Session Number: Parallel Session 7A
Time: Friday, August 27, AM

Combining of Survey and Administrative Data for Cross-Border Areas

Zofia Barbara Liberda, Marek Cierpiał-Wolan, and Władysław Wiesław Łagodziński

For additional information please contact:

Name: Zofia Barbara Liberda
Affiliation: Warsaw University and Statistics Poland

Email Address: barbara.liberda@uw.edu.pl

Combining of Survey and Administrative Data for Cross-Border Areas

Zofia Barbara Liberda[1], Warsaw University and Statistics Poland
Marek Cierpiał-Wolan, Statistics Poland
Władysław Wiesław Łagodziński, Statistics Poland

Abstract

This paper presents an attempt at modeling a combination of survey and administrative data for cross-border areas. Regional statistics come across various data collection problems: limited data availability for areas located on both sides of the national border, lack of information on a certain level of aggregation in individual countries, low level of data comparability. The data model describing cross-border areas consists of survey data at micro level and of administrative, standardized sources of information. The survey data comes from detailed household surveys in cross-border regions, from some special surveys of persons crossing borders at particular places, as well as some surveys conducted in local enterprises. The administrative data sources are: official registers for local regions; bank registers; geo-coded data, e.g. land parcels registers, housing construction permits; other administrative sources of data, e.g. car registrations, tourist services registration; automatic measurement of traffic, etc. The paper is based upon original data collected for the cross-border areas of Poland, Ukraine and Slovakia. We focus on the problem of special data sensitivity due to different nationalities of respondents as well as legal issues of confidentiality due to different national laws and border regulations. The results of the combined data system research may serve for the national accounts statistics.

---

[1] **Corresponding author: barbara.liberda@uw.edu.pl; z.liberda@stat.gov.pl**

Introduction

The processes of integration in Europe (the most recognizable among them being the extended Schengen zone) create the need for more and more information concerning transborder areas, hence growing interest in regional statistics.

Regional statistics often come across different problems, such as limited availability of data for areas located on both sides of the national border, lack of information on a certain level of aggregation in individual countries or low level of data comparability, especially one pertaining to economic issues.

It is essential then, to create a coherent research system for cross-border areas that would allow gathering and analyzing comparable data from sample survey and administrative registers.

Polish official statistics resumed in 2008 the survey of goods and services turnover in border traffic. Initially, it was carried out at the Polish-Ukrainian border, and since 2010 at the European Union's whole external border on the territory of Poland. The results of this survey turned out to be unexpected. In the first quarter of 2010, the expenses incurred on the purchase of goods in Poland by Ukrainians amounted to 407 mln zl whereas by Belorussians 205, which constituted 16 % and 23 % of Polish exports to these countries, respectively. Therefore, it came out that the scale of unregistered trade is surprisingly high.

It seems to be important to launch a survey system for those areas, of which a crucial part is monitoring the socio-economic phenomena and the results' linkage of this monitoring to the information analyses included in other administrative registers.

This paper presents an attempt at modeling a linkage of survey and administrative data for cross-border areas. The objective of this paper is developing a survey methodology for cross-border areas throughout integration of administrative data sources sample surveys data from the Border Guard, as well as the Labor Force Survey, the Household Budget Survey, the Survey of Goods and services turnover in Border Traffic and Social Diagnosis.

1.1. Border Traffic Register

The system of border traffic record is one of the components of the entire system of border and visa control. In this system, the following information is collected: time of the event (crossing); country, name of a border crossing point, time of entry/exit, full name, date of birth, citizenship, type and number of a document (passport, visa, MRG),

pedestrian/vehicle (registration number, vehicle type, brand, whether the person is a driver/passenger), additional information (e.g. entry purpose, detention decision). This system is linked with other databases – the Police, Interpol, etc.

1.2 Labour Force Survey

The Labour Force Survey has been carried out since 1992 by the Central Statistical Office, in accordance with the methodology of the International Labor Organization, thus the comparability of the data on an international scale is assured. The main objective of the Labor Force Survey is to obtain data on the size and structure of labor resources. The results of the survey are used primarily for:

- determining the balance of the workforce divided into three basic categories of the population - employed, unemployed and economically inactive,

- analysis of changes in the economic activity of population in different socio-professional groups, and territorial sections,

- analysis of the situation on the labor market, including the assessment of scale and intensity of changes in unemployment in spatial covers,

- socio-demographic characteristics of the unemployed,

- analysis of the structure of employment by socio-demographic and professional characteristics.

The situation regarding the economic activity of members of households, i.e. performing work, being unemployed or economically inactive in the reference week is the object of the survey.

The Labor Force Survey is a probability sample survey, which means that results from the survey are generalized over the whole population. The sample covers all people at the age 15 and more, who are members of households in the sampled dwellings. The survey does not cover persons absent from the household when the actual period of absence exceeded 3 months, while the total duration of absence (actual and planned) is 12 months or longer. Since the first quarter of 2004, the survey includes foreigners who are also members of the sampled households.

The Labour Force Survey is carried out quarterly as a continuous observation, i.e. in each of the 13 weeks of the quarter, 1/13 of a quarterly sample of dwellings is surveyed. In the Podkarpackie voivodship in each quarter of 2010 survey covered 2808 dwellings.

Selection of the quarterly samples is performed according to the rotation system, which is as follows:

- the sample for each quarter consists of four elementary samples, which are divided into 13 weekly samples,

- partial rotation of elementary samples is carried out in every quarter: in a given quarter there are two elementary samples surveyed in the previous quarter, one elementary sample introduced into the survey for the first time and one sample which was not surveyed in the previous quarter and was introduced into the survey exactly a year before,

- each elementary sample is selected independently.

The survey has been carried out by interviewers of Statistical Offices in randomly sampled dwellings, and the responses to the questions should be provided by the person to whom a questionnaire is addressed. In special cases, it is allowed that another member of the household, well-informed about the professional situation of the person absent, answers the questions.

The survey is carried out in the form of paper questionnaires and LFS applications for laptops.

Two paper questionnaires are used:

- Household File for each household in a randomly sampled dwelling; these ZG files constitute a register of households during the whole survey cycle.

- Household questionnaire which is filled in on a quarterly basis for each person covered by the survey, i.e., persons 15 years-old and more present in a household or absent, if the actual period of absence is no longer than 3 months, or persons absent for longer than 3 months but the total duration of absence (actual and planned) is no longer than 12 months.

Additional questionnaire was introduced and filled out in case the survey was not carried out (e.g. refusal, absence of inhabitants).

1.3 Household Budget Survey

The survey of household budgets is a voluntary survey based on the sampling method that allows for generalization of the results for the whole population of households within the margin of error.

The adopting sample scheme was a two-stage one, geographically stratified, with different probabilities at the first selection stage. The sampling units for the first stage were the area survey points, and those for the second stage were dwellings.

The area survey point is a statistical area or set of areas. It is assumed that an urban area survey point should consist of at least 250 dwellings, while a rural one – 150 dwellings. In the household budget survey, the adopted model is full rotation with a monthly replacement of the sample. This means that for each month 2 dwellings per area survey point are selected, and all the households from these dwellings participate in the survey. Every selected dwelling participates in the survey in the same month of the two consecutive years, i.e. in 2009-2010 for the second subsample and 2010-2011 for the first subsample. In 2010, 162 dwellings have been selected for the monthly survey in the Podkarpackie voivodship.

The survey unit is a one-person or multi-person household. The object of the survey is primarily its budget, i.e. the amount of receipts and expenses (in cash and in kind) of all members of the surveyed household, and a quantitative consumption of selected goods and services. In addition, within the frame of the survey, information is collected on demographic and socio-economic characteristics of people within the household, as well as information on living conditions, household's equipment with durables and the subjective evaluation of the financial situation of the household.

In the household budget survey the following questionnaires are used:

- Household Statistical Sheet,
- Information on participation in the survey,
- Budget Diary,
- Additional information on the household,
- Income in kind from hired work,
- List of dwellings for the household budget survey in the area survey point,
- List of randomly selected dwellings for the household budget survey in the area survey point.

Household budget surveys are conducted by interviewers in randomly selected area survey points, while the records of expenditures and revenues of a household are carried out by the household in their "Budget Diary."

Data obtained from the survey allow for the analysis of the living conditions of the population and evaluation of the impact of various factors on the living conditions of the main population groups and their differentiation.

The household budget survey provides detailed information on:

- the demographic structure of households, i.e. the number of household members, their age, gender, education, disability and economic activity;

- the level and sources of income;

- the level and structure of expenditure as well as sources of goods and services;

- the consumption level of basic food products according to quantity, but also energetic value and nutrients

- prices at which households purchase selected goods and services;

- households' equipment with durable goods

- dwelling conditions

- the subjective evaluation of the financial situation of households.

The results of household budget surveys are mainly used for: the analysis of the level of living conditions and differentiation for the main socioeconomic groups of households; the analysis of the level of living conditions and differentiation for the main socioeconomic groups of households according to dynamic approach; natural consumption and poverty surveys; studying nominal and real income levels and changes observed in this area for the particular groups of households, and for the detailed surveys of the consumer market forecasts, and other economic analyses.

1.4 The survey of goods and services turnover in border traffic at the European Union's external border in Poland.

The survey of goods and services turnover in border traffic covered persons who cross the Polish-Ukrainian border, Polish-Belarusian border, and Polish-Russian border, that is, foreigners (who permanently reside abroad) leaving Poland and Poles (who permanently reside in Poland) returning to the country. The survey deals with persons crossing the border overland: by cars, buses, motorbikes, on foot and by train. Persons who crossed the border on foot also included cyclists and disabled persons using wheelchairs. The survey covers expenses incurred by foreigners in Poland and by Poles abroad for the purchase of goods, as

well as other expenses, such as, costs of accommodation and gastronomic services. The survey examines the expenditures on the purchase of goods not registered in customs documents. The survey investigated also: the distance from the place of residence and the place of purchase to the border; frequency of crossing the border; purpose and duration of stay; country of residence - in the case of foreigners; country of visit - in the case of Poles; possession of Pole's Card – in the case of foreigners.

The basic sources of information on goods and services turnover in border traffic are the questionnaires gathered by pollsters of the official statistics directly on the border crossings. The results of this survey are estimated on the basis of the secondary results of border traffic survey and supplementary information of the Borders Guards on the border traffic recorded on the days when the survey was carried out. The questionnaires are filled in by respondents unaided, or by pollsters during the interview. The questionnaires for foreigners are prepared in the following languages: Ukrainian, Russian, English, French and German. The survey covers the assortment of goods and services most popular with Poles and foreigners who cross the border, respectively. For this reason, there is a difference in the assortment of goods and services the foreigners and the Poles are questioned about. If the volume of expenses is given by foreigners in foreign currency, it is converted into zlotys according to the average exchange rate of the National Bank of Poland at the day of the survey.

The survey of goods and services turnover in border traffic is carried out using the representative method, which allows generalization, with specified error, of the obtained results for a total number of persons who cross the surveyed Polish-Ukrainian border, by foreigners and Poles. A two-stage scheme was used for drawing a sample and determining the layers. First, the days (time intervals) undergoing survey are drawn, and then persons are drawn out of those who cross the border. The layers were determined according to the days of the week, the border crossings, and the kind of traffic. For each of the layer one, selected at random, 12-hour interval is surveyed in a quarter coinciding with a day shift of the border guards. Drawing of a sample is the same for Poles as for the foreigners. For each of the selected shift (a unit of the first stage of drawing participants in the survey) a sample of persons undergoing the survey is selected by means of systematic sampling. In case a selected person refuses to participate in the survey, a successive person is surveyed. The questionnaire survey is carried out in quarter periods, on selected days of a week chosen from the total number of days in a given period (7 times in the quarter). Non-representative days are not included in the drawing, e.g. national and religious holidays. The questionnaire survey is

conducted simultaneously at all border crossings covered by the survey. The estimation of survey results is based on the data gathered in the questionnaires and the information of the Border Guards on border traffic, which concern respective crossings, including the way of crossing the border. These data cover the number of Poles and foreigners who cross the border according to the crossing, direction and kind of traffic (the manner of crossing the border) in a surveyed quarter and in 12-hour shifts during which the questionnaire surveys were carried out. Data are generalized separately for the Poles and foreigners in each layer. Results for voivodships (regions) are calculated on the basis of the results from all layers.

The survey covers all land and railway border crossings with passenger traffic at the European Union's external border in Poland. The survey was conducted on about 1% sample.

1.5 Social Diagnosis

The objective of the research project is to describe conditions and quality of life of the society: its development potential and the direction of changes, threats and challenges. The project comprises many aspects associated with the situation of households and individual citizens, which can be divided into three classes: the demographic and social structure of households; the living conditions of households associated with their material status, access to healthcare services, culture, recreation, education and modern communication technologies; the subjective quality of life, lifestyle, beliefs, attitudes and behaviors of individual respondents.

Demographic and social structure of the households serve as means of stratifying the groups of households and population in order to compare the conditions and quality of life according to various social categories, such as gender, age, level of education, place of residence, social and professional status, main source of income, civil status, type of household (created on the basis of the number of families and biological family type), and other criteria.

The measurements of living conditions of the household include: household income and their way of managing income, nutrition, material affluence of the household, including modern communication technology equipment (mobile phone, computer, Internet access), housing conditions, social benefits received by the household, education of children, participation in culture and recreation, taking advantage of healthcare services, household situation on the labor market, taking advantage of social benefits, insurance and retirement security, poverty and other aspects of social exclusion.

Indicators of the quality of life and lifestyle of individual respondents included:

- general psychological well-being (including: the will to live, sense of happiness, satisfaction with life, depression),
- satisfaction with different areas and aspects of life, subjective evaluation of the material standard of living,
- various types of stress (including: stress associated with visits to public administration offices, stress associated with health condition, stress associated with parenting, financial stress, stress associated with work, ecological stress, marital stress, problems associated with taking care of older people, stress associated with traumatic events, such as assault, burglary, or arrest),
- psychosomatic symptoms (the measurement of distress treated as a general measurement of health conditions),
- strategies of coping with stress, evaluation of experience with the healthcare system, personal finances (including: personal income, insurance and retirement security),
- system of values, risk-seeking, lifestyle and individual behaviors and habits (such as smoking, excessive use of alcohol, use of drugs, or religious practices),
- civic attitudes and behaviors, social support, general evaluation of the transformation process and its influence on the lives of the respondents,
- use of modern communication technologies (computers, the Internet, and mobile phones),
- situation on the labor market and professional career, and problems of handicapped persons.

The distinction between social indicators of living conditions and the individual quality of life is more or less consistent with the distinction between the objective description of the situation (conditions) and its psychological meaning, expressed by the subjective opinion of the respondent (quality of life). This substantial distinction is generally consistent with the type of unit examined and the measurement method. For the living conditions, the examined unit is the household as a whole, and for the quality of life – its individual members. The living conditions were measured by conducting an interview with one representative of the household (a well-informed person; most often, it was the head of the household). The quality of life, on the other hand, was measured by a self-report questionnaire

addressed to all available members of the examined households who have reached the age of 16.

The sample for the survey was drawn through a two-stage stratified sampling method with different probabilities at the first selection stage. Census districts were the units of the first drawing stage, whereas households constituted the second drawing stage. National Official Register of Territorial Division of the Country - called TERYT served as a drawing statement. Before the draw, census districts were grouped according to the strata created with respect to region classes and subregions (NUTS3). In all subregions, three strata were created. The sample was divided between the strata through the allocation method proportional to the number of households in the strata. The allocation of a sample between the first and the second drawing stage resulted from an established number of second stage units (households drawn from previously chosen first stage units – districts). Districts were drawn with the probabilities of choice proportional to the number of households in the district.

The first measurement was conducted in the year 2000, and the subsequent one – three years later. The next two projects were conducted in two-year intervals. The present report shows not only the current image of Polish society; it also allows us to monitor changes in the same households and among the same people in a period of seven years.

2. Legal and methodological aspects of linkages

Central Statistical Office and the Border Guard work on introducing a joint research related to creating an inventory and merging of available sources of information on cross-border areas. The first stage of this venture would be the linkage between Border Traffic Register and a sample survey (Labor Force Survey, Household Budget Survey, Social Diagnosis)

The sets of randomly chosen households for the Labor Force Survey, Household Budget Survey, and Social Diagnosis are of homogeneous structure, that is, include an address, which consists of region's symbol, name of the region, street symbol, name of the street, household number and street number and are supplemented with a unique number, PESEL (Universal Electronic System for Registration of the Population), for every member of the household. The Border Traffic Register, which consists of a name and a surname, date of birth, nationality and passport number, will also be supplemented with the PESEL number. The PESEL number is the linkage key between sets of data. Subsequently, the Labor Force Survey and the Household Budget Survey's sets of data will be paired with the Border Traffic

Register. Thus, the database of households, whose members have crossed the border, will be built. Later on, an analysis of the identified households will be conducted, taking into account the data collected during the survey interviews.

In Poland, the Personal Data Protection Act regulates its rules of processing. According to the act, data processing is only allowed when the data subject signs his/her agreement to have their personal data processed. In case the data is collected from other sources, the administrator is obliged to inform the data subject. It has to be emphasized that this obligation is not required when the data are necessary for scientific, didactic, historical, statistic or public opinion research, and the processing of such data does not violate the rights or liberties of the data subject, especially when the fulfillment of the terms related to informing such a person would involve disproportionate efforts or endanger the success of the research. Hence, the new statistical study of linking the survey and administrative data could be realized by obtaining the PESEL number from administrative sources once such study constitutes a part of the yearly program of Polish statistics.

3. Expected results of the linkage

Central Statistical Office is conducting works on the creation of a coherent research system of the socio-economic phenomena occurring on the cross-border area. The survey on goods and services turnover in border traffic, at the whole external European Union border within the territory of Poland, is the element of the system. The research has shown that a considerable majority of people crossing the border live within the 50-kilometers border area (about 70% of surveyed Poles), and the purchase of goods is the main cause of the border crossing (about 90%). The percentage of persons who often cross the border is also substantial (80% of persons cross the border at least few times per month). These data, in an indirect way, prove the general opinion that a group of people living in the border areas, especially the unemployed, takes advantage of the neighboring border in order to gain considerable income or at least to purchase goods for their own needs for favorable prices.

The use of the existing administrative source of the Border Guard database and its linkage to the results of a conducted sample survey (Labour Force Survey, Household Budget Survey, Social Diagnosis) will enable to supplement the information on the situation of households on the borderland territory. It will provide the knowledge on the number and frequency of border crossing of individual members of households. Moreover, it will allow for broader characteristics of people crossing the border, taking into account chosen socio-

economic aspects based on data included in the Labor Force Survey questionnaire, for example: sex, age, level of education, status on the labor market, income, working experience, job search, and registration in the job agency. It will enable to verify the general belief that unregistered borderland trade can be a basic source of income. According to the standards of the Labor Force Survey, a person who often crosses the border and is engaged in unregistered trade is treated as unemployed. The verification of his or her status on the labor market can be achieved, for example, by comparing persons who answered "no" when asked the following question in the Labour Force Survey questionnaire: *Did you carry out any work bringing profit or income, or did you help in family business free of charge in the surveyed week* and simultaneously frequently crossed the border? The linkage of the information from the Border Traffic Register about people frequently crossing the border (and therefore with high degree of probability in engaging in an unregistered trade) to the information from the survey on unregistered work (module questionnaire Labour Force Survey) will help answer, indirectly, the question about reasons for such an activity. The comparison of those two sources of information will also help to check against the answer to the question about conducting unregistered work and its duration.

Additionally, linking the Border Traffic Register data to the information from the Household Budget Survey and the Survey of Goods and Services turnover in border traffic at the European Union's external border in Poland will broaden knowledge on the financial situation of people frequently crossing the border, especially in the scope of income and expenditures.

On the other hand, taking advantage of the Border Traffic Register database along with the Social Diagnosis results will enable for thorough description of households, taking into consideration the attitudes, state of mind, and behavior of the people constituting the households on the borderland. It will help determine, for example, if the individuals engaged in borderland trade experience higher level of life stress or are more resourceful in life. We will obtain the answer to the question if the general psychological well-being, including the will to live, sense of happiness, satisfaction with life, or symptoms of depression, has influence on professional activity in borderland areas.

Conclusions

The specificity of cross-border areas requires conducting numerous surveys of different kinds related to the socio-economic phenomena. The establishment and effective functioning of a coherent survey system requires for the system to be supplied

by standardized sources of information (office registers, other administrative sources, bank registers, automatic motion measurement, and so on), and production of projects that will cover not only border surveys, but also concentrate on the processes taking place in the border area.

In such a context, the analysis and the linkage of data from sample surveys to available administrative registers are immensely important. First and foremost, it reduces the survey burden put on respondents, and enables to identify scattered information sources owned by many institutions. The sources of information, after being inventoried and merged, constitute a new value. This venture is of a special significance for monitoring and the analysis of the processes taking place in the cross-border area, for which it is important to develop an interdisciplinary data backs.

The statistical public survey system references the opinions and behaviors of citizens only to a small degree, and is based more on micro and macroeconomic aspects of socio-economic phenomena. Therefore, when merging administrative registers with sample surveys, it seems important to join information from surveys such as the Social Diagnosis. In the future, it may lead to a development of new indicators, distinct for border areas, which will comprehensively describe real processes taking place in those areas.

Moreover, pairing registers from different institutions can be used for dynamic delimitation and for forming a typology of households participating in the survey (where the frequency of border crossing is an important criterion). Works in this field can constitute the basis for the expansion and change in sample selection in sample surveys, for example, adoption of additional stratifying criteria.

Linking information from different sources would also create the possibility of conducting works on the verification of the quality of results acquired from conducted surveys (for example, status on the labor market, unregistered jobs, household income) in addition to the factors connected with the border influence (border trade). These activities would enable to create a unique system, within the framework of public statistics, which would provide an extensive and complex information scope on socio-economic phenomena taking place at the external European Union border within the territory of Poland.

The creation of coherent survey system for cross-border areas would, in the first place, require an inventory of sources of information of public statistics and other administrative data sources in particular countries. A detailed analysis of this information and, most importantly, minor corrections in existing surveys (modification of answers and the questionnaire, as well as new modules) can be effective. It is worth emphasizing that the

introduction of new surveys is often very expensive. Therefore, what should be taken into consideration is the possibility of taking advantage of already existing sample surveys through sample and *objective scope* extension, so as to allow for generalization of the results for cross-border areas. A homogenous information infrastructure of cross-border areas (meta-information, databanks, methodological reports, and so on) would be the foundation of this system. Its formation should be coordinated by the public statistics.

References

Border Traffic and Movement of Goods and Services at the Polish-Ukrainian Border in 2009, Central Statistical Office and Statistical Office in Rzeszów, Warsaw and Rzeszów 2010.

Cierpiał-Wolan M., *Cross-border surveys – some methodological aspects*, Statistic in Transition new series, Vol.9, No. 2, Warsaw 2008, pp. 361-366.

Cierpiał-Wolan M., *The pilot survey of goods and services turnover in border traffic at the Polish-Ukrainian border – results and methodological aspects. Public Statistics for European Integration Processes with particular reference to issues concerning transborder areas*, Library of Statistical News, Vol. 60, Warsaw 2008, pp. 131-151.

Czapiński,J., Panek, T. Eds, *Social Diagnosis 2009. Poles Living Conditions and Quality of Live*, Ministry of labour and Social Policy, Warsaw 2009.

Household Budget Surveys in 2008, Central Statistical Office in Poland, Warsaw 2009.

Jenkins**,** S. P., Lynn, P., Jäckle, A., Sala, E., *The Feasibility of Linking Household Survey and Administrative Record Data: New Evidence for Britain*, Int. J. Social Research Methodology*, Vol. 11, No. 1, February 2008, pp. 29–43.

Labour Force Survey in Poland IV quarter 2009, Central Statistical Office in Poland, Warsaw 2010.

Liberda**,** Z. B., *How Statistics Shows the Crisis?* First International Conference, Quantitative Methods in Economics QME'2009, Warsaw University of Life Sciences, Century Manufaktura Center in Lodz, June 18-20, 2009.

Liberda, Z. B. and Grochowska A. Eds., 2009, *Civilizational Competences and Regional Development in Poland*, Warsaw University Press, Warszawa, pp. 292.

Liberda Z. B. and Pęczkowski, M. 2009, *Regional Distribution of Household Income and Household Human Capital* in: Civilizational competences and regional development in Poland, pp. 105-130.