139-171.

Breusch, T. (2005). Estimating the underground economy using MIMIC models. Working Paper. Available at: econwpa.wustl.edu/eps/em/papers/0507/0507003.pdf.

Breusch, T. (2016). Estimating the underground economy using MIMIC models. *Journal of Tax Administration*, 2(1), 1-29.

Cagan, P. (1958). The demand for currency relative to the total money supply. *Journal of Political Economy*, 66(4), 302-328.

Cichocki, S., & Torój, A. (2023). Estimating the size of informal economy in a post-transition country – the case of Poland. *Baltic Journal of Economics*, 23(1), 91-116. https://doi.org/10.1080/1406099X.2023.2228589.

Czapkiewicz, A., & Brzozowska-Rup, K. (2024). The Measurement of the Gross Domestic Product Affected by the Shadow Economy. *Statistics in Transition New Series*, 25(3). https://doi.org/10.59170/stattrans-2024-028.

Dell'Anno, R. (2007). The shadow economy in Portugal: An analysis with the MIMIC approach. *Journal of Applied Economics*, 10(2), 253-277.

Dell'Anno, R. (2016). Analyzing the determinants of the shadow economy with a "separate approach": The MIMIC model and the currency demand approach. *Journal of Economic Studies*, 43(4), 538-556.

Dell'Anno, R. (2022). The shadow economy: Measurement methods and empirical findings. Springer.

Dell'Anno, R., & Schneider, F. (2006). Estimating the underground economy using MIMIC models: A response to T. Breusch's critique. Working Paper, No. 0607. Johannes Kepler University of Linz, Department of Economics.

Dolby, G. R. (1976). The Ultrastructural Relation: A Synthesis of the Functional and Structural Relations. *Biometrika*, 63, 39--50.

Dybka, P., Kowalczuk, M., & Wesołowski, G. (2023). An assessment of the shadow economy in Central and Eastern Europe using the MIMIC approach. *Applied Economics Letters*, 30(4), 1225-1237.

Elgin, C., & Erturk, F. (2019). Informal economies around the world: Measures, determinants, and consequences. *Euroasian Economic Review*, 9(2), 221-237. https://doi.org/10.1007/s40822-018-0105-5.

Feld, L. P., & Schneider, F. (2010). Survey on the shadow economy and tax evasion. *IZA Discussion Paper* No. 4556.

Frey, B. S., & Weck-Hanneman, H. (1984). The Hidden Economy as an "Unobserved" Variable. *European Economic Review*, 26(1), 33-53.

Gërxhani, K., & Cichocki, S. (2023). Formal and informal institutions: Understanding the shadow economy in transition countries. *Journal of Institutional Economics*, 19, 656-672. https://doi.org/10.1017/S1744137422000522.

Giles, D. E. (1999). Modelling the hidden economy and the tax-gap in New Zealand: An econometric analysis. *Empirical Economics*, 24(4), 621-640.

Hassan, M., & Schneider, F. (2016). Size and development of the shadow economies of 157 countries worldwide: Updated and new measures from 1999 to 2013. *IZA Discussion Paper* No. 10281.

Lyeonov, S., Moroz, A., Dudzik, I., & Chuluunbaatar, E. (2025). Green rules & grey markets: Do environmental policies influence the informal economy?. *Economics and Sociology*, 18(1), 313-338. https://doi.org/10.14254/2071-789X.2025/18-1/17

Lewis, W. A. (1955). The theory of economic growth. London: Allen and Unwin.

Medina, L., & Schneider, F. (2018). Shadow economies around the world: What did we learn over the last 20 years? IMF Working Paper No. 18/17.

Medina, L., & Schneider, F. (2021). The evolution of shadow economies through the 21st century. The Global Informal Workforce: Priorities for Inclusive Growth. International Monetary Fund, Washington DC, USA. https://doi.org/10.5089/9781513575919.071.

Schneider, F. (2005). Shadow economies around the world: What do we really know? *European Journal of Political Economy*, 21(3), 598-642.

Schneider, F., & Enste, D. H. (2000). Shadow economies: Size, causes, and consequences. *Journal of Economic Literature*, 38(1), 77-114.

Schneider, F., Buehn, A., & Montenegro, C. E. (2010). Shadow Economies All Over the World: New Estimates for 162 Countries from 1999 to 2007. World Bank Policy Research Working Paper No. 5356.

Schneider, F. (2011). The Shadow Economy and Shadow Labor Force: What Do We (Not) Know?. *IZA Discussion Paper* No. 5769.

Schneider, F., Medina, L. (2018). Shadow Economies Around the World: What Did We Learn Over the Last 20 Years? *International Monetary Fund Working Paper*, WP/18/17.

Schneider, F. (2022). New COVID-related results for estimating the shadow economy in the global economy in 2021 and 2022. *International Economics and Economic Policy*, 19, 1-19. https://doi.org/10.1007/s10368-022-00537-6.

Torgler, B., Schneider, F. (2007). The Impact of Tax Morale and Institutional Quality on the Shadow Economy. *Journal of Economic Psychology*, 28(2), 132-152. https://doi.org/10.1016/j.joep.2006.11.003

Yarovenko, H., Vasilyeva, T., Ustinovichius, L., & Remsei, S. (2024). Illicit practices: Experience of developed countries. *Journal of International Studies*, 17(2), 146-177. https://doi.org/10.14254/2071-8330.2024/17-2/8