UNIWERSYTET
EKONOMICZNY
W POZNANIU

Summary

mgr Tomasz Józefowski

The labour market in functional urban areas of provincial capital cities in the light of Structure Preserving Estimation

Doctoral dissertation

Supervisor: dr hab. Grażyna Dehnel, prof. UEP
Auxiliary supervisor: dr hab. Marcin Szymkowiak, prof. UEP

Institute of Informatics and Quantitative Economics

Department of Statistics

The monitoring and assessment of the labour market are among the key elements of effective state administration. Given reliable information in this regard, one can take measures aimed at counteracting negative effects of ongoing changes. As a member of the EU, Poland is obliged to provide information about the labour market that can be used for reliable comparisons between member states. It is therefore necessary to develop a survey methodology capable of characterising the labour market in Poland which complies with the common guidelines.

In accordance with its statutory goals, Statistics Poland (GUS) conducts the Labour Force Survey (BAEL), whose purpose is to obtain information about the size and structure of employed, unemployed and economically active populations, taking into account their socio-demographic characteristics. However, in the analysis of LFS results, one needs to consider the limitations of the representative method. Because of its design, the LFS cannot provide reliable estimates for more detailed domains, since smaller sample sizes in domains of interest can cause large estimation errors. This is why the LFS results are published for provinces, which is the lowest level of spatial aggregation, where, so far, there is no item nonresponse for basic characteristics of the labour market (GUS, 2018).

The current availability of data is also the result of limited funds for conducting the LFS survey and the desire to reduce response burden for respondents, who are less and less willing to participate in surveys. On the other hand, it is worth noting that recipients of statistics tend to expect a broader scope of information than what the LFS survey can provide.

Faced with the growing demand for information, expressed by national and international users, official statistics is expected to provide reliable information about previously unplanned domains in representatives surveys. One of such domains are cities with their functional areas: there is a big demand for all kinds of information about such functional urban areas, especially as regards their labour markets. The literature of the subject includes a number of ways for defining and delimiting such areas (Śleszyński, 2015). Given their growing importance, national strategic documents have established certain regulations in this respect. In the National Spatial Development Concept 2030, a functional urban area is defined as "a spatially continuous settlement system consisting of administratively separate units. It covers a compact urban area with a functionally linked

urbanised zone" (NSDC 2011:184) [1]. The category is divided into four subtypes: functional areas of provincial (including metropolitan), regional, sub-regional and local centres. The demand for information about these territorial units results mainly from the fact that urban policy is an integral part of the territorial dimension in the implementation of the social cohesion programme for 2014-2020 and one of the priorities of financial support provided by the EU. In this context, functional urban areas have become particularly important, and the change in the way development planning is approached nowadays has opened new possibilities for using EU funds and instruments to support functional urban areas. One of the instruments devised by the European Commission for this purpose are Integrated Territorial Investments (ITI). In the period 2014-2020, ITIs were implemented in functional urban areas of Poland's provincial capitals. The relevance of this problem is evidenced by numerous research articles, e.g. (Batóg i in., 2015; Kociuba, 2015, 2017b,a, 2018b,a; Szafranek, 2015, 2017b,a; Szafranek i Kociuba, 2018; Szlachta, 2017; Śleszyński, 2015) devoted to ways of delimiting FUAs and investigating aspects of their functioning.

For such spatial units only information from sample surveys is currently available but these do not provide detailed statistics about the labour market because the sample size in detailed domains is too small. In this situation, the standard method of direct estimation used to produce official statistics would yield estimates of low precision.

These limitations can be overcome by methods of small area estimation (SAE))[2], which make it possible to obtain reliable estimates at lower levels of spatial aggregation or for more detailed domains. SAE methods can be applied even in situations when there are no representatives for certain cross-classifications (Rao i Molina, 2015). SAE methods prove useful in many fields of application, such as poverty, disability, economic statistics or the labour market (Dehnel i Wawrowski, 2020; Gołata, 2010; Szymkowiak i in., 2017; Klimanek i in., 2017; Wilak, 2020; Józefowski i Szymkowiak, 2013).

It is worth noting at this point that while estimates obtained by means of various SAE methods are generally characterised by better precision compared to direct estimation, estimates produced for smaller territorial units, when aggregated, often are not consistent

[1] `https://www.kooperation-ohne-grenzen.de/wp-content/uploads/2016/05/NSDC-2030.pdf`

[2] In the literature of the subject, this field of statistics is also referred to as indirect estimation.

with direct estimates at higher levels of aggregation[3]. This is one of the reasons why SAE methods are not commonly applied to data from official statistical surveys. One way in to overcome this inconsistency problem is to use Structure Preserving Estimation (SPREE). SPREE estimators retain the association structure for auxiliary variables and domains of interest, which is known from administrative registers or censuses, while ensuring consistency with estimates at higher levels of spatial aggregation. It is a very desirable property of SPREE estimators, especially when one takes into account the requirements regarding the quality of outputs published by national statistical institutes (Paradysz, 2009).

The SPREE technique seems to particularly useful when estimating characteristics for unplanned domains that do not overlap with the administrative division. So far, for example, no method has been developed to produce information about employed, unemployed and economically active populations (as defined by the International Labour Organisation) at the level of functional urban areas of provincial capitals (MOF OW).

The purpose of this dissertation is to partly bridge the existing gap in this respect. It is an attempt to develop a method of estimating characteristics of the labour market in functional urban areas of provincial capitals by applying SPREE estimators, which ensure consistency with direct (including calibration) estimates at a higher level of spatial aggregation. It is also one of the first studies in Poland that provides such a comprehensive analysis of this class of indirect estimators, and, as far as I know, the first one devoted to the study of the urban market in functional urban areas.

Aims and research hypotheses

The main objective of the dissertation was to measure and assess the labour market situation in functional urban areas of provincial capitals in 2017 using Structure Preserving Estimation. The analysis is based on three indicators: the economic activity rate, the employment rate and the unemployment rate, which were calculated for domains cross-classified by sex and age group (15–29, 30–39, 40–49, 50 and older). In order to arrive at these indicators, counts of employed, unemployed and economically inactive people in these domains had to be estimated and their precision had to be assessed. Another im-

[3]For instance, the number of unemployed persons in districts (Pol. *powiat*) of a particular province, estimated by means of SAE methods, may not add up to the direct estimate of the total number of unemployed persons in the province.

portant goal of the dissertation was to provide a comprehensive description of Structure Preserving Estimation.

The structure of the dissertation consisting of five main chapters reflects the process in which its goals were achieved. The first chapter contains an introduction to the subject of functional urban areas (FUA), including a review of their definitions and methods of delimiting them. In particular, the analysis focuses on functional urban areas were Integrated Territorial Investments have been implemented, taking into account their importance for the UE's cohesion policy. The chapter also presents the role of FUAs in the light of national strategic documents. The last section of this chapter is devoted to an overview of available statistical information about FUAs, especially regarding the labour market, and the analysis of user expectations in this respect.

The next stage of the study involved analysing sources of data about the labour market in FUAs. Accordingly, the second chapter covers topics such as data availability depending on the spatial level of aggregation, frequency of data publication and consistency of definitions, as well as the quality of information about the number of employed, unemployed and economically active persons obtained from surveys, censuses and administrative registers. Various sources of auxiliary data were compared (censuses and administrative registers), especially in terms of their timeliness, definitional consistency and correlations with estimated variables. Based on this examination, potential sources of auxiliary variables were identified that could be used in the process of indirect estimation at a later stage.

The third chapter contains a comprehensive overview of small area estimation and the underlying idea of structure preserving estimation. Various SPREE estimators, both classic and model-based, were described in detail, together with their mathematical formulas and properties. A brief summary of the evolution of these methods was also provided. Other aspects covered in the chapter include ways of assessing estimation quality and recommendations regarding the use of indirect estimation techniques. Selected examples of applying SPREE estimators in various areas of socio-economic research were also presented.

The fourth chapter describes the entire research procedure, data sources and variables used in the empirical study (including the target set of functional urban areas), variants of estimators employed in the estimation process and methods of assessing their quality, as

well as the structure of output tables. The chapter contains a detailed analysis of estimated quantities (the number of employed, unemployed and economically active persons) for all estimator variants in terms of estimation quality. The assessment was made for detailed domains defined by spatial division, labour market status, sex and four age groups. After reviewing the quality of obtained estimates, the "optimal" estimator was selected, which was used to calculate the indicators of interest: the economic activity rate, the employment rate and the unemployment rate.

The fifth and final chapter is devoted to analysing the quality of the estimated indicators for the target cross-classification. In addition, taxonomic techniques were used to assess the spatial variation in the labour market situation within functional urban areas of provincial capitals (MO FOW). The assessment included a comparison between values of the labour market indicators in the FUAs and those observed in the remaining part of the respective provinces. The estimates were assessed with respect to the research hypotheses put forward at the beginning.

The goals set out in the dissertation were related to the following four research hypotheses:

1) Structure preserving estimation can be used to estimate basic characteristics of the labour market at the level of spatial aggregation for which no statistical information is currently available,
2) Estimates of basic characteristics of the labour market obtained by applying SPREE estimators have a better precision compared to calibration techniques used in the LFS,
3) When one accounts for sex and age group, the labour market situation of functional urban areas of provincial capitals is better than that observed the remaining part of the province,
4) There is clear evidence of spatial variation in the labour market situation across functional urban areas of provincial capitals.

The novelty of the dissertation consists in the author's proposal of a research procedure involving indirect estimation, which is based on international guidelines. The application of the procedure in the empirical study described in the dissertation is, as far as can be established, the first attempt in Poland to obtain and analyse estimates at this level of aggregation. Another contribution of the dissertation is the first application of structure

preserving estimation to data about the labour market with regard to domains for which no official statistics are currently available. It is also worth noting that the methods tested and applied in the dissertation can be used as a tool for estimating other characteristics of the labour market at the level of functional urban areas.

Research methods and data sources

The SPREE estimators used in the study are a generalised group of synthetic estimators, which rely on information provided by direct estimates. The key element used in the calculation of these estimators is the iterative proportional fitting procedure. The IPF algorithm is used to update counts inside cells of a multidimensional contingency table to ensure that adjusted counts sum up to known marginal totals. Initial counts inside the contingency table can be taken from various data sources, such as a census or an administrative register. Marginal totals are usually filled with direct estimates calculated from survey data. Purcell and Kish (1980) were among the first researchers to propose the use of a structure preserving estimator to update data in a contingency table containing data from the last census and the latest survey and using the IPF method. Their approach was based on the work by Deming and Stephan (1940). Another seminal article on structure preserving estimation was published by Zhang and Chambers (2004), who defined a class of generalised linear structural models (GLSM) and extended the existing approach to mixed models with normally distributed random effects (GLSMM).

Different variants of SPREE estimators based on GLSM and GLSMM models were considered in the dissertation in order to estimate the key characteristics of the labour market for functional urban areas of provincial capitals. The estimation was based on data from the LFS and the Integrated Register of Employed and Unemployed Persons (ZROPiB), compiled using various administrative sources (among others those maintained by the Social Insurance Institution (ZUS), The Agricultural Social Insurance Fund (KRUS) and the unemployment register) as well as updated population estimates. The spatial variation in the labour market situation across FUAs was assessed using selected techniques of taxonomic analysis: the TOPSIS method and Ward's method of hierarchical clustering.

Selected results and conclusions

The estimated indicators of labour force participation, employment and unemployment for the target output tables made it possible to assess the labour market situation for

the domains of interest, especially within functional urban areas of provincial capitals in relation to the research hypotheses. Based on these indicators, it was possible to identify domains of concern, defined by province, sex and age, where the labour market situation should carefully be monitored. For example, with respect to the unemployment rate, one such domain of concern were people aged 15—29. In terms of the economic activity rate and the employment rate, the situation of men under 30 is worse in FUAs of provincial capitals than in the remaining part of their respective provinces. Another reason for concern are low levels of these two indicators observed in the group of women aged 50+, although their situation in FUAs of provincial capitals is better than in the remaining parts of their respective provinces.

The results of the taxonomic analysis were used to create a ranking of FUAs of provincial capitals and identify more homogenous clusters in terms of the labour market situation. The ranking made it possible to illustrate how the phenomenon in question varied across the FUAs in 2017, especially given the fact that no statistical outputs are currently published for such areas.

The proposed integration of data from the LFS survey, administrative registers and current population estimates is in line with the guidelines set out by Eurostat regarding the process of data integration in order to increase the scope of statistical outputs. It can also be viewed as a contribution to the development of non-classical estimation methods in statistical practice.

SPREE estimators can be expected to play a growing role in the production of official statistics in Poland. There are a number of basic reasons for this belief. First of all, in the system of official statistics in Poland the emphasis is increasingly placed on the use of non-statistical data sources (mainly administrative registers). Secondly, in the context of the global pandemic caused by the SARS-CoV-2 virus, it is necessary to conduct comprehensive studies of how the labour market situation is affected by changes taking place in the economy and in society. Finally, faced with the growing demand for information at low levels of spatial aggregation and for more detailed domains survey administrators will have to resort to more sophisticated methods. The SPREE approach described in the dissertation can help to meet this demand.

References

Batóg, J., Mojsiewicz, M., Batóg, B. (2015). Monitorowanie strategii rozwoju w świetle delimitacji miejskich obszarów funkcjonalnych. *Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania Uniwersytetu Szczecińskiego*, 42(1):11–30.

Dehnel, G. Wawrowski, Ł. (2020). Robust estimation of wages in small enterprises: The application to Poland's districts. *Statistics in Transition*, 21(1):137–157.

Deming, W. E. Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, 11(4):427–444.

Gołata, E. (2010). Estymacja pośrednia aktywności ekonomicznej na potrzeby spisu opartego na rejestrach. *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Poznaniu*, 149:85–104.

GUS (2018). Zeszyt metodologiczny. Badanie Aktywności Ekonomicznej Ludności. Główny Urząd Statystyczny, Warszawa.

Józefowski, T. Szymkowiak, M. (2013). Zastosowanie estymatora typu SPREE w szacowaniu liczby osób bezrobotnych w przekroju podregionów. *Studia Oeconomica Posnaniensia*, 1(10):120–136.

Klimanek, T. i in. (2017). Estymacja pośrednia niepełnosprawności–badanie symulacyjne. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 469:91–99.

Kociuba, D. (2015). Miejskie obszary funkcjonalne - wyzwania planistyczne. *Studia Miejskie*, 18:39–53.

Kociuba, D. (2017a). Delimitacja miejskich obszarów funkcjonalnych ośrodków wojewódzkich w realizacji zintegrowanych inwestycji terytorialnych w Polsce - teoria versus praktyka. *Studia Regionalne i Lokalne*, 69(3):54–78.

Kociuba, D. (2017b). Zintegrowane Inwestycje Terytorialne jako narzędzie zarządzania terytorialnego w miejskich obszarach funkcjonalnych. *Studia KPZK*, 174:144–153.

Kociuba, D. (2018a). Implementation of Integrated Territorial Investments in Poland - Rationale, Results, and Recommendations. *Quaestiones Geographicae*, 4:81–98.

Kociuba, D. (2018b). Teoria i praktyka wsparcia zrównoważonego rozwoju miejskich obszarów funkcjonalnych w Polsce. *Biuletyn KPZK PAN*, 272:316–327.

Paradysz, J. (2009). Ocena dobroci estymacji dla małych obszarów. *Zeszyty Naukowe/Uniwersytet Ekonomiczny w Poznaniu*, 128:19–28.

Purcell, N. J. Kish, L. (1980). Postcensal Estimates for Local Areas (Or Domains). *International Statistical Review / Revue Internationale de Statistique*, 48(1):3–18.

Rao, J. M. K. Molina, I. (2015). *Small Area Estimation*, John Wiley and Sons, New Jersey, 2nd edition.

Śleszyński, P. (2015). Problemy delimitacji miejskich obszarów funkcjonalnych w Polsce. *Rozwój Regionalny i Polityka Regionalna*, 29:37–53.

Szafranek, E. (2015). Zintegrowane Inwestycje Terytorialne jako narzędzie wspierające kształtowanie centrów rozwoju regionalnego. *Rozwój Regionalny i Polityka Regionalna*, 32:81–94.

Szafranek, E. (2017a). Idea a praktyka wdrażania Zintegrowanych Inwestycji Terytorialnych w Polsce, [w:] Teoria i praktyka rozwoju obszarów funkcjonalnych. Uniwersytet Opolski. czasopismo. czasopismo.

Szafranek, E. (2017b). Miejskie obszary funkcjonalne a kształtowanie spójności terytorialnej. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 467:113–132.

Szafranek, E. Kociuba, D. (2018). Development of urban areas in the conditions of territorial-oriented policy–theoretical assumptions and experience in functional areas of Polish cities. *Studia Ekonomiczne*, 361:53–75.

Szlachta, J. (2017). Obszary funkcjonalne w Polsce – perspektywa europejska. *Studia KPZK*.

Szymkowiak, M., Młodak, A., Wawrowski, Ł. (2017). Mapping poverty at the level of subregions in Poland using indirect estimation. *Statistics in Transition*, 18(4):609–635.

Wilak, K. (2020). Ocena skali bezrobocia biernego w Polsce. *Wiadomości Statystyczne*, 65(6):11–38.

Zhang, L. C. Chambers, R. L. (2004). Small area estimates for cross-classifications. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 66(2):479–496.