UNITED
NATIONS

E

Economic and Social Council

Distr.
GENERAL

CES/2001/21
4 April 2001

ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Forty-ninth plenary session
(Geneva, 11-13 June 2001)

ADVANCES IN SMALL AREA ESTIMATION FOR THE POLISH LABOUR FORCE SURVEY APPLIED IN THE STATISTICAL OFFICE OF £ÓD• [1]

Supporting paper submitted by the Central Statistical Office (GUS), Poland

I.   Introduction

1.   In Poland in recent years, the level of unemployment increased to the value, that can cause serious socioeconomic problems, especially in some regions of Poland, where unemployment has rather structural, than frictional nature. Such character of unemployment is observed for example in ³ódzkie voivodship. It is caused by the increasing interest in regional statistical data of the unemployment. In countries in transition, where political, economical and social system transformation take place, the competencies of the central government are delivered to the regional authorities. This in particular is connected with growing demand of the statistical information in the local scale, as it was pointed out by Kordos (1991, 1997). Because of the regional differentiation of the social and economical phenomenon, in the case of making the decisions, the local government should be familiar with the specific parameters that describes the situation in particular region. One of such parameter is the local unemployment size.

2.   In recent years in the Statistical Office in £ódŸ some efforts was made to provide more reliable estimates for Polish Labor Force Survey applied to the regional data. First results were presented in 1998 and have discussed theoretical aspects of that problem (Kubacki (1998)). In this work the proposal of the using of some small estimation techniques to obtain estimates of the size of the unemployment at the regional scale was presented. Next the evaluation of these methods was presented. This was done at the IASS Satellite Conference "Small Area Estimation", that was organized in August, 20-21, 1999 in Riga, Latvia (Kubacki (1999)). Here the results of the computation of the variance were presented. Extended version of this paper was also presented in Statistics In Transition (Journal of the Polish Statistical Association) (Kubacki (2000a)). More results were also presented at the conference "Statystyka regionalna w s³u¿bie samorz¹du lokalnego i biznesu" organised by the Poznañ School of Economy in Poznañ-Kiekrz in June, 5-7, 2000. (Kubacki (2000b)).

---

[1] Paper prepared by Mr Jan Kubacki, Statistical Office in Lodz, Central Statistical Office (GUS), Poland

3.   To obtain the results presented in the papers mentioned above special computer programs was prepared. These programs were integrated with the system that processed the data for Labour Force Survey at the regional level. All programs were prepared in Microsoft FoxPro (the same database system as used in the system processing the individual LFS data). For some analysis one of the module of the Statsoft's Statistica program was used.

4.   In Poland there is relatively big spatial differentiation of the unemployment size. Both Witkowski (1992), Go³ata (1996), Dehnel and Go³ata (2000) pay attention in their works to this fact. There are regions (voivodships) where relatively high unemployment rate occur (20%) and regions where this rate is rather low (5%). There is also high differentiation in unemployment rate within regions. Because of this, it will be useful to obtain the information of the unemployment for areas smaller then region, as Witkowski (1992) suggested. It seems plausible, that the smallest territorial unit for which the analysis of the unemployment could be performed is the local labour market. It is rather well-founded that such market is determined by the area of convenient everyday commuting. In Poland it can be assumed, that this area is delimited by the borders of the County Labour Offices (formerly – before the latest administrative reform – Local Labour Offices).

5.   There are two main available sources of the information of the unemployment in Poland. One of them is the data from the registration of the unemployed, that is carried out by the Labour Offices, the second one is the results of the Polish Labour Force Survey (PLFS) that is performed by the Central Statistical Office. Each of these sources has its own specificity. It is the result of the criterion of ranking of the person among unemployed. In PLFS such criteria are consistent with International Labour Organization classification and are different than these used by Labour Offices. Because of this each source provides different information of the size of unemployment. On the other hand PLFS is the survey which is designed to obtain reliable data for the country, what additionally causes difficulty in obtaining credible results for the local areas (counties).

6.   However in PLFS questionnaire there is a question about whether the interviewer is registered in the Labour Office. Answers to this question can join the information about the registration of the unemployed and the answers to the question, that allows to classify particular person as unemployed in the PLFS sense. It gives the possibility to improve the quality of estimation of the unemployment parameters performed for the local areas.

7.   The aim of this paper is to present the methodology of estimation the size of unemployment in local scale (for the counties) based on the individual data from PLFS and data for the counties obtained from Local Labour Offices. Brief methodology of the PLFS is also presented.

II. Techniques of estimation

8.   Techniques of estimation presented below are based on works of Bracha (1994) and the book of Särndal, Swensson and Wretman (1992).

9.   The basic assumption used in application of the following methods is the availability of the local data of the registered unemployment in Poland. These data are published quarterly by the Central Statistical Office and contained aggregated information for the regions and counties. Local data are approachable for age, sex, education and the duration of unemployment. There is also another advantage in using administrative data of unemployment – relatively high correlation coefficient between the fact that a person is classified as unemployed (in PLFS sense) and the positive answer to the question "Are you registered in Labour Office?" This value for the region considered below is equal to 0.67.

10.   It is assumed that population is divided into non-overlapping geographical domains (counties). Smaller domains could form larger domains (regions) but particular smaller domain belongs only to the one larger region. The same assumption is valid also for the primary sampling unit and secondary sampling unit. It is also founded that population can be classified using non-overlapping age-sex group. Here are some assumptions about notation using in this paper

d – for the geographical domain (in this case – county or region)
g – for the age-sex group
i – for the element of the sample within domain or group
s – for the element in the s elementary sample

11.   There is another assumption for the synthetic estimators where •g denotes the summation over all smaller geographical units (counties) that belongs to the larger unit (region), dg denotes the summation over the cell which is the intersection of the domain d and group g, dgs fixes the aggregation over the cell which is the intersection of the domain d, group g and elementary sample s.

12.   The quantities written in upper-case are related to the population, while quantities written in lower case are related to the sample. In this paper the N_{dg} means the number of persons that belong to the particular age-sex group, the y_i is equal 1 if the respondent is classified as unemployed (0 in the opposite case), the x_i is equal 1 if the respondent answer positively to the question "Are you registered in Labour Office ?" and 0 otherwise.

A.   Horvitz-Thompson estimator

13.   General form of the Horvitz-Thompson estimator is given as follows:

$$y_{dg}^{(1)} = \sum\nolimits_{dg} \frac{y_k}{\pi_k} \tag{1}$$

where summation are taken over all elements of the sample, that belongs to the domain d and group g. To determine the values of π_k it can be useful to use the value of the sample fraction f . However it could lead to biased estimates and often together with fraction is used the coefficient of response. This reduces the bias, but makes the computation more complex. In this technique no auxiliary information is used, so often this estimator is called direct estimator.

B.   Post stratified estimator

14.   As it is mentioned by Särndal, Swensson and Wretman (1992) one can often observe that variation of the population is related with the division of the population into some number of the groups. These groups can be often described by the demographic or education factor. These factors often better explains the characteristics of the population. They can be used with conjunction of the territorial information. Such group model is used in post-stratification estimation. Three main types of the post-stratified estimators can be distinguished.

15.   First of them is the ordinary post-stratification estimator, where only additional information is the size of the population of the interest.

$$y_{dg}^{(2)} = N_{dg} y_{dg}^{*1)} / \hat{N}_{dg} \tag{2}$$

16.   In work presented here we used demographic information from administrative sources. The second is the ratio estimator where auxiliary information of the cell total X_{dg} for the dg cell is used

$$y_{dg}^{(3)} = \frac{\sum\nolimits_{dg} y_i}{\sum\nolimits_{dg} x_i} X_{dg} \tag{3}$$

The third is the regression estimator

$$y_{dg}^{(4)} = N_{dg} (\bar{y}_{dg} + b_{dg} (\bar{X}_{dg} - \bar{x}_{dg})) \tag{4}$$

C. Synthetic estimator

17. When the synthetic estimator is applied to the specific domain/group it is assumed that the structure in the larger domain/group is similar to domain/group of interest. Using this class of estimators is often useful when there is no data from sample for an individual cell in the domain/group. In our case we use sample data from larger administrative units. As in the case post-stratified estimator we distinguish three types of synthetic estimators:

The first one is the ordinary synthetic estimator:

$$y_{dg}^{(5)} = N_{dg} \frac{\sum_{\bullet g} \frac{y_i}{p_i}}{N_{\bullet g}} \qquad (5)$$

The second is the ratio estimator

$$y_{dg}^{(6)} = X_{dg} \frac{\sum_{\bullet g} y_i}{\sum_{\bullet g} x_i} \qquad (6)$$

The third is the regression estimator

$$y_{dg}^{(7)} = N_{dg} (\bar{y}_{\bullet g} + b_{\bullet g} (\bar{X}_{\bullet g} - \bar{x}_{\bullet g})) \qquad (7)$$

D. Modified generalized regression estimator (MRE)

18. Here we use only one type of MRE estimator. It uses the synthetic regression estimator as its basis plus an adjustment term involving regression residuals.

$$y_{dg}^{(8)} = N_{dg} (\bar{y}_{\bullet g} + b_{\bullet g} (\bar{X}_{\bullet g} - \bar{x}_{\bullet g})) + \frac{N_{dg}}{\hat{N}_{dg}} \sum_{dg} \frac{e_i}{p_i} \qquad (8)$$

III. Estimation of variance

19. In evaluation of stochastic performance of the above estimators random group method was used. Here the independence of the elementary sample and availability of the individual data was exploited. It will be explained using the post-stratified estimator as an example. As it was presented by Wolter (1985) the variance estimator $\hat{s}_{dg}^{(l)2}$ can be obtained using the following formula

$$\hat{s}_{dg}^{(l)2} = \frac{1}{k(k-1)} \sum_{s=1}^{k} (y_{dgs}^{(l)} - y_{dg}^{(l)})^2 \qquad (9)$$

where l denotes the kind of estimator and s the elementary sample. In the case of post-stratified estimation $y_{dgs}^{(l)}$ can be obtained as follows:

For ordinary estimator:

$$y_{dg}^{(2)} = N_{dg} y_{dgs}^{*1)} / \hat{N}_{dg} \qquad (10)$$

For ratio estimator:

$$y_{dgs}^{(3)} = \frac{\sum_{dgs} y_i}{\sum_{dgs} x_i} X_{dg} \qquad (11)$$

For regression estimator:

$$y_{dgs}^{(4)} = N_{dg} (\bar{y}_{dgs} + b_{dgs} (\bar{X}_{dg} - \bar{x}_{dgs})) \qquad (12)$$

20.	Similar conclusion can be obtained for the remaining estimators

IV. Conclusion

21.	Further improvement of this methods can be made by applying more advanced techniques. The choice of more advanced method should take into consideration the compromise between preserving the structure of the estimates, minimizing the estimates variance and bias, and also availability of the sample and the auxiliary data at the local scale. However the properties of such methods should be always referred to the more basic methods of estimation, so the above results can be treat as a basis for future work, when the model approach can be applied.

- - - - -

REFERENCES

BRACHA C. (1994) Metodologiczne Aspekty Badania Ma³ych Obszarów (Methodological Aspects of Small Area Research). series "Z Prac Zak³adu Badañ Statystyczno-Ekonomicznych" Central Statistical Office, Warsaw

GO£ATA E. (1996) Statystyka ma³ych obszarów w analizie rynku pracy (Small Area Statistics in the Labour Market Analysis), Wiadomoœci Statystyczne, 3, 45-60

DEHNEL G., GO£ATA E. (2000) An Attempt to Small Area Estimation in the Largest Region in Poland, Statistics in Transition, **4**, 5, 753-768

KORDOS J. (1991) Statystyka ma³ych obszarów a badania reprezentacyjne (Small Area Statistics and Sample Surveys), Wiadomoœci Statystyczne, 4, 1-5

KORDOS J. (1997) Efektywne wykorzystanie statystyki ma³ych obszarów (Efficient Use of Small Area Statistics), Wiadomoœci Statystyczne, 1, 11-19

KUBACKI (1998) Szacowanie bezrobocia w skali lokalnej, Wiadomoœci Statystyczne (Estimation of the value of unemployment on the local scale), 2, 10-20

KUBACKI (1999) Evaluation of Some Small Area Estimation Methods for Polish Labour Force Survey in One Region of Poland, Proceedings of the IASS Satelite Conference on Small Area Estimation, Riga, Latvia, 245-249

KUBACKI (2000a) Some Small Area Estimation Methods for Polish Labour Force Survey in One Region of Poland, Statistics in Transition, **4**, 5, 769-777

KUBACKI (2000b) Wybrane metody statystyki ma³ych obszarów zastosowane w szacowaniu bezrobocia dla powiatów, (Selected Methods of Small Area Statistics Applied to the Estimation of the Size of the Unemployment For Counties - Poviats) in proceedings of the conference: "Statystyka regionalna w s³u¿bie samorz¹du lokalnego i biznesu", Kiekrz, 5-7, June, 2000.

SÄRNDAL C.E., SWENSSON B., WRETMAN J. (1992) Model Assisted Survey Sampling, Springer-Verlag

SZARKOWSKI A. and WITKOWSKI J. (1994) The Polish Labour Force Survey. Statistics in Transition, 4, 467-483

WITKOWSKI J. (1992) Szacowanie bezrobocia dla ma³ych obszarów (Assessment of Unemployment for Local Areas), Wiadomoœci Statystyczne, 11, 1-9

WOLTER K.M. (1985) Introduction to Variance Estimation, Springer-Verlag