SAMPLE SELECTION WHEN ESTIMATING WELFARE MEASURES

Dominik Sieradzki, MSc[1]

Faculty of Economic Sciences, Warsaw University of Life Sciences – SGGW

ABSTRACT

In this paper issue of estimation of welfare measures using survey sampling is discussed. Two strategies of sample selection were described. Precision of estimation of the measures depending on chosen sample selection was presented. First sample selection strategy that was discussed is popular simple random sampling without replacement scheme. The second is stratified random sampling. Both sample selection methods were compared with respect to precision of estimation of unknown parameter in the population. Precision of estimation is measured by variance of estimator. It is desirable that the variance of estimator is as small as possible. In such case estimation of welfare measures is the most accurate.

Keywords: welfare measures, sampling methods, efficiency of estimators
JEL codes: D60, C83

INTRODUCTION

Measurement of wealth is a very important issue when talking about theory of welfare economics. Until now among economists there is no definite agreement on which aspects should be taken into account when measuring socio-economic wealth. In this paper issue of accuracy of estimation of welfare measures depending on sample selection strategy is discussed. In the first part theoretical basis of measurement of wealth are presented. Subsequently, two methods of sample selection together with estimators of welfare measures were showed.

In the empirical part accuracy of estimation of welfare measures depending on chosen sample selection strategy was presented with the help of computer simulation. The research hypothesis assumes, that chosen sample selection strategy has a great impact on precision of estimation of welfare measures.

THEORETICAL BACKGROUND

Theoretical reflections concerning wealth economy are centered around wealth measurement. Measuring wealth and expressing it with a number brings much needed knowledge and lets draw conclusions. Some economists tend to interpret welfare as a psychic sphere. A basic tool they use to measure welfare are surveys. Surveys bring questions that enable subjective evaluation of income utility or level of satisfaction of respondents. It soon became clear that social welfare is not only related to income. What is crucial are also non-economic aspects. That is why specific indicators and measures were introduced.

In economic research basic knowledge is brought by social indicator (Panek, 2007). Several definitions are distinguished. One of them establishes, that social indicator is a numeric rating of social effects of economic growth and is considered to be understood

---

[1] Corresponding author: Nowoursynowska 159, 02-787 Warsaw, Poland, dominik_sieradzki@sggw.pl, +4822 593 73 20

broader than a measure (Luszniewicz, 1982). The other one defines social indicator as a measure of direct effects – mainly as welfare, which can be comprehended as a level of satisfying the needs of an individual and a family (Przeciszewski, 1987). Some say that social indicator is the key, strategic or summary measure of social changes (Jońca, 1991). In a quantitative approach social indicators are distinguished as numbers which describe studied phenomenon. These numbers are obtained due to different cognitive methods (Frąckiewicz and Frączkiewicz-Wronka, 2001). Many studies treat indicators and measures as synonyms, but some researchers distinguish one from another. Specific numeric rating is understood as a measure, however if this rating is used to describe changes of e.g. social states, it is defined as an indicator (Panek, 2007). It can be therefore assumed, that numeric information that is measured empirically is called a measure, whereas if this information is interpreted on account of rating social-economic phenomenon, it should be called an indicator (Panek, 2007). Three types of social indicators can be identified: single indicators, such as e.g. unemployment rate, level of education; synthetic (indices), generally an average from simple indicators; multiple indicators, e.g. set of some number of observable empirical indicators, which describe unobservable feature (Panek, 2007).

Specific indicator values are obtained when concluding survey sampling. They include, for instance, percentage of the population that received higher education or percentage of households that possess at least two houses. In such cases, the main task comes down to estimating structure indicator of the studied population.

MATERIALS AND METHODS

In this paper data from Central Statistical Office of Poland was used (GUS, 2018). Data concerns number of population broken down by place of residence – city, village and year of birth as of 31 December 2017. Number of people aged 25 or more at that time equalled 29,018,116 people, out of which 17,875,105 people lived in cities and 11,143,011 people lived in the village. The aim of the research is to estimate number of people with higher education.

Estimation of this type of measures can be made with tools of survey sampling. This method is one of mathematical statistics sections and its main interest concerns finite populations (Bracha, 1996). To formalize the problem of estimation of percentage of people with higher education, let us consider finite population $U = \{u_1, \ldots, u_N\}$ – which contains people aged 25 or more. In this population objects (people) with higher Education are observed (these objects are somehow differentiated). Let $\boldsymbol{Y} = \{Y_1, \ldots, Y_N\}^T$ denote vector of variable value in a population, where for $k = 1, \ldots, N$

$$Y_k = \begin{cases} 1, & \text{if } k\text{-th person has higher education} \\ 0, & \text{otherwise} \end{cases}$$

Let $M = \sum_{i=1}^{N} Y_k$ denote number of people with higher education. The purpose is to estimate, based on the sample drawn, percentage of individuals with a higher education $\theta = \dfrac{M}{N}$. After establishing the main aim of the research it is crucial to decide which sampling scheme will be used to draw sample from the population.

One of the most popular techniques of sample selection is simple random sampling without replacement scheme. In such a case every object of the population can be drawn only once. In the sample the number of people with higher education is a random variable with a hypergeometric distribution. Further in this research this random variable will be denoted as a ξ. Unbiased estimator with minimal variance of percentage of people with higher education equals

$\hat{\theta}_c = \dfrac{\xi}{n}$. Variance of such an estimator equals for

$$D_\theta^2 \hat{\theta}_c = \frac{\theta(1-\theta)}{n} \frac{N-n}{N-1}$$ each value of percentage

of people with higher education θ.

Another broadly used technique related to sample selection is stratified random sampling. It is based on dividing researched population into separate groups called 'strata' and drawing a sample out of each strata independently. Let us assume that we divide popula-

tion U into two strata – depending on a place of residence: city/village. First strata contains N_1 people, second strata – N_2 people and naturally $N = N_1 + N_2$. When taking into account such a division of the population into strata, real percentage of people with higher education equals (Cochran, 1977)

$$\theta = \frac{N_1}{N}\theta_1 + \frac{N_2}{N}\theta_2$$

where θ_1 and θ_2 denote percentage of people with higher education in cities and in the village, accordingly. Let n_1 and n_2 denote sample sizes from the first and the second strata, respectively. The whole sample size equals $n = n_1 + n_2$. Now it is necessary to consider two random variables describing number of people with higher education in samples drawn from each strata

$$\xi_1 \sim H(N_1, \theta_1 N_1, n_1), \xi_2 \sim H(N_2, \theta_2 N_2, n_2)$$

It seems intuitively obvious to take as our estimate of proportion of people with higher education θ

$$\widehat{\theta}_w = \frac{N_1}{N}\frac{\xi_1}{n_1} + \frac{N_2}{N}\frac{\xi_2}{n_2}$$

The variance of the estimator $\widehat{\theta}_c$ equals:

$$D^2_{\theta_1,\theta_2}\widehat{\theta}_w = \frac{\theta_1(1-\theta_1)}{n_1}\frac{N_1 - n_1}{N_1 - 1} + \frac{\theta_2(1-\theta_2)}{n_2}\frac{N_2 - n_2}{N_2 - 1}$$

When stratified random sampling is used, issue of dividing sample into strata occurs. This problem is called a sample allocation. One of best known approaches is proportional allocation (Bowley, 1926), which assumes division of the sample that is proportional to size of the strata, i.e.

$$n_i - n\frac{N_i}{N} \text{ for } i = 1, 2$$

In both methods of sample drawing estimators that are used are unbiased, therefore to make estima-

tion more accurate, it is needed to compare efficiency of estimators. Let

$$\text{reduction} = \left(1 - \frac{D^2_{\theta_1,\theta_2}\widehat{\theta}_w}{D^2_{\theta}\widehat{\theta}_c}\right) \cdot 100\%$$

denote relative reduction of variance. In the next part of the research accuracy of estimation with use of both estimators will be compared in the numerical study with the help of computer simulation.

RESULTS AND DISCUSSION

The standard way of estimation of percentage of people with higher education is to take a sample of size n due to the scheme of simple sampling without replacement. In the sample the number of answers 'yes, I have higher education' is counted. Let us denote this number as x. Evidently the standard estimator of the θ is $\dfrac{\xi}{n}$.

Suppose that size of the sample equals 1,000 and in the whole sample 600 'yes' answers were obtained (i.e. 600 people have higher education). The point of estimation of percentage of people with higher education equals $\widehat{\theta}_c = 0.6$ and its variance may be estimated as

$$\hat{v}(x) = \frac{\widehat{\theta}_c(1-\widehat{\theta}_c)}{1,000}\frac{29,018,116 - 1,000}{29,018,116 - 1}$$

where x is the number of answers 'yes' in the sample. Hence $\hat{v}(600) = 0.000239987$. If the stratified random sampling with proportional allocation is used, it is necessary to determinate sizes of samples: $n_1 = 616$ and $n_2 = 384$. To exemplify, in the sample from first strata there were 250 'yes' answers and in the sample from the second strata there were 350 'yes' answers. The point estimate of the proportion would be $\widehat{\theta}_w = 0.81$. The variance of the estimator $\widehat{\theta}_w$ may be estimated as

$$\hat{v}(250,\,350) = \frac{\dfrac{250}{616}\left(1 - \dfrac{250}{616}\right)}{616} \cdot \frac{17{,}875{,}105 - 616}{17{,}875{,}105 - 1} +$$

$$+ \frac{\dfrac{350}{384}\left(1 - \dfrac{350}{384}\right)}{384} \cdot \frac{11{,}143{,}011 - 384}{11{,}143{,}011 - 1} =$$

$$= 0.000238369$$

The relative reduction of estimated variance equals

$$\text{reduction} = \left(1 - \frac{\hat{v}(250,\,350)}{\hat{v}(600)}\right) \cdot 100\% = 0.67\%$$

Table 1 shows other possible answers results, assuming that the overall 'yes' answers equals up to 600.

Tables 2 and 3 contain the same results, assuming that the overall 'yes' answers equal to 700 and 800, respectively.

It can be observed, that whatever results of the poll in strata are given, the estimator $\widehat{\theta}_w$ is better than the estimator $\widehat{\theta}_c$.

Table 1. Possible results for $\xi = 600$, $\hat{v}(600) = 0.000239987$

ξ_1	ξ_2	Variance	Reduction (%)
250	350	0.000238369	0.67
300	300	0.000219514	8.53
350	250	0.000179522	25.19
400	200	0.000118393	50.67
450	150	0.000036126	84.95

Source: own calculations.

Table 2. Possible results for $\xi = 700$, $\hat{v}(700) = 0.000244495$

ξ_1	ξ_2	Variance	Reduction (%)
350	350	0.0001821	13.27
400	300	0.0001372	34.65
450	250	0.0000712	66.1

Source: own calculations.

Table 3. Possible results for $\xi = 800$, $\hat{v}(800) = 0.000159991$

ξ_1	ξ_2	Variance	Reduction (%)
500	300	0.00000285	98.22
450	350	0.00007379	53.88

Source: own calculations.

CONCLUSIONS

Based on concluded considerations, it seems that stratified random sampling with proportional allocation is a better sample selection strategy. Accuracy of estimation for a sample selected in such a way was up to 98% better than not including information about dividing population into strata. Given the results obtained it should be established, that set research hypothesis is true, therefore strategy of sample selection in fact has impact on accuracy of estimation of welfare measures. In further research including other methods of sample selection and sample allocation in stratified random sampling is planned.

REFERENCES

1. Bowley, A.L. (1926). Measurement of the precision attained in sampling. Bulletin of the International Statistical Institute, 12.
2. Bracha, C. (1996). Teoretyczne podstawy metody reprezentacyjnej [Theoretical foundations of the survey sampling]. PWN, Warszawa.
3. Cochran, W.G. (1977). Sampling Techniques. 3rd ed. John Wiley and Sons, Hoboken, NJ.
4. Frąckiewicz, L., Frączkiewicz-Wronka, A. (2001). Metody pomiaru zjawisk społecznych w skali makro- i mikroregionalnej [Methods of measuring social phenomena on the macro- and micro-regional scale]. Prace Naukowe. Akademia Ekonomiczna im. Karola Adamieckiego. Wydawnictwo Uczelniane Akademii Ekonomicznej im. Karola Adamieckiego w Katowicach.
5. GUS (2018). Ludność według płci i roku urodzenia w 2017 roku [Population by sex and year of birth in 2017]. Retrieved from: http://demografia.stat.gov.pl/bazademografia/Downloader.aspx?file=pl_lud_2017_00_04.zip&sys=lud [Accessed 27.05.2018].
6. Jońca, S. (1991). Wskaźniki społeczne [Social indicators]. Biblioteka Wiadomości Statystycznych. Dział Wydawnictw Statystycznych GUS, Warszawa.
7. Luszniewicz, A. (1982). Statystyka społeczna [Social statistics]. PWE, Warszawa.
8. Panek, T. (2007). Statystyka społeczna [Social statistics]. PWE, Warszawa.
9. Przeciszewski, T. (1987). Planowanie społeczne a polityka społeczna [Social planning and social policy]. PWN, Warszawa.