



# STATISTICS IN TRANSITION

*new series*

*An International Journal of the Polish Statistical Association*

## CONTENTS

Editor's note and acknowledgements .....	425
Submission information for authors .....	429
<b><i>Development of the European Statistical System within Eastern Partnership – Directions and Strategy:</i></b> The international conference held on 18-19 October 2011 in Kraków	
Foreword by OKRASA W., CIERPIAŁ-WOLAN M., .....	431
<u>Opening speeches</u>	
WITKOWSKI J., CSO President's welcome address .....	435
BOHATÁ M., Opening speech on behalf of EUROSTAT .....	441
<u>Presentations on behalf of international organisations:</u>	
VOGEL F. A., BELKINDAS M. V. (WB), International cooperation through official statistics on food security, the environment, climate change and poverty .....	445
GAMEZ G. (ECE), Strengthening regional cross-border information systems: the example of Geneva cross-border region .....	453
<u>The Country Presentations</u>	
SMAILOV A., The integration of statistics of Kazakhstan with the European Statistical System .....	461
OLSSON B., Quality in official statistics using European Statistics Code of Practise as a framework .....	469
KOSTEVICH I., Statistical information system of the Republic of Belarus: experience and perspectives .....	475
ZBARSKAYA I., Statistical system of the countries-members of the Commonwealth of Independent States: 20 years on the way to integration (CIS-STAT) .....	485
OLENSKI J., Statistical parainformation platform – foundation of information transparency in heterogeneous international environment (with special reference to the Eastern Partnership initiative) .....	493
SPOIALA L., Quality – key element of development of official statistics .....	511
OSAULENKO O., Experience and expectations of the European Neighbourhood Countries: the example of Ukraine .....	519
POTISEPP P., SÖSTRA K., RANDLEPP A., Transformation of Estonian statistical system to the European statistical system – main aspects and experiences .....	527
CIERPIAŁ-WOLAN M., Directions for development of transborder areas – state and prospects .....	537
MATKOVSKYY S., Cooperation in the framework of transborder and regional statistics.....	547
DRÁPÁL S., New challenges within cross-border statistics .....	553
TSEKVAVA T., International cooperation and capacity development in Geostat .....	559
SIMCHERA V. M., KOYUNCUGIL A. S., Challenges to European statistics' credibility – why European statistics is still not transparent and its credibility unsatisfactory .....	567
<b>Other articles</b>	
MISRA G. CH., YADAV S. K., SHUKLA A. K., A class of regression type estimators in survey sampling .....	579
SINGH R., KUMAR M., CHAUHAN P., SAWAN N., SMARANDACHE F., A general family of dual to ratio-cum-product estimator in sample surveys .....	587
STRAWIŃSKI P., An improvement of quality of statistical matching for survey data using dynamic caliper .....	595

STATISTICS IN TRANSITION-new series, December 2011  
Vol. 12, No. 3, pp. 425–428

## EDITOR'S NOTE AND ACKNOWLEDGEMENTS

With the end of the year 2011 concluded was another year (which was the fourth year of being operated under the current Editorial Office) of our journal's successful cooperation with numerous members of the community of statisticians from all over the world – researchers, policy analysts and public statistics leaders and practitioners. This marks the firmness of the journal's mission and the vital role it plays as a platform for sharing ideas concerning statistical methods and their application, and for providing quantitative knowledge for better-informed decision-making by the relevant bodies in the government, business and general public. I would like to take this opportunity to express, also on behalf of the Editorial Office, our gratefulness to all the journal's patrons and supporters – with the Editorial Board chaired by Professor Janusz Witkowski, President of the Central Statistical Office, and the Polish Statistical Associations with its President, Professor Czesław Domański, and Associate Editors.

Also, I would like to express our sincere gratitude to all the journal's collaborators during the past year. While the major actors at the publication stage always are the authors of the articles, it is the silent cohort of their counterparts, peer-reviewers, whose skills and expertise make it possible to maintain the high quality standards of the journal. I would especially warmly like to thank the people who acted as referees of the papers submitted for publication during the past year for their anonymous contribution to the journal's continuous striving toward excellence. Their names are listed below, following the brief presentation of the contents of this issue.

\*

This issue is composed of two major sections. The first one contains papers presented at the international conference held on October 18-19, 2011, in Kraków as an event organized by the Central Statistical Office of Poland within the Polish Presidency of the EU. According to its title, *Development of the European Statistical System within the Eastern Partnership – Directions and Strategy*, the aim of the meeting was to bring the countries of the region closer to the standards of the European Statistical System, and to discuss various aspects of the possible cooperation and integration. More details are given in the introduction to this section.

The second section, which contains regular articles (submitted papers) starts with *A Class of Regression Type Estimators in Survey Sampling* by **Govind Charan Misra, Subhash Kumar Yadav, Alok Kumar Shukla**, who propose a class of linear regression models for the estimation of population mean and total when information regarding auxiliary variables is available in survey sampling.

They use regression method of estimation by introducing a new auxiliary variable  $z$ , which may also be a function of the auxiliary variable  $x$ . The proposed model leads to reduction in mean squared error as compared to ordinary regression method of estimation. The improvement has been demonstrated over ordinary regression estimator and also on ratio estimator with the help of an empirical example.

In the next paper, *A General Family of Dual to Ratio-Cum-Product Estimator in Sample Surveys* is discussed by **Rajesh Singh, Mukesh Kumar, Pankaj Chauhan, Nirmala Sawan, Florentin Smarandache** for the finite population mean. Under simple random sampling without replacement (SRSWOR) scheme, expressions of the bias and mean-squared error (MSE) up to the first order of approximation are derived. The authors show that the proposed family is more efficient than usual unbiased estimator, ratio estimator, product estimator, Singh estimator (1967), Srivenkataramana (1980) and Bandyopadhyaya estimator (1980) and Singh et al. (2005) estimator. An empirical study is carried out to illustrate the performance of the constructed estimator over others. In conclusion, the authors suggests that (for future research) the family suggested here can be adapted to double sampling scheme using Kumar and Bahl (2006) estimator.

**Pawel Strawiński's** paper *An Improvement Of Quality Of Statistical Matching For Survey Data Using Dynamic Caliper* discusses one of the key issue in evaluation studies which arises when matching procedure is used to estimate program net effects. The goal of the method is to establish a counterfactual state by choosing from the control pool a group that is similar to those in the treatment group. The author proposes a modification of the matching with caliper procedure while setting the caliper value as a fraction of estimated propensity score. The simulation results and examples are presented. Using data from Dehejia and Wahba (1999) some advantages of the proposed approach are stressed. The obtained results suggest that proposed approach is more efficient than the one traditionally used.

Włodzimierz OKRASA  
Editor-in-Chief

## ACKNOWLEDGEMENT TO REVIEWERS

The Editor and Editorial Board wish to thank the following persons who served since 31 December 2010 as reviewers of manuscripts for the *Statistics in Transition new series* – Volume 12, Numbers 1–3; the authors' work has benefited from their feedback.

**Aleksandra Baszczyńska**, University of Łódź, Poland

**Jacek Bialek**, University of Łódź, Poland

**Agata Boratyńska**, Warsaw School of Economics, Poland

**Katarzyna Bolonek-Lasoń**, University of Łódź, Poland

**Denis Conniffe**, University of Ireland - Maynooth, Ireland

**Czesław Domański**, University of Łódź, Poland

**Elżbieta Gołata**, University of Economics, Poznań, Poland

**Henryk Gurgul**, AGH University of Science and Technology, Kraków, Poland

**Alina Jędrzejczak**, University of Łódź, Poland

**Cem Kedilar**, Hacettepe University of Ankara, Turkey

**Jan Kordos**, Warsaw School of Economics, Poland

**Jerzy Korzeniowski**, University of Łódź, Poland

**Barbara Kowalczyk**, Warsaw School of Economics, Poland

**George Menexes**, Aristotle University of Thessaloniki, Greece

**Andrzej Młodak**, Statistical Office Poznań

**Włodzimierz Okrasa**, University of Cardinal Stefan Wyszyński in Warsaw, Poland and Central Statistical Office of Poland

**Walenty Ostasiewicz**, Wrocław University of Economics, Poland

**Iannis Papadimitriou**, Aristotle University of Thessaloniki, Greece

**Dorota Pekasiewicz**, University of Łódź, Poland

**Mariusz Pilch**, University of Łódź, Poland

**Waldemar Popiński**, Central Statistical Office, Poland

**Agnieszka Rossa**, University of Łódź, Poland

**Mauro Scanu**, ISTAT, Roma, Italy

**Divakar Shukla**, Dr. H. S. Gaur University, India

**Grażyna Trzpiot**, Academy of Economics, Katowice, Poland

**Vijay Verma**, University of Siena, Italy

**Jacek Wesolowski**, Warsaw University of Technology, Warsaw, Poland

**Leon Willenborg**, Statistics Netherlands

**Feliks Wysocki**, Poznań University of Life Sciences, Poznań, Poland

**Benhuai Xie**, Takeda Global Research and Development, Minnesota, USA

**Janusz Wywiał**, Academy of Economics, Katowice, Poland

**Janusz Żądło**, Academy of Economics, Katowice, Poland

STATISTICS IN TRANSITION-new series, December 2011  
Vol. 12, No. 3, pp. 429

## SUBMISSION INFORMATION FOR AUTHORS

*Statistics in Transition – new series (SiT)* is an international journal published jointly by the Polish Statistical Association (PTS) and the Central Statistical Office of Poland, on a quarterly basis (during 1993–2006 it was issued twice and since 2006 three times a year). Also, it has extended its scope of interest beyond its originally primary focus on statistical issues pertinent to transition from centrally planned to a market-oriented economy through embracing questions related to systemic transformations of and within the national statistical systems, world-wide.

The *SiT-n*s seeks contributors that address the full range of problems involved in data production, data dissemination and utilization, providing international community of statisticians and users – including researchers, teachers, policy makers and the general public – with a platform for exchange of ideas and for sharing best practices in all areas of the development of statistics.

Accordingly, articles dealing with any topics of statistics and its advancement – as either a scientific domain (new research and data analysis methods) or as a domain of informational infrastructure of the economy, society and the state – are appropriate for *Statistics in Transition new series*.

Demonstration of the role played by statistical research and data in economic growth and social progress (both locally and globally), including better-informed decisions and greater participation of citizens, are of particular interest.

Each paper submitted by prospective authors are peer reviewed by internationally recognized experts, who are guided in their decisions about the publication by criteria of originality and overall quality, including its content and form, and of potential interest to readers (esp. professionals).

Manuscript should be submitted electronically to the Editor:  
sit@stat.gov.pl., followed by a hard copy addressed to  
Prof. Włodzimierz Okrasa,  
GUS / Central Statistical Office  
Al. Niepodległości 208, R. 287, 00-925 Warsaw, Poland

It is assumed, that the submitted manuscript has not been published previously and that it is not under review elsewhere. It should include an abstract (of not more than 1600 characters, including spaces). Inquiries concerning the submitted manuscript, its current status etc., should be directed to the Editor by email, address above, or w.okrasa@stat.gov.pl.

For other aspects of editorial policies and procedures see the SiT429 Guidelines on its Web site: [http://www.stat.gov.pl/pts/15\\_ENG\\_HTML.htm](http://www.stat.gov.pl/pts/15_ENG_HTML.htm)



STATISTICS IN TRANSITION-new series, December 2011  
Vol. 12, No. 3, pp. 431–434

## FOREWORD

Włodzimierz Okrasa, Marek Cierpiał-Wolan

*Development of the European Statistical System within Eastern  
Partnership – directions and strategy*

International Conference, 18-19 October 2011, Kraków, Poland

An international conference “Development of the European Statistical System within Eastern Partnership – directions and strategy” was held on 18-19 October 2011 in Kraków, Poland. It was organized by the Central Statistical Office of Poland in cooperation with the Ministry of Foreign Affairs as one of the central events of the Polish Presidency in the Council of the European Union, under the honorary patronage of the Deputy Prime Minister and the European Parliament.

The main goal of the conference was to discuss possible challenges to, and put forward the more advanced ideas for strategies of bringing the partners of EU Member States from Eastern Europe, Caucasus and Central Asia closer to the European system of official statistics, and to develop a mechanism for maintaining the growing cooperation between the two groups of countries.

The vital importance of such an ambitious goal was emphasized by Professor **Janusz Witkowski**, President of the Central Statistical Office in his welcome speech; he has also forwarded greetings sent to the meeting participants by the highest officials of the Polish Government and of the European Parliament, while stressing the significance of the presence of EUROSTAT, represented by the Eurostat’s Deputy Director, Marie Bohatá, and of other international organizations.

Two groups of the papers have been presented during the meeting sessions. The first one comprises papers presented on behalf of the international organizations: EUROSTAT, The World Bank, the United Nations Economic Commission for Europe; the other consists of country presentations.

Involvement in extensive cooperation which aims at harmonisation of statistical methods, comparability of statistics and strengthening of state and cross-regional information systems between countries covered by the Eastern Partnership were the key aspects of speeches presented by representatives of international organizations. **Marie Bohatá** from Eurostat pointed to numerous activities carried out with the participation of Eurostat, which seek to build closer relationship with all European countries, among others, High-Level Seminars for 11 countries from Eastern European and Central Asian countries, which also



covers the project on Adapted Global Assessments supported by of EFTA and UN/ECE, short-term technical assistance missions, as well as seminars on subject matter areas and Twinning projects. Busan Action Plan for Statistics, a new initiative which aims to improve the statistical systems in developed and developing countries was discussed **Misha Belkindas** and **Frederic Vogel** from the World Bank, whereas the example of developing cross-regional information systems in the Geneva cross-border region was presented by **Gabriel Gamez** from the United Nations Economic Commission for Europe.

The development of statistical systems from country perspective, with the focus the important role that the European Union and international organizations such as UNECE, OECD, IMF and World Bank played in supporting these advancements, was highlighted in the presentations of representatives of statistical institutions from Russia, Georgia and Belarus. **Irina Zbarskaya** representing the Interstate Statistical Committee of the Commonwealth of Independent States described the processes of integration of the statistical systems of the Member States of the CIS. **Tengiz Tsekvava** of the State Statistical Office of Georgia, apart from stressing the impact of international cooperation on the development of potential of the Georgian statistics, spoke about the problem of outflow of skilled workers from the statistics because of low wages in the department. **Irina Kostevich** from the National Statistical Committee of the Republic of Belarus presented the experiences and perspectives of statistical information system of the Republic of Belarus, with particular focus on creating a harmonized information system within the Eastern Partnership.

The issues of new challenges facing cross-border statistics were the central themes of presentations by **Marek Cierpiał-Wolan** from the Statistical Office in Rzeszow and **Stanislav Drpal** from the Czech Statistical Office. The former pointed to developing a coherent system of surveying cross-border areas and outlined Polish public statistics' experience in monitoring the socio-economic phenomena, whereas the latter one highlighted the important role of the exchange of experience and knowledge between countries.

Different approaches to problems associated with establishing a coherent information system of the Eastern Partnership were considered in the next group of papers. **Berit Olsson** from the Swedish Statistical Office focused on the principles of the European Statistics Code of Practice as the ones of crucial importance to the public statistics quality. Professor **Józef Oleński** from the University of Warsaw presented novel concept of statistical parainformation platform used as a tool for exchanging information in order to provide their transparency - through the observance of standards and standardized methodology - by all the parties involved in the statistical processes. **Vasily Simchera** from the Russian Academy of Economic Science and Custom Union of the EurAsEC referred to challenges that lie ahead of the reliability of statistics, in particular, to imperfections in the functioning of European statistics, especially with regard to information at different levels of aggregation, and their consistency. **Rune Rafaelsen** from the Norwegian Barents Secretariat addressed the need to develop

an index for a number of cross-border regions in Europe that takes into account economic, social and cultural processes.

The presentations that were devoted to the directions of transformation of statistical systems in the Eastern Partnership countries and their harmonization with the European Statistical System were presented by representatives of Estonia, Ukraine, the Republic of Moldova and the Republic of Kazakhstan. **Priit Potisepp** from Statistics Estonia spoke about the process of centralization of Estonian official statistical system, updating the legal-institutional basis and implementing the statistical programme. **Oleksandr Osaulenko** from State Statistics Service of Ukraine discussed the main aspects of the National Statistical System Transformation Process and priorities of further development. **Lucia Spoiala** from the National Bureau of Statistics (NBS) of the Republic of Moldova elaborated on adherence of NBS to the European Statistics Code of Practice as well as on key problems associated with that. **Alikhan Smailov** from the Agency of Statistics of the Republic of Kazakhstan (ASRK) referred to activity and legal basis of the Agency.

In his closing speech, the President of CSO **Janusz Witkowski** expressed the hope that it was a kind of platform to exchange experiences and to present current issues of the statistical services of different countries. The President said that the contacts made during the conference will help in the future to better coordinate the necessary actions for moving towards the integration of the statistical systems of the Eastern Partnership members countries with of the European Statistical System.

Below is the Agenda of the conference:

### **Session 1: Official statistics as a dynamic form of cooperation between Eastern Partnership countries**

Chair: **Lidia Bratanova** (UNECE Statistical Division)

1. **Misha Belkindas** and **Frederic A. Vogel** (World Bank) - "International Cooperation through Official Statistics on Food Security, the Environment, Climate Change and Poverty".
2. **Irina Zbarskaya** (Interstate Statistical Committee of the Commonwealth of Independent States) - "Statistical system of the countries-members of the Commonwealth of Independent States – 20 years on the way to integration".
3. **Tengiz Tsekvava** (National Statistics Office of Georgia) - "International cooperation and capacity development in Geostat".

### **Session 2: Role of official statistics in establishing a coherent information system within the frame of Eastern Partnership**

Chair: **Lars Thygesen** (Statistics Denmark)

1. **Berit Olsson** (Statistics Sweden) - "Quality in official statistics – using European Statistics Code of Practice".

2. **Irina Kostevich** (National Statistical Committee of the Republic of Belarus) - “Statistical information system of the Republic of Belarus – experience and perspectives”.
3. **Józef Oleński** (University of Warsaw, State High School of Technology and Economics in Jarosław, Poland) - “Statistical parainformation platform – foundation of information transparency in heterogeneous international environment (with special reference to the Eastern Partnership initiative)”.
4. **Vasily Simchera** (Russian Academy of Economic Science, Statistics Division of the EurAsEC Customs Union, Russia) - “Challenges to European statistics’ credibility – why European statistics is still not transparent and its credibility unsatisfactory”.

### **Session 3: Cooperation in the field of cross-border and euro-regional statistics**

Chair : **Ms Irina Zbarskaya** (Interstate Statistical Committee of the Commonwealth of Independent States)

1. **Marek Cierpiał-Wolan** (Regional Statistical Office in Rzeszów, Poland) - “Directions for development of transborder areas – state and prospects”.
2. **Semen Matkovskyy** (Main Statistical Office in L’viv Region, Ukraine) - „Cooperation within the cross-border and regional statistics”.
3. **Gabriel Gamez** (UNECE, Statistical Division) - “Strengthening regional cross-border information systems – the example of Geneva Cross-Border Region”.
4. **Rune Rafaelsen** (The Norwegian Barents Secretariat) - “Poland Border Dialogues. Developing a CBC Index for Europe”.
5. **Stanislav Drápal** (Czech Statistical Office) - “New challenges within cross-border statistics”.

### **Session 4: Transformation of statistical systems within the Eastern Partnership countries – towards harmonization with the European Statistical System – session organized by Eurostat.**

Chair : **Mr Pieter Everaers** (Eurostat).

1. **Priit Potisepp** (Statistics Estonia) – “Transformation of EstSS to the ESS – main aspects and experiences”
2. **Oleksandr Osaulenko** (State Statistics Service of Ukraine) - “Experience and expectations from the European Neighbourhood Countries – the example of Ukraine”.
3. **Lucia Spoiala** (National Bureau of Statistics of the Republic of Moldova) - “Quality – a key element of development of official statistics”.
4. **Alikhan Smailov** (The Agency of the Republic of Kazakhstan on Statistics) - “The integration of statistics of Kazakhstan with European Statistical System”.

*STATISTICS IN TRANSITION*-new series, December 2011  
Vol. 12, No. 3, pp. 435—439

## **CSO PRESIDENT'S WELCOME ADDRESS**

**Janusz Witkowski**

Distinguished Guests, Dear Colleagues and Friends of Statistics, Ladies and Gentlemen,

It is my great pleasure to welcome you to this conference. As its co-organiser, I have the honour to host you in the beautiful city of Kraków, the uniqueness of which stems from its long and glorious history marked by splendid and original monuments, and from hearty hospitality of its citizens. As the former capital of Poland and one of the oldest academic centres in Europe, Kraków has become a fascinating place for many prominent artists and scientists. Its local atmosphere is conducive to reflections on the past inspiring also to forward-looking thoughts. Therefore, it seems an ideal place to host our conference. I hope that all of you will feel well accommodated, and that you will spend a pleasurable and unforgettable time in this extraordinary city.

We have gathered here to deliberate on the subject matter that is of key significance, both to common Europe-wide system of statistics and to each country-specific statistical system. I would like to greet all the attendees to this conference, devoted to the development of the European statistical system among the members of the Eastern Partnership. Thank you for your interest and for coming in such a great number, representing so many different countries and international institutions and organisations.

The importance of this event is also emphasized by the honourable patronage of the European Parliament, confirmed by the signature of Professor Jerzy Buzek, President of the European Parliament, and the honorary patronage of Mr. Waldemar Pawlak, Deputy Prime Minister and Minister of the Economy. In recognition of the weight of this initiative, the President of the Republic of Poland, Mr. Bronisław Komorowski, has also addressed this event's participants and organisers. Finally, the interest expressed by EUROSTAT, which organises one of the sessions, also underscores the importance of this meeting.

The main focus of this conference relates to one of the priorities adopted by the Polish Government, i.e. "The Relationship with the East." For this reason, it has been entered onto the list of the official events under the Polish Presidency. Its thematic scope is of utmost practical importance to the implementation of this priority, and to the strengthening of international cooperation at the European level. It is beyond any doubt that the process of strengthening our cooperation,

identifying development strategies in various countries, and coordinating them on the European scale, as well as assisting and cooperating in their implementation, and contributing to European integration, should be based on an accurate, modern and harmonised system of official statistics. Therefore, our aim will be to debate about directions and strategies to be adopted in order to overcome the challenges of bringing closer European official statistical systems and enhancing cooperation between EU Member States and their partners from Eastern Europe, Caucasus and Southern Asia.

While emphasising the need to pursue partnership cooperation, in order to upgrade the statistical systems in individual countries and their integration within common European statistics, one should agree to the necessity to get through a number of problems being faced by global and European statistics. The responsibility for facing those challenges is shared by the entire statistical community and it essentially requires adequate international cooperation. Such conclusions arose from the ISI Satellite Conference, which took place in Kraków two months ago and was devoted to Improving Statistical System Worldwide – Building Capacity.

The aforementioned conference - as well as the discussions held at other international meetings, including the recent World Statistics Congress in Dublin - indicates that each region and country is struggling with various statistical problems, which finds confirmation in our individual experience. This also concerns the European Statistical System, which is tackling new challenges.

The dynamics of socio-economic changes in the contemporary world, the supra-national threats of economic crisis, and the globalisation of economies and social life patterns, have highly elevated overall expectations towards official statistics.

This refers both to the scope of information and to the methodological standards, as well as to information comparability patterns in the international dimension. And it provides an incentive for the international community to intensify discussion on how official statistics should be modernised and improved. These issues fall down into the general expectation, which is to strengthen the essential and continually-increasing role of official statistics in today's world. Quality statistics provide the basis for an accurate diagnosis of the socio-economic situation, indispensable to monitoring the developing changes. It also constitutes a crucial element for the decision-makers and expands our knowledge of the reality we live in.

Such expectations are resulting in expanding statistical objectives, obliging us to implement innovative methods of data collection, processing and dissemination tools. At the same time, it is necessary to ensure a decreased statistical burden on respondents and good quality of statistical output.

The necessity to improve, modernise and revitalize the European Statistical System has been reflected in a number of clear-cut common European measures. Such types of measures are, however, both important and demanding, as the statistics in individual countries differ in their level of development - which

requires different ways of redevelopment or modernisation - and different support from various international institutions and organisations, through strong bilateral (partnership) cooperation. In addition, they function under diversified conditions, being determined by, among others, diverse historical experience. They also differ in methodological awareness and research experience, as well as in institutional arrangements, the type of legal regulations concerning statistical operations, and the scope of information produced and disseminated. Such a situation creates considerable challenges related to redeveloping and harmonising European statistics, including international cooperation and the anticipated effective support given to various groups of countries.

Irrespective of State borders, political divisions or location of our countries, socio-economic systems or other disparities, we all constitute one statistical family and we should aim at creating a consistent European statistical system, regardless of the fact that the ways of achieving this objective may be different. While respecting the historical distinctiveness and values of national statistical systems, we should all focus on accurate official statistics. In consideration of those objectives, we should establish a system of closer relationships with each other, learn about our respective problems and jointly work on suitable solutions, providing mutual support. **This will be our primary intention and the essence of this conference.**

\*

As a comprehensive understanding of the statistical systems in various partner countries paves the way for effective cooperation, it is, therefore, essential to determine specific stages of cooperation. This should start with identifying the key characteristics of statistical systems of partner countries, in terms of their strengths and weaknesses. This would lay the ground for creating development strategies, and for improving national statistics, by way of constructing a cooperation programme with all the elements necessary for it to work, including financial support. Afterwards, we should proceed with creating consistent cooperation, i.e. with the implementation of our common development strategy.

The principles of cooperation are essential to the successful modernisation of any statistical system. Representing Polish official statistics, we can state this with all our might, as we have ourselves experienced immense international support in the process of redeveloping Polish statistics during the transition period. That is, how we have learnt that such cooperation is both advantageous and effective, which is well-proven by the current shape of our statistics. We have benefitted from this support for many years in many ways, through a variety of forms, including bilateral (partnership) support, international projects, twinning agreements, study and expert visits, training sessions and workshops, and institutional support (from Eurostat, the World Bank, ILO, and FAO). The type and form of cooperation should reflect the needs of the partner countries and the specific stage of development in their respective statistical systems.

\*

For many years now, we have pursued cooperation with several Eastern European countries. Our principal objective is to share our experience with other countries which are redeveloping their official statistics, with a view to creating a system that is consistent with European Statistical System, what also is one of the priorities of Polish official statistics. We hope that this cooperation is beneficial and effective for both sides. For our partners, due to our familiarity with their statistical systems and with the redevelopment (modernisation) related problems they are facing. It is also beneficial, for a number of reasons, to Polish official statistics, as it fosters the consolidation of research-oriented working relations between our countries. The best example of it is being provided by cross-border research cooperation, which has led to interesting results in the surveys in fields which have not yet been sufficiently covered by official statistics so far.

Let me stress that cross-border cooperation is a very promising area for future cooperation in the field of regional statistics, due to the growing demand for such information among regional and local authorities. Our previous cooperation experience regarding regional statistics is very encouraging. It takes various forms, starting with joint publications, informational materials and databases, through jointly-organised conferences, seminars and substantive meetings, up to joint statistical surveys on border traffic and movement of goods and services. These are the fields in which our colleagues from regional statistical agencies cooperate with statisticians from neighbouring countries, much to our appreciation. I hope that our cooperation will continue, to the benefit of all partners. Therefore, we can draw a conclusion that it is necessary to strengthen cooperation with neighbouring countries and with countries having common roots of their statistical systems.

\*

For all these reasons, our conference – its topic and time and place - seems particularly well destined to play a significant promotional role. It is also meant to inspire further activities, aimed at improving statistics in Eastern Partnership countries. I hope this conference will provide a valuable insight, as there are still many essential questions to be answered in the foreseeable future, including:

- What needs to be done in each of the participant country to integrate them into European system of statistics, without a risk of being deprived of their historical identity and distinctiveness?
- How can international cooperation be coordinated in most effective way?
- What kind of target objectives should be attributed to national and international institutions, including Eurostat (programme-based cooperation, as well as methodological and educational cooperation), and in what form should they be produced under the current socio-economic conditions?

The conference programme is structured so as to attempt to reach practically relevant answers to those questions. And I hope that it will set-up directions for facilitating our further cooperation in efforts toward improving official statistics.

Let me once again welcome you to this meeting. Thank you for coming. I wish all of us have interesting discussions, inspiring further activity, and allowing us to create an efficient vision of international cooperation, aimed at excelling statistical systems in our respective countries, and fostering the development of common European statistics.

I wish you a good stay in Poland, and especially here, in Kraków.





*STATISTICS IN TRANSITION*-new series, December 2011  
Vol. 12, No. 3, pp. 441—443

## **OPENING SPEECH ON BEHALF OF EUROSTAT**

**Marie Bohatá<sup>1</sup>**

The Eastern Partnership – an initiative launched by Poland and Sweden – should contribute to the development of stability, better governance and economic development at the Eastern borders of the European Union and aims at building an increasingly closer relationship with all European countries. It is an important initiative for the European Union and for the European statisticians in particular, the proof of which is the organisation of this Conference.

Eurostat is actively involved in and highly encourages cooperation with our colleagues in the national statistical institutes of the countries covered by the Eastern Partnership. The main aim of this cooperation is harmonisation of statistical methods and data with European standards but also support to institution building and infrastructure development in statistics.

One of our flagship initiatives is the organisation of annual High-Level Seminars for 11 countries from Eastern European and Central Asian countries during which topical issues relevant for the top management of the national statistical institutes are discussed. It is a forum at which the participating countries can exchange their opinions and ideas on the improvement of their respective statistical system. The topic of the 2010 seminar was the coordination of the national statistical system while this year the participants discussed various aspects of the quality in statistics. The EU member states, the Commission/Eurostat, EFTA and UN/ECE also contribute with their experience to those seminars.

Both topics that were dealt with at those seminars are of crucial importance. Coordination was stressed in the ESGAB report and is one of those issues that are considered for the envisaged revision of Regulation (322) on the ESS.

Sponsorship on quality deals with the common ESS quality assurance framework and the alignment with the ESCB quality management.

Eurostat is also implementing – with the support of EFTA and UN/ECE – a project on Adapted Global Assessments of the statistical systems in these countries. These assessments cover all institutional aspects of a statistical system, the cooperation with other authorities and the coordination role of the NSI but also a number of statistical areas. The European Statistics Code of Practice and its principles are the basis of these assessments as well as international and European standards for the production of statistical information. The outcome of such an

---

<sup>1</sup> Eurostat, European Commission.

assessment is a report containing a number of recommendations for improvements for institutional aspects (e.g. the independence of the statistical office, quality management, the confidentiality principle, the websites and dissemination of statistics and others) but also for statistical production processes and outputs. These recommendations will help the partner countries to improve and come closer to the principles enshrined in the European Statistics Code of Practice and to European standards. Assessments were already done for Armenia, Azerbaijan, Kyrgyzstan and Kazakhstan, one assessment is ongoing for Ukraine and others will be implemented in Moldova, Belarus, Tadjikistan and possibly Georgia in 2012.

It has to be mentioned that this is not a one-off exercise: the CoP is a living tool, which has been recently amended. Some principles, especially professional independence, should become even legally binding.

Other cooperation activities include short-term technical assistance missions, seminars on subject matter areas and Twinning projects. Currently, Twinning projects in statistics are implemented in Armenia, Azerbaijan and Ukraine and other countries consider applying for it. Eurostat also takes note and encourages bilateral cooperation between the Member States and Norway and individual countries in the region. There are good examples of such cooperation, like the involvement of Statistics Sweden (SIDA) in the development of the national statistical system in Georgia or the technical assistance provided by Statistics Norway to the Moldovan NSI.

Eurostat actively participates in the process of negotiation of the Association Agreements which are held at present individually with Armenia, Azerbaijan, Georgia, Moldova and Ukraine. Each Association Agreement contains a chapter on statistics which specifies the basis, the areas and the ways and means of the cooperation in statistics. For all the countries the chapter on statistics is provisionally closed but this is only the beginning of a closer cooperation with the aim to come closer to the European statistical system and its members.

I briefly mentioned all those activities and forms of cooperation, because I believe that this Conference is a competent forum to discuss in detail further cooperation in the framework of Eastern Partnership. We have a unique opportunity to discuss the far reaching visions, intermediate objectives and short-term goals related to the cooperation with the region and in the region. We can share our ideas about the best ways of working together and the results we would like to achieve in the foreseeable future.

We are all aware that in addition to the capacity building of the national statistical systems needed to face the challenges of the modern societies, to strengthen trade ties and to support the sustainable economic growth, especially at the time of global financial crisis, statistics has an important role to play in all areas of social, political and economic life. It supports the aspiration of deepening democracy and establishing a civil society. It is the backbone of several sectoral policies. We are witnessing that the needs for European Statistics (ES) are constantly increasing and their nature is changing as they are massively used for

EU policy making. Their importance and volume are growing, and they are becoming increasingly complex. EU policy is addressing more complicated issues than in the past and also European governance is changing. There is currently much more interest on the side of politicians in statistics which makes us proud of our work on the one hand, but on the other there are also some problems and challenges linked to that.

Moreover, some statistics are directly used for regulatory or administrative purposes and this specific usage has important implications for quality of statistics/indicators.

More and more we have to bear in mind that quality of official statistics is not an absolute feature to be dealt with in a stand-alone manner. Instead, it is relative, defined by the users' needs, and multifaceted, and thus should be dealt with in a context and in an integrated manner.

We as official statisticians are not very much at ease with the consequences of these new developments. The new challenges linked to proximity of ES to politics and to their high visibility raise concerns regarding our independence and credibility. Statistics are an important tool for policy makers, businesses, scientists and the general public. Users need to be well informed and aware of the strengths and the weaknesses of statistics they use. This is not an easy and straightforward task, as we have to explain not our compliance with quality standards, but also the actual standards themselves. These are not set in stone and they need to be continuously refined in order to reflect new developments in statistics and in our societies.

I hope that during this Conference we will create ideas to bring about closer regional cooperation and further integration of statistical activities in the region, to further strengthen the national statistical systems and to intensify the contacts between the region and the key international players in statistics. In the margins, we can also brainstorm on a full-fledged implementation of the fit-for-purpose paradigm for official statistics.



STATISTICS IN TRANSITION-new series, December 2011  
Vol. 12, No. 3, pp. 445—452

## **INTERNATIONAL COOPERATION THROUGH OFFICIAL STATISTICS ON FOOD SECURITY, THE ENVIRONMENT, CLIMATE CHANGE AND POVERTY**

**Frederic A. Vogel, Misha V. Belkindas<sup>1</sup>**

### **ABSTRACT**

The ongoing debate about the causes and effects of food price volatility, especially on the number of people living in poverty, is occurring at the same time there is increased concern about the effect of agriculture on the environment and global warming. Practices to increase food production have consequences on the environment and global warming. From a statistical point of view, these are issues that transcend national boundaries. International cooperation will be essential to resolve these issues; it is dependent on official statistics.

This paper will describe statistical tools that could be used to describe the intertwined relationships between food price volatility, food security, poverty, the environment, and global warming. These statistical tools and outcomes need to be developed to provide results comparable across countries. The lessons learned from the well-known International Comparison Program will be applied to suggestions how to develop official statistics about the above issues to shape policy decisions across national boundaries.

The current world financial crisis that came on top of the food price crises is evidence enough about the globalization of the world economies with decisions made by one country that cause a rippling effect across the world. More than ever before, good statistics are needed to deal with these issues and the stakes are even higher to make sure that the information be comparable across national boundaries.

The best example of international cooperation through official statistics is the International Comparison Program (ICP) which is also the world's largest and most complex international statistical program. The next section provides a brief overview of the cooperative effort required. That is followed by a review of how this can impact other areas of official statistics. The paper concludes with recommendations on the way forward.

---

<sup>1</sup> The World Bank.

## **The ICP Standard for International Cooperation**

The ICP involves an international collection of prices for an agreed upon basket of goods and services. These prices must be for products representative of each country's consumption, but also comparable across countries. Agreed upon methodology is used to compute Purchasing Power Parities (PPPs) that are a form of an exchange rate used to convert national estimates of their Gross Domestic Product (GDP) and aggregates into a common currency. While there are many uses of these PPPs, a main one is to establish the international poverty lines. The cooperative nature of the ICP is described in Vogel (2009).

*No country can produce PPP comparisons with other countries by itself. The ICP is a joint effort where countries work together to determine what will be priced and when. The data analysis and estimation methods require that data be pooled across countries. The very essence of the ICP is the comparability of results between countries, strict adherence to time schedules, and a common understanding of data sharing and confidentiality requirements. There is no other statistical program requiring so much cooperation between national, regional, and international organizations.*

*The Consumer Price Index (CPI) provides a useful example because it shares a common technical language and conceptual framework with the ICP and the output of national accounts. Each country determines the scope and coverage of its CPI from the content of the product basket, to the coverage of the country and all other aspects leading to the publication of the results. Each country has full responsibility for its final CPI estimates for which it is accountable to its public and other governmental entities. They are also accountable for ensuring its country's data confidentiality rules are followed.*

*The ICP takes the process to another level where decisions are shared with other countries, a regional coordinator, and the ICP Global Office. Purchasing Power Parities are estimated by price comparisons between countries; therefore, inter-country coordination is required. Each country must abide by standards accepted by other countries and follow the same data collection and national accounting procedures. An essential feature is that they are required to send their data to a regional coordinator where they are subject to review by other countries. Regional coordinators then send data on to the Global Office where the data review ensures consistent procedures are being used across regions. A sense of partnership and overall trust had to be established between countries and regions for assurance that other countries and regions were applying the same guidelines and standards. Countries had to follow similar methods and procedures in spite of the fact they differ widely in size, culture, and diversity of goods and services available to their people. They also have different levels of statistical capacity. Furthermore, not every country speaks the same language adding another dimension to the requirement for comparable methodology and procedures.*

The distinguishing feature of the ICP is that each country sends its prices off to the regional and global coordinators and has to await their review before it knows its PPP levels. The process works because of the complete transparency of the methods and procedures used to compute the PPPs. Each country is allowed to review not only their data, but also the data for other countries to challenge inconsistencies. This provides trust in the system which is essential.

### **International Cooperation in Official Statistics—other candidates.**

The wide use of PPPs points to other but related areas in national official statistics where the conceptual framework shifts from producing national statistics to measures at regional and global levels. Examples follow.

*The Consumer Price Index.* The Consumer Price Index (CPI) is an index that measures the month to month or quarterly rate at which prices of consumption goods and services are changing. It provides a measure of price inflation also widely used as a proxy for a general index of inflation for the economy as a whole. It is a statistic for economic policy making, especially monetary policy. The CPI is widely used for index linking of social benefits such as pensions and other government payments, and also as escalators for measuring changes in GDP growth rates. Benchmark PPPs are extrapolated forward for non benchmark periods using these growth rates relative to the US growth rates. This has an impact at the international level. For example, international poverty lines are converted to the national level using benchmark PPPs, then national poverty rates for non benchmark years are a result of the extrapolation. These extrapolated PPPs are also widely used by health and education organizations and also those making cost of living assessments. Experience has shown that these growth rates are not consistent with structural changes taking place over time; therefore, there is a divergence in CPI price levels and PPPs that increases as the time between benchmarks increases. As a result a gap in the international data system is a set of CPIs' harmonized across countries. Ideally, countries within a region would determine a set of products and services in common between them which would be priced.

The theory and methods for the CPI are well developed and fully documented in the Consumer Price Index Manual (2004.) However, each country independently prices a set of nationally determined goods and services with different base periods making it nearly impossible to compare rates of inflation taking place over a group of countries. A recent paper by Biggeri and Laureti, (2011) presented the properties of the CPI and PPP estimators and concluded that a complete integration of the PPP and CPI data collections and estimations is essential to obtain increased coherence between PPP and CPI results. This would allow PPPs to be based on the more frequent CPI data collections and avoid the difficulties linked to the use of CPI's for temporal adjustments in PPPs.



*Household Surveys.* Household surveys such as the World Bank Living Standard Measurement Studies are the most effective way to obtain information related to well-being and poverty. Household surveys are used to determine the number of people living below the national and eventually the international poverty line. Household surveys provide the main input to the underlying weights used in the estimation of the Consumer Price Index across the major aggregates. These weights need to be established in a way to ensure more comparability of the CPI's across countries.

The household surveys are complex and expensive; in developing countries they are often donor driven, of an ad hoc nature with little coordination of the activities of different donors. The World Bank together with several other international organizations formed the International Household Survey Network (IHSN) to develop tools helping increase comparability of surveys, as well to contribute to better coordination between donor agencies and the receiving organization.

The scope, content, and coverage of household surveys needs to be made consistent and comparable across countries. A minimum set of core data needed should be identified, and a standard set of survey methodology identified.

*Food Security.* Food security involves a wide set of data and indicators ranging from calorie and nutrition intake to the cost of food. However, the single most important set of information required to make policy and marketing decisions affecting food security are timely estimates of current supplies and forecasts of production prior to harvest. Data on maize production that becomes available a year after harvest is of no use with a surplus crop rotting in storage because it is too late to seek export markets, or a crop disaster led to people either starving and/or staging food riots.

Remote sensing technology provides a powerful way to monitor crop conditions and obtain early warnings of disasters. Many of the weather satellites contain a monitor that provides a vegetative index which is a single number that quantifies plant biomass and vigor. While the main use is for early warning purposes, research is underway to use the indices for crop yield forecasting. (Panda, et. al. 2010).

While the methodology is complex preventing its use by many countries, groups of countries could combine resources whereby a regional remote sensing organization provided the satellite imagery and early warning materials that each country could use. Because the satellite imagery crosses national boundaries, countries with overlapping production areas and water sources would each be able to assess the full picture.

*Monitoring the Environment and Climate change.* These are complex issues which like food security involve a large array of indicators. However, one standard that is comparable across the world is land use and the capability to monitor it using satellite imagery. It can show changes in land cover and land use over time, analysis of population density, distribution, and growth. It provides analysis of watersheds, another example of the many uses.

Again, the complex methodology may prevent its use in many countries, but this is an area in which cooperative efforts across countries would enable their use of the technology.

*Improved Statistical Methodology.* The above technology will allow a total rethinking of statistical methodology used by national statistical systems. The imagery described above can be used to provide classifications of urban and rural areas into strata based on population density. This could be used as a master sample frame for most of the data collections made by a country. The use of periodic population censuses could be replaced by a continuous cycle of household sample surveys that provide current up to date data needed at the time in addition to the information about population provided by vital registration systems.

*Data Collection Through Mobile Phones.* The need of comparable cross – country and sub-regional data requires not only cooperation of statistical agencies but employment of new data collection methods and tools. One of these endeavors is a pilot study for Crowd-Sourced Data Collection through Mobile Phones.

Crowd-sourcing is the act of taking a job traditionally performed by a designated agent and outsourcing it to an undefined, generally large group of people in the form of an open call. Crowd-sourced data are collected and reported by the user community using commonly a mobile phone. At least two surveys within the World Bank can potentially benefit from this new method of data collection.

The first one is the International Comparison Program (ICP), which as described above is a worldwide statistical partnership to collect comparative price data to estimate purchasing power parities (PPPs) of the world's economies which are used to and compile detailed real expenditure values of countries' gross domestic products (GDP). The overall project is led and coordinated by the ICP global office at the World Bank and data are collected by the national statistics offices of individual countries and or regions. These data are currently collected around every 6 years. The prices needed must be the national annual average; this can require monthly price collections across the country for food products. The cost of the latest round is estimated at around \$37 million. It is important to look for means to improve data quality and increase the frequency of data collection while lowering the overall program costs.

The second related survey is food price monitoring. Prices for food are measured systematically mainly from the producer side, for example, with the Commodity Markets Review. In the current economic environment of rising food prices, a need to measure food prices from the consumer side as well is becoming equally essential. Currently there is no access to reliable and frequently produced retail price data for food commodities.

The objective of the pilot is to study the feasibility of crowd-sourced data collection especially for the uses of the ICP and food price monitoring. The method involves usage of non-professional price collectors (NPCs) and mobile

phones as the means to collect price data for the food commodities within 9 pilot countries. The World Bank has contracted a private company, txteagle, to conduct the pilot study. The pilot countries are Brazil, Egypt, India, Indonesia, Jordan, Kenya, Nigeria, Uganda and Zambia. The survey framework for these countries is developed both by the World Bank and txteagle.

Using the crowd-sourcing method, prices are collected for 30 basic food commodities, such as rice, vegetables, and sugar. The World Bank has provided specifications for these products using the ICP product list as a base. Each product specification includes 7-15 price determining parameters.

Central to the activities carried out is the web micro-site, developed specifically for the pilot. Communication between the consultancy and the NPCs, for training and data submissions are channeled through the site. Additional means to recruit and communicate are email, Skype, and Twitter.

The results, i.e. price data, reports and gained experience, will provide insight to the potential use of alternative ways to sample and collect price data across the countries. Depending on the results of the pilot, the developed platform can potentially be used for a wide variety of applications in individual countries and the international statistical community.

## **The way forward**

The Global Strategy to Improve Agricultural and Rural Statistics (2010) provides the framework to meet data demands for emerging data requirements. The Global Strategy contains several elements that relate to the initiatives presented above and also to the role of the international statistical system. The Global Strategy outlines methodology for the integration of sector statistical systems into the national statistical system through the development of a master sample frame and its use in an integrated survey framework. The Global Strategy calls for the linkage of economic and social indicators with the land use and environment. Technology will play a key role.

The current national and global statistical system is largely based on methodology that is decades old and does not consider the huge development of technology and how that can be used to more efficiently improve the quality of data. It is important that with the decreasing resources for statistical work and increasing demands more thought is given to increase efficiency of statistical operations (Belkindas, 2011)

The new initiative to improve statistical systems worldwide the Busan action plan on statistics (BAPS) is concentrating on six actions (Statistics .. 2011):

*Action 1* - Strengthen and re-focus national and regional statistical strategies with a particular emphasis on improving statistical systems that support country-level development priorities. Strategies should be updated to reflect new challenges and opportunities. A key aim is to increase in-country capacity – including through training - to respond to emerging and unforeseen needs.

*Action 2* - Ensure that outcomes of global summits and high-level forums specifically recognize the need for statistical capacity development, including technical assistance, training, and financial support. Implementing and monitoring global initiatives requires collaboration between national and international statistical organizations.

*Action 3* - Implement standards for data preservation, documentation, and dissemination that permit full public access to official statistics.

*Action 4* - Develop programs to increase the knowledge and skills needed to use statistics effectively for planning, analysis, monitoring, and evaluation, thus increasing transparency and accountability.

*Action 5* - Improve accessibility of statistics and research produced by international organizations and bilateral development agencies.

*Action 6* - Ensure financing for statistical information is robust and that funding instruments and approaches reflect the new modalities and actors in development finance.

The plan calls for improvement of statistical systems in developed and developing world, to refocus them to cater to the most important needs of the Governments and society and to increase financing for statistics. Such systems should be able to provide data comparable across the Globe.

## REFERENCES

- VOGEL, FA, 2009, "Lessons Learned from ICP 2005" Report to the ICP Technical Advisory Group, the World Bank, October 2009. ([www.worldbank.org/data/icp](http://www.worldbank.org/data/icp)).
- ILO/IMF/OECD/Eurostat/the World Bank, (2004), Consumer Price Index Manual, Theory and Practice, International Labor Office, Geneva.
- BIGGERI, LUIGI and LAURETI, IIZIANA (2011), "Extrapolation of PPPs over time using CPI data: Methods and Interpretation". Proceedings of the 58<sup>th</sup> World Congress of the International Statistical Institute, Dublin, Ireland, August 2011.
- PANDA, S.S; AMES, D.P. and PANIGRAHI, S. (2010) "Application of Vegetation Indices for Agricultural Crop Yield Prediction using Neural Network Techniques, Remote Sensing, 2010.
- "The Global Strategy to Improve Agricultural and Rural Statistics" Report no. 56719-BLB, The World Bank, September 2010.

MISHA V. BELKINDAS (2011) “Rising Demands in Era of Shrinking Budgets  
Proceedings of the 58<sup>th</sup> World Congress of the International Statistical  
Institute, Dublin, Ireland, August 2011.

“Statistics for Transparency, Accountability and Results”, Draft, September 2011.

*STATISTICS IN TRANSITION*-new series, December 2011  
Vol. 12, No. 3, pp. 453—460

## **STRENGTHENING REGIONAL CROSS-BORDER INFORMATION SYSTEMS: THE EXAMPLE OF GENEVA CROSS-BORDER REGION**

**Gabriel Gamez<sup>1</sup>**

### **ABSTRACT**

This paper in the first place briefly introduces the emerging issue of cross-border regions throughout Western and Eastern Europe and discusses the definition of cross-border cooperation. Secondly, it analyzes the development of joint institutional mechanisms for the design and monitoring of cross-border regional policies in all relevant fields including the cooperation in the field of statistics. Thirdly, the paper will describe in more detail the institutional and practical organization of the Statistical Observatory of the Geneva cross-border region. The Observatory was formally established in 2001 by the France-Geneva Regional Committee, the governance committee of the Geneva cross-border region, and produces more than 50 socio-economic and environmental indicators for a cross-border urban area of more than 800,000 inhabitants over two countries: 1/3 in France and 2/3 in Switzerland with close to 75,000 cross-border workers. The paper presents some important indicators produced by the observatory and concludes with major achievements and remaining challenges.

### **1. Cross-Border Regions (CBRs): development and definitions**

Until recently, in Europe, state borders were considered as hermetic and securitized boundaries subject to strict immigration (and sometimes emigration) controls and restrictive customs regulations. It is only after the Second World War in the case of West European countries and in the 90's for East European countries, that the increasing mobility of goods, people, knowledge, capital and information changed the characteristics and functions of borders. Enterprises and citizens from contiguous regions from two or more nation states established gradually an economic model based on the comparative advantages of each of the states. The European integration process intensified the implementation of this new economic pattern where for example firms can set up their activities depending on conditions in terms of the national regulations, access to

---

<sup>1</sup> United Nations, Economic Commission for Europe, Statistical Division, Regional Adviser.

international markets and tax burden on one side of the border while at the same time benefitting from lower property prices and labour costs on the other side of the border. In well-integrated cross-border regions, studies confirm that the strong asymmetry in terms of salaries as well as goods, services and property prices lead an increasing number of workers to keep their job in a central urban area while moving their residence to the periphery, on the other side of the border. This phenomenon is reinforced further in transnational regions with a strong shared cultural identity, such as Geneva, since access to education and services (e.g. health care) on the other side of the border is not hindered by any language barrier.

**Figure 1.** Geneva metropolitan cross-border region

	<b>Total</b>	<b>CH</b>	<b>FR</b>
Area, in km <sup>2</sup>	2,097 <i>100%</i>	553 <i>26%</i>	1544 <i>74%</i>
Municipalities	232 <i>100%</i>	92 <i>39%</i>	140 <i>61%</i>
Population	812,346 <i>100%</i>	528,299 <i>65%</i>	284,047 <i>35%</i>
Employment	412,444 <i>100%</i>	333,490 <i>81%</i>	78,954 <i>19%</i>
Housing	394,401 <i>100%</i>	251,779 <i>64%</i>	142,622 <i>36%</i>

## 2. Cross-Border Cooperation

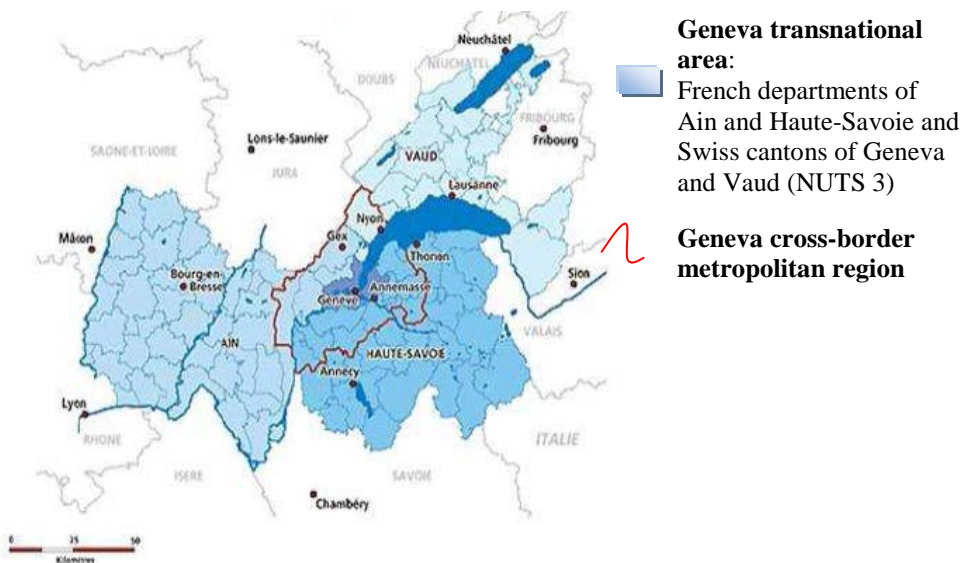
CBRs not only developed their own cultural, social and economic identities but were also gradually confronted with practical issues such as land use, transport network and the environment. The CBRs were increasingly concerned with practical problem-solving in a wide range of domains that needed a coordinated administrative response and developed, in some cases, original forms of public governance between contiguous territorial authorities across national borders.

Thus, cross-border cooperation can be defined as: cross-border collaboration mechanisms and agreements established between contiguous sub-national territorial authorities (municipalities and districts) within their own national legal systems.

Cross-border cooperation is often based on informal arrangements among participant authorities that, in some cases, can be acknowledged or even empowered by the relevant national authorities. In the European Union, the INTERREGIO programme helped to institutionalize these agreements but one should not forget that cross-border cooperation does not create some kind of supranational authorities. Furthermore, transnational metropolitan areas do not

always cover entirely two or more sub-national administrative units (regions or districts) on each side of the border. This makes the establishment of cooperation mechanisms even more difficult and complicates the production of statistical indicators since the latter are often based on sub-national administrative unit at a rather aggregated level; nomenclature of territorial units for statistics (NUTS) in the EU. It is also important to distinguish cross-border cooperation from cooperation programmes that aim to promote cooperation among larger regions or even among two or more countries in Europe.

**Figure 2.** The Geneva cross-border region and related sub-national administrative territories



While many indicators and statistical information are available on the mobility of cross-border workers within transnational regions in Europe, (and in some cases on shuttle trade), private and public stakeholders are confronted with scarce regional cross-border statistical information in other areas such as health, education, security, housing and real estate market, trade, and environment. The main challenge for the implementation of cross-border agreements and cooperation mechanisms in these areas is to go beyond the traditional institutional organization of official statistics based on countries and national administrative regions and instead to propose statistical information and indicators according to new territorial divisions in all relevant fields for the monitoring of the development of cross-border regions.

The rest of this paper will examine the practical case of the Geneva metropolitan cross-border region which established in the 70's, cooperation



mechanisms and institutions that deal with all aspects of the daily-life and future developments of the region based on indicators produced by a regional and independent statistical body.

### **3. Geneva cross-border cooperation and its Statistical Observatory**

#### **3.1. Institutional setup and organization**

The Geneva metropolitan region institutionalized its cooperation already in 1973 by setting up a cross-border coordination body: the France-Geneva Regional Committee (Comité Régional Franco-Genevois / CRFG). The CRFG was created by an agreement between the French and the Swiss governments and foresaw the participation in the Committee of representatives from local authorities from the departments of Ain and Haute-Savoie, and the region of Rhone-Alpes on the French side, and on the Swiss side, from the Canton of Geneva.

The aim of the CRFG is to define common strategies for the development of the cross-border region and to provide the necessary impetus for the realization of joint projects. The Committee meets in general once a year and defines main priorities and strategic programmes. For the implementation of the programmes, five sub-committees have been created that deal respectively with issues concerning travel and Security, economy and employment, housing and environment, health and social cohesion, and culture, education and sport. The functions of the CRFG and its five subordinated bodies were gradually reinforced by the entry into force of Swiss-EU bilateral agreements on e.g.: free movement of persons, overland transport, Schengen/Dublin, education and statistics.

In order to monitor the effects of the bilateral agreements between Switzerland and the EU on the regional integration process, the CRFG decided in 2001 to set-up a regional cross-border statistical observatory (Observatoire Statistique Transfrontalier / OST) jointly financed and managed by the statistical office of the Canton of Geneva (OCSTAT) and the INSEE Rhone-Alpes regional statistical office. Data from the Geneva metropolitan cross-border region situated on the district of Nyon are provided by the statistical office of the Canton Vaud, even though the latter is not formally part of the agreement. In 2007, Swiss and French partners in the CRFG signed a framework agreement that clarifies the mission, activities, organization and funding of the Statistical Observatory. The agreement stressed that the Observatory complies with the principles of official statistics. It benefits from professional independence, ensures impartiality in the compilation and publication of statistical information, follows the principle of confidentiality, and provides to all users simultaneous equal access to the information. Statistical information compiled by the Observatory is available on its website free of charge. The CRFG approves the annual statistical programme of work and guides the Observatory in its strategic development and research activities.

Until 2007, the activities of the Observatory were co-financed by the INTERREG III programme but since then the French and Swiss public authorities involved in the framework agreement subsumed responsibility for the funding of the Observatory. The annual budget of the cross-border statistical body is approximately €200 000. Staff costs, equivalent to two full-time positions, stand for 80% of the total budget. The remainder covers additional costs, for example, printing, shipping, service delivery to third parties and maintenance of the website of the Observatory (<http://www.statregio-francosuisse.net/>). INSEE Rhone-Alpes and OCSTAT provide additional contributions in kind by making available IT infrastructure, technical and scientific management and data access.

### **3.2. Territorial scope and domains of observation**

First of all, it is important to stress that the Observatory is in general not conducting statistical surveys. Indicators produced by the cross-border statistical body are based on statistical information obtained from the relevant national and regional statistical offices in France and Switzerland, often at NUTS 3 level, and also administrative data transmitted by public authorities within the Geneva metropolitan cross-border region.

This is one reason which explains why the statistical information compiled by the Observatory covers two different territorial divisions (see also figure 2):

- (1) The transnational area, covering the departments of Ain and Haute-Savoie on the French side, and the cantons of Geneva and Vaud on the Swiss side (NUTS 3);
- (2) Geneva metropolitan cross-border region.

The Geneva metropolitan cross-border region accounts for about one third of the population, employment and housing of the transnational area. The metropolitan cross-border cooperation is evolving and covers gradually a larger share of the transnational area. Thus, the region covered by the cross-border cooperation does not correspond exactly to the boundary of the area covered by the Observatory. The latter outlined the Geneva metropolitan cross-border region based on employment criteria and cross-border flows of workers. For statistical reasons (comparability and coherence) it is not possible for the Observatory to change the boundary of the metropolitan region. Nevertheless, the data produced by the Observatory remain relevant for the analysis of major economic, social and environmental phenomena in the metropolitan area.

In order to provide statistical information for the cross-border cooperation process and the monitoring of joint metropolitan projects, the Observatory, after consultation with the CFRG, decided to compile more than 50 statistical indicators; some available for the transnational region and the metropolitan area, some others only for one of these two territorial divisions.

The indicators cover the following domains:

- |                           |  |
|---------------------------|--|
| 1. Population             | 8. Real Estate Market                  |
| 2. Labour market          | 9. Land                                |
| 3. Social balance         | 10. Transport and mobility             |
| 4. Education and training | 11. Economic framework and development |
| 5. Health                 | 12. International Trade                |
| 6. Housing                | 13. Public procurements                |
| 7. Environment            |  |

The list of indicators can be found on the website of the Statistical Observatory.

### **3.3. Dissemination**

All statistical indicators, charts, maps and publications produced by the Statistical Observatory are made available free of charge on its website. As requested by the framework agreement, the Observatory produces every year in September a statistical yearbook. The yearbook is the flagship publication of the Observatory where statistical indicators for all relevant domains are presented and complemented with short analyses, charts and maps. Besides the regular chapters, the yearbook describes and comments every year on a specific aspect of the development of the cross-border region. The yearbook covers the Geneva metropolitan cross-border region as well as the transnational area.

Thematic analyses are also carried out by the Observatory and published in Fact Sheets. Two to four fact sheets are produced every year. The domains covered by these thematic analyses are approved by the CRFG as part of the annual working programme. The Observatory also issues, on a regular basis, press-releases that inform the media and the public at large about new publications and fact sheets. On the website, detailed spread-sheets can be downloaded in Excel format. A list of all the publications and dissemination material to be issued during the year, as detailed in the annual statistical programme of work, are available on the website. It is nevertheless still not to be considered as a pre-announced release calendar since dates and time of dissemination are not mentioned.

### **3.4. Primary data sources**

The major issue for the Observatory in the compilation of cross-border regional statistics is to some extent the comparability of the primary data sources. If the French and Swiss national statistical institutes are both following international guidelines and recommendations, a major achievement in the comparability of national statistics was the entry into force of the Swiss-EU bilateral agreement in Statistics. Through this agreement, Switzerland formalized its collaboration with EU authorities in the field of statistics (Eurostat) with the aim to achieve the compliance of Swiss statistics with the EU standards in all fields relevant for the cooperation between both partners. The comparability of

official statistics between France and Switzerland has been gradually enhanced including data at NUTS 3 level for the compilation of indicators for the transnational area: French departments of Ain and Haute-Savoie, and Swiss cantons of Geneva and Vaud. For the Geneva metropolitan cross-border region, however, most of the indicators rely on administrative data. The latter are by definition only seldom comparable which complicates the production of reliable statistical indicator for the metropolitan region.

#### 4. Conclusions

The availability of reliable statistical indicators is a prerequisite for the development of cross-border regions. It is important that cooperation mechanisms and institutions for the development and follow-up of joint programmes and projects can rely on diagnostic tools that provide a common reference framework and a comprehensive picture of the cross-border region in all relevant domains.

Institutionalized cross-border cooperation with joint public governance mechanisms can reinforce the legitimacy of an independent cross border statistical body. A joint Cross-Regional Committee, representing the interests of the region and its major stakeholders, should not only be responsible for the establishment of a subordinated statistical body but should also be involved in the adoption of the multi-year and the annual programme of work of this institution. This secures the production of relevant regional indicators for the development and monitoring of joint programmes and projects in the cross-border region. Sub-national (territorial) statistical offices of the regions involved in the cross-border cooperation should make available and combine their respective know-how, competences and information technologies in the production of cross-border statistics. Last but not least, the sustainable development of a cross-border regional office requires budgetary contributions from all public partners involved in the project.

Nevertheless, while goodwill and political commitment from cross-border regional authorities are an important factor for the set-up of a statistical body, there are still some outstanding issues that need to be addressed: for example:

- Cross-border regions, and in particular metropolitan areas, do not match exactly sub-national administrative regions on each side of the border and the size of territorial administrative units (districts or municipalities) can diverge considerably.
- Cross-border statistical information and indicators often rely on regional statistics produced by the national statistical systems on both sides of the transnational area. It is therefore important for the comparability of the information used in the compilation of cross-border statistics that countries comply with international and European recommendations and guidelines. The capacity to produce cross-border statistics can be seen as an indicator of the degree of international harmonization of national statistics.

- Some indicators cannot rely on national or regional statistical information and need to be compiled out of administrative data. However, administrative data are often not comparable across borders. In this case, comprehensive national population and business statistical registers, regularly updated according to international recommendations, could be an excellent source of information for the compilation of cross-border statistics.
- Cross-border statistical bodies are confronted with different legal provisions concerning the transmission of or access given to third parties of confidential data. In some cases, the national legal framework does not allow, national or regional statistical institutes to transfer or give access to confidential statistical information to a cross-border statistical institution. Unfortunately, national and European legislations do not consider cross-border statistical bodies as part of the system of official statistics.

It is obviously still a challenge for cross-border regions in Europe to extend beyond the traditional institutional organization of official statistics which is based on national and sub-national divisions. However, compared to other fields with cross-border cooperation mechanisms and tools, statistics can build on a strong common professional culture and a longstanding successful international cooperation. This explains why statistics is among the most successful fields of cooperation in the Geneva metropolitan cross-border region.

## Acknowledgements

This contribution greatly benefited from the dissemination material available on the website of the *Observatoire Statistique Transfrontalier* (OST / [www.statregio-francosuisse.net/](http://www.statregio-francosuisse.net/)) and extensive discussions with Mr. Dominique Frei, Director of the Statistical Office of the Canton of Geneva (OCSTAT / [www.ge.ch/statistique](http://www.ge.ch/statistique)).

## REFERENCES

- MARKUS PERKMANN, *Cross-border regions in Europe*, in *European urban regional studies* 2003 (10:153): <http://eur.sagepub.com/content/10/2/153>.
- JEN NELLES and OLIVIER WALTHER, *Changing European borders: from separation to interface? An introduction*, in *Articulo - Journal of Urban research* 6/2011: <http://articulo.revues.org/1658>.
- URSZULA ROMAN, *Cross-border region in Europe: An impediment or opportunity? A case of German-Polish and Slovakian-Polish border*, UACES Student Forum 11 Annual Conference University of Bath, 29-30 April 2010: <http://www.uws.ac.uk/schoolsdepts/business/cces/documents/BathConferencePaper.pdf>.

STATISTICS IN TRANSITION-new series, December 2011  
Vol. 12, No. 3, pp. 461—468

## THE INTEGRATION OF STATISTICS OF KAZAKHSTAN WITH THE EUROPEAN STATISTICAL SYSTEM

Alikhan Smailov<sup>1</sup>

### ABSTRACT

1. Legal basis of integration of the Agency of Statistics of the Republic of Kazakhstan (further – ASRK) with international statistical community:
  - law «About the state statistics», principles, positions, international agreements;
  - strategic plans for development.
2. General information on interaction of ASRK with international statistical community:
  - global estimations;
  - agreement with World Bank;
  - cooperation with international organizations and national statistical offices.
3. Activity of ARKS on harmonization of standards, qualifiers and methodology with the international analogs:
  - Harmonization of statistical qualifiers.
  - Introduction of the international standards in statistics branches:
    - national accounts;
    - price statistics;
    - manufacture and others (according to information of departments).
4. Introduction of the modern technologies in processes of gathering, processing and distribution the data on the basis of international experience.

Let me present you the report about the **integration of statistics of Kazakhstan with the European statistical system.**

It seems appropriate to begin with information about the **legal basis** for the integration of the Agency of Statistics of the Republic of Kazakhstan with international statistical community.

On March 19, 2010 a new **Law of the Republic of Kazakhstan "On State Statistics"**, corresponding to modern realities was adopted. This Law regulates public relations arising in the process of state statistical activity and aims to meet the needs of society, state and international community in official statistical information.

---

<sup>1</sup> The Agency of Statistics of the Republic of Kazakhstan.

Article 9 of the Law defines a general procedure for international cooperation in the field of statistics. In particular, it establishes the right of state statistical bodies on dissemination of statistical information and exchange of experience in statistical methodology in the framework of cooperation with foreign partners.

Article 5 sets out the principles of state statistics. One of the main principles is to ensure consistency and comparability of state statistics with internationally accepted standards, classifications and methods.

Thus, we can say that in the basic document of the statistical system of the Republic there is a necessary mechanism for its harmonization with European and international standards.

International agreements have an important role in establishing and expanding partnership relations. Thus, the Partnership and Cooperation Agreement signed in 1995 is the basic document upon which the partnership of European Union and Kazakhstan is based. Article 67 of this agreement is devoted to cooperation in the field of statistics.

In addition, in 2010 a Memorandum of understanding between the Agency of Statistics of the Republic of Kazakhstan and Eurostat was signed. The Agency also has cooperation agreements with national statistical offices of many European countries.

For effective and progressive implementation of state policy in the field of statistics the Agency prepares the **strategic plans**. Strategic plan of the Agency of Statistics of the Republic of Kazakhstan for 2010-2014 identifies *obtainment of international experience in the field of statistics* as a priority area of activity. Through the achievement of target indicators in this area it is planned to increase the degree of integration of Kazakhstan statistics into international statistical space.

Speaking about long-term development plans, the assistance provided by international institutions in the analysis of statistical system of the country should be noted. In particular, in 2002-2003 Eurostat and UNECE and in 2007-2008 UNECE and ESCAP conducted a **global assessment** of the statistical system of the Republic of Kazakhstan. The report of the assessment of 2007-2008 marked a good level of cooperation of the Agency with national statistical offices and international organizations aimed at improving the quality of statistical information and introduction of international standards.

**Interaction of the Agency of Statistics of the Republic of Kazakhstan with international statistical community** is implemented through continued **cooperation with international organizations and national statistical offices** in such areas as introduction of international statistical methodologies, standards and classifications, development of modern tools, information and communication technologies, improvement of knowledge and skills of personnel and their professional competencies.

One of the brightest examples of such cooperation is the European Union TACIS program. In 1993-2007, during implementation of the TACIS program in Kazakhstan dozens of projects in areas such as statistical infrastructure, economic

and social statistics were carried out. In the course of the program, European experts have made a significant contribution to the harmonization of the statistical system of Kazakhstan with European counterparts.

It is necessary to mention also the efforts of organizations of the United Nations system and Interstate Statistical Committee of the CIS on assistance in introduction of the international standards in Kazakhstan, and in the whole region. Thus, during the period since 1992 to 2010 569 recommendations, instructions, classifications and other methodological materials were developed by the CIS Statistical Committee in line with international standards.

At the present time the Agency of the Republic of Kazakhstan on Statistics is carrying out a preparatory work for the implementation of **a wide scale project with the World Bank on strengthening the national statistical system of the Republic of Kazakhstan**. It is planned to start the implementation of this five year project since the beginning of 2012. In the nearest future it is planned to announce the international competitive bidding on twinning-partnership with the Agency for rendering the consulting services.

Dear colleagues, further, let me turn to review of Kazakhstan experience on **harmonization of national standards, classifications and methodologies with international analogs**.

**Classifications** are a basis on which all information systems are constructed. They are considered as basic tools used for maintaining of many various state databases.

In order to ensure statistical data comparability and harmonization with the international classifications, since 1997 the Agency of the Republic of Kazakhstan on Statistics has been carrying out a work on developing and maintaining statistical classifications in line with international recommendations. The state classifications of economic activities (based on NACE) and of products by activity (based on CPA) were the first classifications developed according to the international standards. It should be noted that NACE classification of the European Community is more preferable to Kazakhstan than ISIC classification of the United Nations, because it provides more detailed distribution of the enterprises in mining and manufacturing industries. Also, in the European classification system more convenient for users coding system is used.

In 2006-2008, in connection with the revision of NACE and CPA classifications in the European Community, the transition to new system of international statistical classifications has begun in Kazakhstan. At present, Kazakhstan is one of the first CIS countries which have passed to new system of international classifications.

In 2007-2008, new versions of classifications of economic activities and products by activity were adopted as the state standards.

To the present time the Agency has revised sectoral statistical classifications of goods and services used in industry, agriculture, forestry, fishery, trade, construction, tourism, and service statistics.



Introduction of international classifications into statistical practice is one of the priority directions of development of statistics in Kazakhstan.

Also the Agency widely uses the experience of the European countries on **statistical registers**.

As an example, structure and content of our Statistical business–register were defined according to the standards and recommendations of the European Union in the field of enterprises statistics.

In its turn, structure and content of being created statistical population register are defined taking into account the experience of the North European countries. In particular, recommendations named «Statistics on basis of registers in the European countries. Review of advanced techniques with paying major attention to population and social statistics 2008» have been used.

We permanently work on **introduction of international standards into all areas of statistical activity**.

Great attention is paid to the improvement of **national accounts**. In the frameworks of TACIS program the projects on non–observed economy and quarterly calculations of gross regional product have been implemented. In cooperation with Statistics Norway the process of formation of “supply-use” tables is being improved. And it is not casually that in the Strategic plan of the Agency the necessity of accelerated development of national accounts system is especially mentioned. In particular, special attention will be paid to introduction of new international standard “SNA 2008”, to improvement of the comparability between separate calculations within SNA and to transition to more detailed calculations of intersectoral balance.

**Price statistics** also has a high importance. In many European countries along with a consumer price index also a harmonized consumer price index is calculated, which is formed according to the coordinated approach and on the basis of a uniform set of definitions. According to the Eurostat Guidelines on harmonized CPI, basic requirements to construction of this indicator are the inclusion of such services, as public health, education, insurance, financial and social protection services, as well as housing expenditures in the CPI structure. Since January, 2011 CPI of Kazakhstan includes the all list of the named services.

«Guidelines on producer’s price index: theory and practice» developed by international organizations are used by the Agency as a methodological basis for the construction of a PPI. According to this Guidelines and the «Methodological manual of Eurostat on constructing the producer’s price indexes on services 2005», along with estimation of prices movement in the production, the price changes of the services of productive character should be also considered. At present, in Kazakhstan the services of productive character are partially considered in the CPI structure, and also independent price indexes on freight transport and communication services are calculated. Since 2011, index of commercial real estate rent price is being constructed.

In the future, construction of a number of price indexes on services in the field of accounting, audit, consulting, computer facilities, scientific research and developments and others in line with the Eurostat methodology is planned.

According to the studied European experience the Agency has formed a system of indicators of fixed capital **investment statistics**, improved **construction statistics**, introduced a classification of construction products on the basis of CPA 2008.

**Energy statistics** is also formed in line with the international standards. In 2002, according to the Eurostat recommendations the «Methodological provisions on definition and formation of indicators of energy statistics» were developed and adopted, in 2008 the indicators of output (consumption) of electricity by kinds of activity were revised and introduced, in 2009 in line with the Guidelines on energy statistics of Eurostat and International energy agency 2007, the Methodical explanations of indicators of energy statistics were revised.

At present, the Agency implements a Plan of actions on development of **ecology statistics** for 2010-2012; and with a view of its implementation, on the basis of UNECE recommendations the «Ecological indicators and reports based on them for EECCA countries (the Eastern Europe, Caucasus and the Central Asia)» the «Methodical recommendations on formation of indicators of environment statistics» were prepared. In 2011, in the framework of implementation of this plan, the Classification of kinds of activity and expenditures on protection of the environment were developed based on the CEPA 2000 international standard (Classification of Environmental Protection Activities and Expenditure).

In **domestic trade** and **commercial services** statistics area, indicators of volumes of retail trade, services by kind of activity, physical volume indexes are developed in line with the Eurostat methodical manual on short-term enterprises statistics 2006. In domestic trade statistics, the introduction into the national practice of the United Nations domestic trade methodical recommendations 2008 is also among planned actions.

According to the OECD Guidelines on measuring non-observed economy, in order to ensure completeness and quality of the service statistics data, the estimation of volume of **non-observed economy** is carried out.

In **foreign merchandise trade statistics**, the Agency since July, 2011 carries out the surveys on mutual trade statistics of participants of external economic activities of the customs union of three countries – Republic of Kazakhstan, Republic of Belarus and the Russian Federation. Thus, in practice the experience of the EU countries on the formation of foreign trade statistics in line with the international merchandise trade statistics methodology is used, and also, introduction of the UN methodology of 2008 is planned.

In **trade markets statistics**, taking into account the international practice and recommendations of the International energy agency, a fuel–energy balance of the country and energy intensity of GDP are calculated annually. In connection with the ratification by Kazakhstan of the Kyoto Protocol to the United Nations

Framework Convention on Climate Change, new methodical recommendations on fuel–energy balance were developed, including the provisions on calculation of energy and carbon intensity of GDP.

Agency also constantly studies the international experience on **tourism statistics**. On the basis of international recommendations on tourism statistics 2008 the Agency has developed the methodical recommendations on formation of tourism statistics indicators. Sample surveys of visitors (tourists) and households for formation of the Satellite tourism account are carried out in line with international experience.

In **information-communication technologies statistics**, on the basis of the «Guidelines on estimating access to ICT and their use by households and separate persons» and «Guidelines on information economy statistics» of UNCTAD (United Nations Conference on Trade and Development) and International telecommunication union the sample surveys about level of use of information-communication technologies in enterprises and households are carried out. In 2012 it is planned to conduct a new survey in the area of secondary education prepared on recommendations of the International telecommunication union and Partnership on ICT estimation for development. On the basis of this survey the indicators characterizing availability of means of interactive training, share of students using computers and the Internet will be formed.

In formation of indicators of **science and innovation statistics** the methodological approaches based on Frascati and Oslo Manuals are used. On annual basis the indicators about level of innovative activity, expenditures on technological innovations, on scientific research and developments, administrative innovations and others are formed. In 2009 the methodical recommendations on innovation statistics which allow forming internationally comparable indicators of innovative activity have been developed taking into account the international standards.

In **statistics of culture**, according to the recommendations of the UNESCO Institute of Statistics 2009 new indicators on objects of the film industry, such as number of created national films, national films with participation of other countries, with prevalence of national or foreign financing and others are formed. Further introduction of UNESCO methodology on the system of culture statistics 2009 in Kazakhstan is planned.

In **demographic statistics** calculations of all indicators are based on methodological recommendations of the UNECE and Eurostat. Currently, we develop a "Methodology of predicting the number and sex composition of population" using the UN guidelines («United Nations Manual III: methods for population projections by sex and age»).

Thus, the 1999 and 2009 population censuses were conducted using UNECE guidelines. The "Recommendations of the Conference of European Statisticians on Population and Housing Census Round 2010" prepared in cooperation with Eurostat were used in developing the toolkit of 2009 Census.

At present, the formation of indicators of **employment, wages and living standards statistics** in Kazakhstan is carried out in line with the recommendations and standards of the International Labour Organization, the UNECE and the World Bank.

Key labor market indicators are collected through the quarterly Labour Force Survey (LFS), which was organized with the support of Eurostat and statistical office of Germany.

In 2010 the Agency studied and applied experience of the national statistical office of Poland in terms of constructing and calculating the index of labor cost.

Ladies and gentlemen, in conclusion, let me comment briefly on the **introduction of modern technologies by the Agency into the processes for collection, processing and dissemination of data on the basis of international experience.**

As part of the government program for the development of "e-government (electronic)" and for improvement of the information infrastructure, in 2010 the project of Integrated information system "e-Statistics" (hereinafter – IIS "e-Statistics") of the Republic of Kazakhstan has been launched.

IIS "e-Statistics" is designed for automated collection and processing of primary statistical data, storage of statistical data, compilation of statistical materials and implementation of statistical data inquiries.

One of the main components of IIS "e-Statistics" is the information system "Metadata" (hereinafter – IS "Metadata"). The main purpose of IS "Metadata" is managing the Agency's information systems, as well as satisfaction of the needs of different groups of statistical information users.

Close internal integration of the IS "Metadata" and IS "CLASS" is provided. IS "CLASS" is designed for automatization of the processes of design, implementation, maintenance and dissemination of statistical classifications, nomenclatures and reference books.

Data from the sources of statistical information are collected in the IIS "e-Statistics" by means of the database "Administrative data" and information system "Storage of primary statistical data."

In IIS "e-Statistics" data collection in the production of statistical information is done by information system "On-line data collection".

To ensure timely and high quality processing of primary statistical information and compilation of statistical indicators in IIS "e-Statistics" the information system "Processing of primary statistical and administrative information" is provided.

In IIS "e-Statistics" the process of dissemination of statistical information is supported by the database of aggregated indicators and information-analytical system.

Last year the work on creation of information-analytical system "Taldau" (hereinafter – IAS "Taldau") has begun. In the current year IAS "Taldau" was put into trial operation and access is granted to all users without restriction. The implementation of this system is aimed on providing access to a wide range of statistical data users at a convenient for them format.

To date, more than 600 key statistical indicators are uploaded into this system. The system includes three main modules – "Section", "Regions" and "Catalog".

"*Catalogue*" is a hierarchical structure of all statistical indicators that are uploaded into the system and formed on the basis of classifier of statistical indicators.

"*Section*" is a list of statistical indicators, grouped by sector.

On page of the "*Regions*" module the user can view the map and brief details on the selected region (area, administrative center, population, etc.), as well as to analyze the development of regions through statistical indicators.

This year as part of IAS "Taldau" the module "Analytics" is planned. It will contain several models of analysis, presumably:

- correlation analysis;
- comparison of indicators with international statistics;
- producing predictive data;
- cross analysis.

Another planned module is "My Account". Personal account of the user will allow system users to create ad hoc reports and store them on their page.

Within the "e-Statistics" project the "On-line data collection" information system is developed using the modern information technology, which allows respondents to submit the reports in real time without leaving the office. This system allows automating the process of completing and submitting the required statistical reports by respondents using the digital signature. The following year, respondents will be given an opportunity to submit reports through this system for 65 statistical forms. By 2013 it is planned to introduce all statistical forms into this system.

Dear colleagues, thank you for your attention. I hope that this report has been useful to you.

STATISTICS IN TRANSITION-new series, December 2011  
Vol. 12, No. 3, pp. 469—474

## QUALITY IN OFFICIAL STATISTICS USING EUROPEAN STATISTICS CODE OF PRACTISE AS A FRAMEWORK

Berit Olsson<sup>1</sup>

### ABSTRACT

To build trust for the statistics in the society and for the national statistical institute (NSI) high quality of official statistics is a key issue. High quality of official statistics is therefore important.

But how does the society know that the quality is good? And how can the NSI prove that the disseminated statistics have good quality?

One way is to follow international guidelines and framework for quality measures and to publish quality indicators. One useful quality framework is **European statistics Code of Practice, CoP**. The European Code of Practice is based on 15 principles. Governance authorities and statistical authorities in the European Union commit themselves to adhere to the principles fixed in this code covering the institutional environment, statistical processes and outputs. A set of indicators of good practice for each of the 15 principles provides a reference for reviewing the implementation of the Code.

It takes time to implement the CoP and to measure all indicators but the CoP may work as guide and some indicators can be chosen as a start.

#### **Institutional environment**

Institutional and organizational factors have a significant influence on the effectiveness and credibility of a statistical authority. Example of principles are professional independence, mandate for collection of data quality, commitment and statistical confidentiality.

#### **Statistical processes**

European and other international standards, guidelines and good practices must be fully observed in the processes used by statistical authorities to organize, collect, process and disseminate official statistics. Examples of principles are sound methodology, appropriate statistical procedures and non-excessive burden on respondents.

---

<sup>1</sup> Statistiska Centralbyran, Statistics Sweden, International Consulting Office.

### **Statistical output**

Available statistics must meet users' needs. The important issues concern the extent to which the statistics are relevant, accurate and reliable, timely, coherent, comparable across regions and countries, and readily accessible by users.

The documentation of the principles is of great importance for the users, especially the principles and indicators of the statistical output. The use of some kind of Metadata system is recommended.

Statistics Sweden has been working with different aspects of quality over the time and CoP is today an important basement. During the last five years much attention has been given to the analysis of the statistical processes.

For the moment the Quality Management system at Statistics Sweden include four key issues:

- Quality framework: **The EFQM Excellence Model**
- Methods for continuous improvement work: **Six Sigma**
- ISO 20252 Market, opinion and social research
- Methods for evaluation and follow-up: **Internal auditing.**

To build trust for the statistics in the society and for the national statistical institute (NSI), high quality of official statistics is a key issue. But how does the society know that the quality is good? And how can the NSI prove that the disseminated statistics have good quality? And how will you define quality?

When talking about the quality of statistics the focus has often been on sampling errors or other indicators of statistical output. But during recent decades more focus has been put on the whole context of the statistical production. Quality work encompasses, to a larger or smaller extent, all activities at a statistical office. Different models have been developed as Total Quality Management, Balance Scorecard, EFQM, Six Sigma and EU's Code of Practice. There is an agreement that quality is dependent on many different factors or components and you have to take into account the quality of the organisation, the quality of the processes used and of course the quality of the products. In order to improve quality it is important to consider the interdependence between quality actions. As an example, a higher organisational quality is important to improve product quality, while at the same time the actions taken to improve product quality will help also to improve organisational quality.

So when you want to define statistical quality you have to take into account institutional and organisation quality, process quality and statistical output quality. It is thus essential to work with all three areas simultaneously. In order to get an overview or a good structure both for users and producers you may follow international guidelines and framework for quality measures and publish quality indicators. The framework has to be well known, accepted and used by several agencies.

One useful quality framework is **European statistics Code of Practice, CoP**. The European Code of Practice is based on 15 principles. Governance authorities and statistical authorities in the European Union commit themselves to adhere to the principles fixed in this code covering the institutional environment, statistical processes and outputs. A set of indicators of good practice for each of the 15 principles provides a reference for reviewing the implementation of the Code.

It should be stressed that the transit to an organisation that works according to the principles of CoP or some other quality framework is a long process. It takes time to implement the CoP and to measure all indicators but the CoP may work as a guide and some indicators can be chosen as a start. The ultimate goal for a quality improvement strategy is that quality work will be an integral part of the ordinary work.

### **Institutional environment (CoP Principle 1 – 6)**

Institutional and organisational quality refers to management, strategies, competence and processes of the organisation. Institutional and organisational factors have a significant influence on the effectiveness and credibility of a statistical authority. Examples of principles are professional independence, mandate for collection of data quality, commitment and statistical confidentiality.

#### *Principle 1: Professional independence*

“The professional independence of statistical authorities from other policy, regulatory or administrative departments and bodies, as well as from private sector operators, ensures the credibility of European statistics.”

The statistical law plays an important role. Independence of the statistical agency has to be specified in law. Transparency of the statistical agency is also of importance with statistical work programs published as well as a follow-up to progress made. Statistical releases have to be clearly distinguished and issued separately from political/policy statement.

#### *Principle 2: Mandate for data collection*

Indicators; examples;

- The mandate to collect information for the production and dissemination of official statistics is specified in law
- The statistical authority is allowed by national legislation to use administrative records for statistical purposes.

#### *Principle 3: Adequacy of resources*

The resources available to statistical authorities – staff, financing, computers - must be sufficient to meet (European) statistics requirement.



*Principle 4: Quality commitment*

“All ESS members commit themselves to work and cooperate according to the principles fixed in the ‘Quality declaration of the European statistical system’”.

*Principle 5: Statistical confidentiality*

It is very important that the statistical confidentiality is guaranteed by law.

*Principle 6: Impartiality and objectivity*

This principle includes some very important indicators:

- Statistics are compiled on an objective basis determined by statistical considerations
- Errors discovered in published statistics are corrected at the earliest possible date and published
- Statistical release dates and times are pre-announced
- All users have equal access to statistical releases at the same time.

**Statistical processes (CoP Principles 7 – 10)**

European and other international standards, guidelines and good practices must be fully observed in the processes used by statistical authorities to organise, collect, process and disseminate official statistics. Examples of principles are sound methodology, appropriate statistical procedures and non-excessive burden on respondents.

**Statistical output (CoP Principle 11 – 15)**

Available statistics must meet users’ needs. The important issues concern the extent to which the statistics are relevant, accurate and reliable, timely, coherent, comparable across regions and countries, and readily accessible by users.

Perhaps one of the most well-known principles is Principle 12: Accuracy and reliability. This principle includes the indicators:

- Source data, intermediate results and statistical outputs are assessed and validated
- Sampling errors and non-sampling errors are measured and systematically documented.

The documentation of the principles and indicators are of great importance for the users, especially the principles and indicators of the statistical output.

The use of some kind of Metadata system is recommended.

**Statistics Sweden’s quality work**

Statistics Sweden has been working with different aspects of quality for some time and CoP is today an important framework. During the last five years much attention has been given to the analysis of the statistical processes.

For the moment the Quality Management system at Statistics Sweden include four key issues:

- Quality framework: **The EFQM Excellence Model**
- Methods for continuous improvement work: **Six Sigma**
- ISO 20252 Market, opinion and social research
- Methods for evaluation and follow-up: **Internal auditing.**

Statistics Sweden works constantly to improve its operations. To assure good quality in our surveys, we have a quality policy and we use standardised routines, guidelines and handbooks in the production of statistics.

Since 2008 statistics Sweden has been using the framework called the EFQM Excellence Model as support for its operational development. The model is based on a number of criteria that are central to the work to improve the quality of operations: leadership, employees, operational planning, partnerships/resources as well as results. This particular framework facilitates cooperation and benchmarking with a number of other organisations.

Since January 2008 a central quality unit has been formed in the Research and Development Department. It coordinates the quality work and today the unit consists of four people. The work is led by the Quality Head who has regular follow-up meetings with the Director-General. At the initiative of the management group, 12 quality coaches were appointed in autumn of 2008 to work part-time with quality issues.

Statistics Sweden strives to standardise parts of the statistics production that have considerable influence on quality and costs. On our intranet we have a feature called Operational support that describes Statistics Sweden's common methods, tools and working methods. Information is available about the various stages in the statistics production process with technical documentation, guidelines, templates and checklists to carry out the statistics production process.

Operational support is a helpful tool for all statistics production at Statistics Sweden. It covers statistical surveys based on direct collection, administrative data, other existing registers or a combination of these sources. The end products are the statistics or final observation registers.

We work continuously to improve the statistics production process. As a part of this goal, Statistics Sweden is working towards certification in 2012 according to the international standard ISO 20252 for marketing, opinion and social research. Among other things, the standard places demands on the system for quality management, customer contacts, data processing, employees' competence and all the stages in a statistical survey.

Statistics Sweden plans to use the tools offered by the Six Sigma system in many of the projects for improvement and development. Three employees from Statistics Sweden have received 20 days of training for the Six Sigma methods. Due to other priorities, work with Six Sigma has been postponed. Statistics Sweden will work on an internal approach for Six Sigma.

When a statistics assignment has been completed, we follow up the results in relation to the expectations of the customer by what we call a customer satisfaction questionnaire. The customer has an opportunity to give his/her viewpoints on how well Statistics Sweden has done the job. The viewpoints of the customers then form the basis for future improvements and development, as well as comprise an important variable in the results section of The EFQM Excellence Model.

STATISTICS IN TRANSITION-new series, December 2011  
Vol. 12, No. 3, pp. 475—484

## STATISTICAL INFORMATION SYSTEM OF THE REPUBLIC OF BELARUS: EXPERIENCE AND PERSPECTIVES

Irina Kostevich<sup>1</sup>

### ABSTRACT

Making optimal decisions in the context of economic integration with the European Union requires a creation of the harmonized information system within the framework of the Eastern Partnership.

Therefore, the official statistics plays a more important role in questions related to improvement of statistical methodology, establishment of information resources, development and implementation of the common standards of statistical data provision and sharing. It is evident that such an information system must be created with consideration for capacities of national statistical systems of the Eastern Partnership member countries.

Presently, the cooperation between the National Statistical Committee of the Republic of Belarus and the Statistical Office of the European Communities has become fairly stable. The Belstat regularly provides the statistics on the following topics: foreign trade, demographics and annual data reflecting socio-economic development of the country (Eurostat questionnaire for countries with economies in transition). The global assessment of Belarusian statistics, carried out in 2007 by Eurostat, showed that Belarus has a comprehensive, consistent, and well established statistical information system which includes all statistical domains allowing for provision of comparable data in most areas.

One of the possible directions for developing the information sharing and creation of the harmonized information system is the use of SDMX international standards (sharing of statistical data and metadata) which are used in the European Statistical System and define a general information model, technical specifications and formats of structured aggregated statistical data and metadata.

Today, within the framework of the State Programme of Creation of the Integrated Information System of State Statistics of the Republic of Belarus, organizational and technical solutions designed to maintain a common metadatabase and a summary information database are under development, with XML format used for data presentation.

---

<sup>1</sup> National Statistical Committee of the Republic of Belarus.

Based on Eurostat's experience of development and implementation of standards, the creation of the harmonized information system within the European Partnership must include stages for analysis of national information systems, employed approaches to metadata structure description, definition of techniques for data conversion in order to meet requirements of technical part of the standard, development of appropriate software for conversion.

It is evident that the creation of the harmonized information system within the framework of the Eastern Partnership, first of all, requires a detailed coordination between statistical offices of the programme's member countries.

## **Introduction**

The initiative for extending cooperation between the European Union with its Post-Soviet eastern neighbours was officially formalised in May 2009 in the declaration adopted by 27 EU member states and Azerbaijan, Armenia, the Republic of Belarus, Georgia, the Republic of Moldova and Ukraine. The declaration defines the Eastern Partnership as measures assumed by the partner countries, based on common interests and mutual commitments as well as on shared ownership and responsibility. The paramount goal of the Eastern Partnership is to create the necessary conditions to accelerate political association and economic integration between the European Union and partner countries and to extend bilateral and multilateral cooperation.

In order to enhance economic integration, the main directions are the extension of trade and investment activity, establishment of free trade areas, development of small and medium sized business ("Economic integration and convergence with EU sectoral policies"), strengthening of energy security, conduction of energy saving and energy efficiency projects ("Energy security").

Certainly, the enhancement of integration and development of these directions require an appropriate information support while the established information resources must be credible and available to general public. In this connection, the vital task is to create a harmonised information system allowing to obtain and efficiently use the data and to support decision making in the context of economic integration of the partner countries with the EU.

To solve this task, the official statistics should focus on such areas of activity as:

- creation of a methodologically coherent and credible information resource on the socio-economic development of the country;
- support of standards on description, dissemination and exchange of summary statistical data.

Therefore, the role of national statistical agencies increases in matters of updating the statistical methodology, extending an information base and providing the necessary level of detail when establishing information resources, developing and implementing the common standards for statistical data presentation and exchange. It is evident that a harmonised information system is

to be created with due consideration for capacities of national statistical systems of the Eastern Partnership countries.

A possible direction for further development of the information exchange and creation of the harmonised information system is the use of international SDMX standards (statistical data and metadata exchange) which are implemented in the European Statistical System and define the general information model, technical specifications and formats of aggregated statistical data and metadata.

### **1. Information cooperation of the National Statistical Committee of the Republic of Belarus with the European Union and the Eastern Partnership countries.**

When speaking of cooperation with statistical agencies of the Eastern Partnership countries, it is important to note a well organised system of statistical information exchange within the framework of the Interstate Statistical Committee of the Commonwealth of Independent States. The CIS member countries provide information to the Interstate Statistical Committee of the CIS according to the Programme of Work, subject to annual review and approval by statistical agencies of the CIS countries. Presently, information on the following topics is provided to the Interstate Statistical Committee of the CIS:

- main macroeconomic indicators and financial activity;
- foreign economic activities;
- main indicators of the real economy (including industry, agriculture, construction, transport and communications, trade and paid services);
- institutional reforms in the economy;
- demographic statistics;
- labour and living standards statistics;
- environmental statistics; etc.

In addition, the Belstat closely interacts with statistical agencies of the CIS countries under current bilateral agreements on statistical cooperation. This type of cooperation allows the Belstat to receive up-to-date information on the socio-economic development of the CIS countries as well as to conduct comparative analysis in such fields as foreign trade statistics, industrial and price statistics, etc.

Lately, the cooperation between the National Statistical Committee of the Republic of Belarus and the Statistical Office of the European Communities has become fairly stable. The Belstat regularly provides the statistics on the following topics: foreign trade, demographics and the annual data reflecting the socio-economic development of the country (Eurostat questionnaire for countries with economies in transition). The global assessment of Belarusian statistics, carried out in 2007 by Eurostat, showed that Belarus has a comprehensive, consistent and well organised statistical information system which includes all statistical domains allowing to provide comparable data in most areas.

I would like to focus on the interaction between statistical agencies of the Republic of Belarus and the Republic of Poland within the initiatives of the

European Partnership. As a part of extension of cross-border cooperation and presentation of the information reflecting the socio-economic development of cross-border territories, statistical booklets “Grodno region and Podlaskie Voivodeship in 2008” and “Grodno and Bialystok in 2009” were published in 2010 and 2011, respectively. A statistical publication “Neman Euroregion” is planned for release.

It is evident that today a development strategy for an information system must cover all spheres of statistical activities, set priorities and directions for enhancement. Its implementation must be based on the statistics development programme in order to achieve goals and actual results including ensuring high quality data and the use of best international practices as well as to consider user needs at all levels and serve as a coherent base for extending partnership relations.

## **2. Potential for implementation of the SDMX standards by the national statistical agency of the Republic of Belarus.**

### **2.1 General characteristics of approaches to the SDMX standards: methodological and technical aspects.**

The International Statistical Data and Metadata eXchange (SDMX) standard is an initiative that was started in 2001 by seven international organisations, among which were the statistical office of the European Union (Eurostat) and the United Nations Statistical Division. These standards define, first of all, guiding principles and formats for data and metadata exchange.

The SDMX standards contain two major components: statistical (content-oriented, based on a common information model) and technical (defining formats and syntax, data exchange procedures).

Approval of the SDMX Technical Standards version 2.0 by international organisations in 2005, their recognition as the preferred standard for data and metadata exchange by the United Nations Security Council in 2008, active use of standardised network services to integrate the existing and to create new resources of statistical organisations as well as further development and implementation of the standards (creation of SDMX Working Groups with global participation for technical and statistical aspects of the standards and release of SDMX Technical Standard version 2.1 in 2011) make it possible to consider them as the most perspective tool for creation of a harmonised information system.

Today, the Eurostat information system is based on the SDMX meaning that description rules and metadata structure based on the use of hierarchical codes and classifications are highly universal. The Euro SDMX Metadata Structure published in March 2009 uses twenty one high-level concepts allowing not only to describe statistical characteristics of data (such as the information on metadata update, classifications used, users of information, territory and survey period, measurement units, accuracy, reliability, validity, comparability) but also to reflect characteristics of the statistical process of data collection, costs and respondent burden, policies of confidentiality, data dissemination and quality control. This approach has a high redundancy and interoperability (an ability for

the system to operate without limitations) enabling integration of the systems of different owners. While the main emphasis in the standards is placed on definition of formats for structured aggregate statistical data presented in time series, cross-sectional data formats are also supported. SDMX technical specifications are based on XML format, widely used today in various spheres.

The results of the SDMX Global Conference held on May 2-4, 2011 in Washington, DC show that more than 80% of organisations are already using or planning to start using these standards. The Belstat also took part in the online survey on acceptance level, challenges and perspectives of the SDMX standards for the purposes of official statistics.

The perspectives of the SDMX standards in international exchange are obvious; however, their use in the national statistical system of the Republic of Belarus and for inter-departmental exchange is limited due to heavy centralisation of the Belarusian system. Therefore, one of the key tasks will be the development of the corresponding software and interfaces allowing for conversion of the existing and developed national bases and their integration with additional information sources to support an automated data exchange with other information systems, e.g. the European Statistical System.

## **2.2. Approaches to metadata description in the Integrated Information System of State Statistics**

Development of a new generation information system is currently under way in Belarusian statistics. Quality changes in approaches to the statistical information system are implemented within the framework of the State Programme of Creation of the Integrated Information System of State Statistics (IISSS), endorsed and approved by the President of the Republic of Belarus.

The State Programme stipulates activities to implement brand new approaches to automate processes of organisation and maintenance of state statistics, i.e. collection, processing, aggregation of raw statistical data; accumulation, storage and protection of statistical information; dissemination of summary statistical data, and is aimed to create an integrated statistical information resource and a set of software and hardware tools and technologies enabling adoption of a common state statistics policy. Within the framework of the Programme developed are organisational and technical solutions for maintenance of a unified metadatabase and summary information databases with XML used as a format for data presentation.

IISSS software components are based on an Oracle Database Management System, involve the development of novel technologies and comprise 5 main subsystems:

**a subsystem for metadatabase management** with the following key elements:

- a statistical register,
- a catalogue of statistical indicator,
- common statistical classifications and references based on national classifications.



**a subsystem for collection and processing of raw statistical data** containing:

- tools for creation of uniform template forms for state statistical observations,
- tools for collection and processing of raw data in electronic form.

**a subsystem for accumulation and storage of statistical information** allowing for management of long-term databases:

- generation of predefined and customizable tables;
- database export for further analytical processing, including external information systems.

**a subsystem for analysis and dissemination of statistical information**, which includes management of a finished document database and a corporate Web system, data visualization tools.

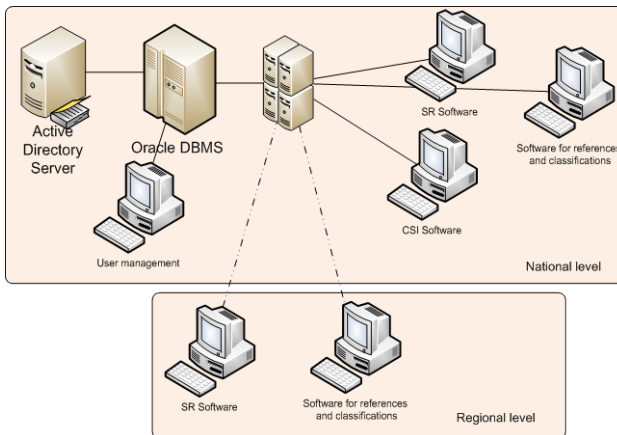
**an information security subsystem**, required for cryptographic protection of transmitted information and serving for the purposes of the state system of public key management, currently under development in the Republic of Belarus, which allows for using the same key to set up a full legally relevant workflow between organisations and state bodies.

A metadatabase is a methodological basis for achieving consistency of data formation principles. A common metadatabase is a tool for methodological coordination of statistical observations allowing to implement principles of integrated statistics.

A metadatabase management subsystem serves for formation and management of metadata used for collection and processing of statistical data at three levels: district, regional and national.

The joint creation and management of the information between different territorial levels are carried out within the Integrated Republican DBMS taking into account interaction of multiple statistical authorities.

A simplified diagram of the interaction is presented below:



Software tools of a metadatabase management subsystem are used for the following tasks:

- management of a statistical register;
- management of a metadatabase for the catalogue of statistical indicators;
- management of a common system of IISSS statistical classifications and references.

Software tools for managing statistical register are designed for registration of statistical observation units and preparation of statistical observation unit list to conduct an observation.

Software tools for managing reference data are designed for centralised management of classification and references used for achieving information integrity of all IISSS subsystems.

Software tools for managing the catalogue of statistical indicators are designed to create and manage a list of indicators upon which an aggregate database (macro data) is compiled.

The catalogue of statistical indicators is a system of metainformation about indicators used in statistical practice and aiming to solve the tasks of:

- formalised description of statistical indicators;
- identifying the values of statistical indicators;
- systematisation of indicators and description of inter-linkages between them;
- search and presentation of reference information on indicators.

Development of the IISSS metadatabase catalogue of statistical indicators employs the following principles for description of indicators within the common methodology of defining an indicator base and its attributive component:

each indicator contains the base reflecting the essence and general distinguishing features of socio-economic phenomena without specifying the place, time and numerical value;

the base must be unambiguous and is characterised by attributes (each base has its own set of attributes from a general range);

hierarchical coding, sensitive to the indicator class, objects under observation, class and base qualifiers, is used for structuring (classification and rubrication);

each indicator is provided with a tutorial containing methodical (methodological) explanations in a text format and a changelog.

In case of macro data the base and attribute of an indicator are assigned with periods and breakdowns (e.g. territorial, by kind of economic activity, etc.) as well as measurement units and directly the indicator value.

The multiple indicator rubrication offers great possibilities facilitating the search of the indicator in question and related meta information allowing for

further export of indicator values for various periods and in the required breakdown from a macro database.

It is worth noting that the principles, implemented in the subsystem for managing the IISSS metadatabase, in general comply with the approaches used in the SDMX standards, however, a wider, full-scale application requires additional works in terms of creation of metadata structure description (e.g. quality management, confidentiality and data dissemination policies, user satisfaction) and development of the required software and hardware tools for integration with information systems of international organisations and national statistical agencies.

### **2.3. Major challenges to the implementation of the SDMX standards on the way to create a harmonised information system**

The programme for establishment of the harmonised information system within the framework of the Eastern Partnership must include stages of analysis of national information systems, approaches used to describe metadata structure, identification of techniques to convert data to meet the requirements of statistical and technical parts of the standards, possibly, creation of special (localised) version of the standard to minimize an impact on internal structures of national statistical agencies, development of relevant software to convert the existing data.

Different approaches of national statistical agencies to the SDMX standards can be viewed as one of the main challenges. Despite overall support (according to the results of the 2011 Global Survey which covered 124 organisations representing six continents), over 10% of the organisations do not plan to use the standard while 50% are not currently using the standard but are planning to implement it in the future. The transition to the standard is slowed by the lack of resources, both financial and human, which is aggravated by the world financial crisis.

The Republic of Belarus is planning to start implementing international standards to set up an exchange of the information with international organisations. Apart from the issues related to the lack of human resources and level of personnel training, necessary technical support, lack of financial resources, which are common among national statistical agencies, we also have to solve the organisational problems of the existing databases, overcome the constraints related to dissemination of the summary information, and fix compatibility issues in order to meet the requirements for data protection.

Eurostat's experience in the development and implementation of the SDMX standards at international level will be of tremendous help to all Eastern Partnership countries.

It is also essential to choose a direction related to the development and further elaboration of harmonised definitions of data and metadata structure. Moreover, it is reasonable to discuss the establishment of a local group consisting of the partner countries to develop common approaches to metadata structure and localized definitions.

It is evident that, in the context of developing the information society and focusing on the use of information resources in management and decision support systems, provision of electronic services, the official statistics must progress by extending the range of information resources and services provided while increasing their quality and ensuring the interaction with information systems of international statistical agencies and organisations. The technological basis for implementing this strategy is the use of web technologies and groupware applications.

Production and wider distribution of an open source and free software as well as an in-depth study of the SDMX potential to be harmonised with other statistical standards can be identified as actual needs of statistical agencies for development and implementation of the SDMX standards.

## **Conclusions and suggestions**

Today, when countries are integrating their economic systems, the immediate issue is to establish a multilateral public information resource on the socio-economic development of the territories and economic relations.

Considering this, the official statistics should tackle the tasks of forming methodologically harmonised and accurate information as well as supporting the standards of description, presentation and exchange of statistical data.

It is evident that the creation of a harmonised information system within the framework of the Eastern Partnership, first of all, requires a detailed coordination between statistical offices of the partner countries.

Implementation of the SDMX standards can be suggested as one of the mechanisms to create a harmonised information system.

It is necessary to study the desire and willingness of the official statistical agencies to use international standards, to make an adequate assessment of the existing methodological and technical challenges and advantages. It is evident that such a transition will be long-term requiring additional expenditures.

Eurostat's experience in the development and implementation of the SDMX standards as well as a more active participation in international workgroups, production and wider distribution of web-oriented free software will be of tremendous help to all participating countries in order to create a joint information resource and harmonised information system.

*STATISTICS IN TRANSITION*-new series, December 2011  
Vol. 12, No. 3, pp. 485—492

## **STATISTICAL SYSTEM OF THE COUNTRIES- MEMBERS OF THE COMMONWEALTH OF INDEPENDENT STATES: 20 YEARS ON THE WAY TO INTEGRATION (CIS-STAT)**

**Irina Zbarskaya<sup>1</sup>**

### **ABSTRACT**

Breakdown of the Soviet Union has clearly revealed insufficiency and in many directions even unsuitability of the existing for many years system of economic and statistical information for assessment of radical changes taking place in economy and social life. Serious reconstruction was needed of the existing fundamentals of primary accounting and statistics by means of reforming common methodological and organizational basis of statistics itself, as well as adjusting it in accordance with international rules and standards.

First of all, a serious step in the movement “towards the west” was the support proposed by The European Union in the field of transmission of “know-how” through the TACIS Program - Program of technical assistance to the CIS countries.

The Program has covered a wide range of directions: reforming of the system of government administration; restructuring of public enterprises and development of private sector, finances and banking sector, development of transport and communication infrastructure, energy sector, creation of efficient production system, etc. But statistics were not forgotten. In the total TACIS budget its share is rather low, not more than 0.2%. However, statistical projects play a special role in the creation of a new information infrastructure and promotion of the efficient exchange of data.

Cooperation with the EU within the framework of the TACIS Program has played an important role in the development of statistics in the CIS countries. However, it should be emphasized that for the development of the statistical potential of these countries it is also very important to interact with statistical services of other leading international organizations, first of all, with UNECE, OECD, IMF, the World Bank, ILO and other.

The solution of the tasks faced by statisticians of the Commonwealth countries was promoted and will be promoted by the discussion of urgent problems at the

---

<sup>1</sup> Interstate Statistical Committee of the Commonwealth of Independent States.

Conference of European Statisticians, active cooperation with statistical organizations of the region.

Big role in building up national statistics was played by the Interstate Statistical Committee of the Commonwealth of Independent States, which at the first stage of CIS existence CIS-STAT became the main moderator of international standards introduction. Today CIS-STAT faces new challenges. First of all, this is methodological unification of different large-scale statistical works.

All the measures taken have enabled the creation and development of the interstate statistical system of the Commonwealth, although not everything took place in favorable conditions. Many national statistical services face serious financial difficulties constantly and for that reason they lag behind in technical equipment and other supplies, can not keep qualified personnel and have to reduce statistical activities. Some national statistical services are forced to give up conduct of a number of important statistical works.

Breakdown of the Soviet Union has clearly revealed insufficiency and in many directions even unsuitability of the existing for many years system of economic and statistical information for assessment of radical changes taking place in economy and social life. Serious reconstruction was needed of the existing fundamentals of primary accounting and statistics by means of reforming common methodological and organizational basis of statistics itself, as well as adjusting it in accordance with international rules and standards.

The last 20 years became the years of fruitful work on the way to integration of statistical system of the countries-members of the Commonwealth into the world standards. Let me briefly dwell upon the main milestones of our common history.

First of all, a serious step in the movement "towards the west" was the support proposed by The European Union in the field of transmission of "know-how" through the TACIS Program - Program of technical assistance to the CIS countries.

The Program has covered a wide range of directions: reforming of the system of government administration; restructuring of public enterprises and development of private sector, finances and banking sector, development of transport and communication infrastructure, energy sector, creation of efficient production system, etc. But statistics were not forgotten. In the total TACIS budget its share is rather low, not more than 0.2%. However, statistical projects play a special role in the creation of a new information infrastructure and promotion of the efficient exchange of data.

The history of the TACIS statistical Program has gone through several stages.

The initial stage of TACIS cooperation in the sphere of statistics took place in 1992-1994 and was based on common interstate priorities. It was focused on mutual acquaintance of the EU and the CIS countries with each others' experience and on distribution of common approaches to the organization of statistics. Statistical programs implemented at this stage were one of important parts of the overall EU program of assisting CIS countries in transition from a centrally

planned economy to a market economy. In general, the first consultations aimed at the transfer of “know-how” in the priority fields of statistics became an important stage in developing close cooperation with the national statistical services of the European Union countries. Specialists from the CIS countries got opportunities to communicate with their European colleagues, to deal with urgent statistical problems in more detail, and to learn from the experience of the European Union.

More target-oriented cooperation in the framework of the TACIS Program became *after the signing of the Agreements on Partnership and Cooperation* with the majority of the CIS countries. These agreements provided new opportunities for the CIS countries to be integrated into the world economic system. They covered a whole range of cooperation possibilities between the EU and the CIS in the political, trade, and economic spheres. Certain articles of these agreements dealt with cooperation in the field of statistics. So, the second stage of the TACIS statistical program implementation was wholly devoted to reaching the objectives set in the bilateral documents with the cooperation being focused at the national priorities and covering a wide range of statistical subjects.

An important stage in extending the cooperation in the framework of the TACIS Program was the implementation of a series of thematic projects on the adaptation of state statistical systems in the CIS countries to new economic conditions. It was closely related to the start of the process to ensure the comparability of the statistical methodologies of the CIS countries with the international standards. Some thematic projects were devoted to specific methodological issues targeting the most urgent problems of the national statistical systems, such as the organization of sample surveys, statistical observation of prices, registers of enterprises, business statistics, etc. At this stage, cooperation was carried out in the form of providing consulting assistance, transfer of “know-how”, and implementation of pilot projects.

Starting from 1998, one of the main requirements to national TACIS programs is *a comprehensive conceptual approach* to their development. Separate projects were replaced by unified programs, where subprojects of the programs were interrelated and subordinate to a common priority task. As a result, the programs implementation period was clearly determined (three years), which allowed the national statistical services to carry out project work in stages with the account of the intermediate objectives.

National programs corresponding to this stage of the TACIS Program were targeted at development of statistical infrastructure (classifications, register of enterprises, regional statistics, dissemination of statistical information) and development of certain important fields of statistics (statistics of industry, national accounts, trade statistics, demographic statistics). A characteristic feature of this stage was the wide use of pilot surveys in the programs implementation, which were targeted at testing the techniques developed by national statisticians together with the EU experts and adapted to the economic peculiarities of the regions. Such an approach allowed a significant increase in the efficiency of



cooperation, when implementation of methods, software products, and statistical forms has become an inseparable part of the Program.

An illustration of it can be the adaptation of the model of demographic forecast proposed by the experts from Groningen University for the calculation of estimated number and age and sex structure of population as a whole for the Russian Federation and for each of its subjects. Thus, the software for this model was developed by Russian specialists. And further on, also in the framework of the TACIS Program the scope of this model was extended under the support of Austrian statisticians which enabled to make a transition from the forecast of population number to the forecast of number and structure of households. At present all the calculations are successfully made by Rosstat.

In 2000's *coordination and harmonization* became the main priorities. In view of the comprehensive approach to programs development, much attention was paid to linking new components of technical assistance to the long-term integrated statistical programs in the countries, to coordination of international programs of assistance to the CIS countries and selected cooperation programs. Consequently, at this stage the TACIS Program has become a tool for leading European experts to test the national methodologies in the key statistical areas. Approved methodologies have become the basis for conducting wider and more costly programs in the framework of national budget financing or with the help of international donors.

The main objective of the Program at the beginning of the new millennium has become the harmonization of the methodological approaches to calculating statistical indicators according to the standards and methodologies used by international organizations, and producing statistical data consistent with the generally practised methodology. This priority was predetermined by further integration of the CIS countries into the world economic system and by the necessity to improve the exchange of data among national statistical services and international organizations.

Another important feature of the recent stages of the Program implementation was the opportunity for the specialists from the CIS countries to take an active part in international statistical activities in Europe. Financial support within the framework of the Program has allowed them to participate in the most important seminars conducted under the auspices of the UNECE and Eurostat.

I cannot but give several figures which clearly show the scale of cooperation. During the past years, about 30 national statistical services participated in the Program implementation. More than two hundred projects, dozens of experimental surveys, and targeted supplies of equipment and software have ensured to a great extent the development of statistics in the CIS countries and have promoted convergence of their statistical methodology to the European standards.

Cooperation with the EU within the framework of the TACIS Program has played an important role in the development of statistics in the CIS countries. However, it should be emphasized that for the development of the statistical

potential of these countries it is also very important to interact with statistical services of other leading international organizations, first of all, with UNECE, OECD, IMF, the World Bank, ILO and other.

As an example, we can recall one of the most fundamental pieces of work carried out in the last decade and related to implementation of a totally new methodology in macroeconomic statistics, that is, implementation of national accounts methodology.

Substantial efforts to promote the implementation of the international standards of the SNA into the statistical practices of the CIS statistical services were made by OECD, whose experts were the first to take an active part in solving this global problem starting from the end of 1980's. Their experience, and their professional approach to the solution of arising questions contributed much to the successful start of this work. Later on, representatives of the Statistics Directorate of the OECD organized methodological seminars in many CIS countries and participated in experimental practical calculations.

Not less significant contribution to the implementation of SNA methodology in the CIS countries was made by representatives of the World Bank. The reports prepared in the mid 1990's, containing the analysis of the SNA implementation process and recommendations for its improvement, played a very important role in efficient continuation of this work. In several countries these recommendations were of a revolutionary character, and their publication greatly assisted national statistical services in advancement of this important work. Rather significant for increasing the general efficiency was the involvement by the World Bank of national experts in preparation of the reports. It was one of the ways to transfer international experience.

Cooperation with the IMF Statistics Department was also very important for successful implementation of the SNA. Detailed consultations and methodological seminars on the SNA implementation provided by its experts were very supportive to the work of the statisticians in the CIS countries.

Consultations and working meetings organized by the UNECE Statistical Division were also very important. Participation in these meetings provided CIS statisticians with a unique opportunity not only to focus on their urgent problems, but to see the whole range of issues related to the modern development of national accounts. It is of an invaluable assistance because it helps to see perspectives more clearly and choose the optimal way.

Many examples can be given from other fields of statistics, for example, participation of ILO experts in the development of labor statistics, price statistics, etc. All these examples are evidence of efficient cooperation with many international organizations with the purpose of solving the main strategic task in the field of statistics - transition from the methodology accompanying the centralized economy to the methodology of data collection and processing consistent with the market economy.

All the efforts described above have enabled the creation and development of the interstate statistical system of the Commonwealth, although not everything

took place in favorable conditions. Many national statistical services face serious financial difficulties constantly and for that reason they lag behind in technical equipment and other supplies, can not keep qualified personnel and have to reduce statistical activities. Some national statistical services are forced to give up conduct of a number of important statistical works.

The solution of the tasks faced by statisticians of the Commonwealth countries was promoted and will be promoted by the discussion of urgent problems at the Conference of European Statisticians, active cooperation with statistical organizations of the region.

The Conference of European Statisticians retains its leading role as "the coordinator of coordinators" in the field of international statistical cooperation. Along with that it, has one more very important significance for statistical services of the CIS countries: it provides a unique opportunity for their heads to participate in general discussion of the most important problems of the modern development of statistics. In rather difficult situation, in which CIS countries statisticians have to work, such the discussion helps them to plan efficient strategies for further development.

Now it is possible to ascertain with confidence that national statistical services of the Commonwealth countries, some in better, some in worse way have managed to form practically new national statistics, adequate to socio-economic situation of the societies in their states and meeting, as a whole, *the Fundamental principles of official statistics approved by the UN Statistical Commission*.

Some words about the activities of the Interstate Statistical Committee of the CIS, which will celebrate its 20<sup>th</sup> anniversary together with national statistical services. The past twenty years were a complicated period when it was needed to put all statistics on a new footing.

At the first stage of CIS existence CIS-STAT was the main moderator of international standards introduction. For a period since 1992 till 1995 more than 160 methodological recommendations prepared with the account of international standards were developed and sent to national statistical services of the Commonwealth.

But, perhaps, most difficult both in the work of CIS-STAT and in practice of national statistical services was to persuade politicians, members of parliaments of the necessity to organize statistical activities on a legislative basis. **And it can be stated with satisfaction today that laws on statistics have been adopted in all the CIS countries; laws on population and agricultural censuses have been adopted in a number of countries as well.**

Now methodological function of CIS-STAT fades into the background because national statistical services of the CIS have gained experience and have highly professional specialists who participate in the elaboration of standards at international level.

Today CIS-STAT faces new challenges. First of all, it tries to combine different large-scale statistical works in methodological and time unity.

In particular, this refers to 2010 round of population censuses in the CIS countries. By initiative of the Council of Heads of Statistical Services of the CIS the Council of Heads of Governments of the Commonwealth has adopted a decision to conduct population censuses in the CIS countries in maximum close dates. In the framework of specially created Coordination Council a number of model organization and methodological documents were elaborated and summary tabulation was reconciled for the purpose of unified presentation of information.

Unfortunately, the countries couldn't manage to realize to the full extent time unity in the conduct of population censuses. At present seven CIS countries have conducted the censuses. Other countries will conduct their censuses in 2012-2013.

We are waiting impatiently for the results of the censuses. Demographic situation in the countries remains rather complicated. Intensive migration processes are going on, the structure of population is changing. The process of aging and depopulation has put the problem of labor force on the first place. Complex analysis of the results of national population censuses is needed to understand and forecast the situation at labor market and migration trends not only in the Commonwealth countries but in the European region as a whole.

Speaking about the population censuses one more problem should be mentioned. The entire statistical community is preoccupied with perspectives of preparation for 2020 round of population censuses. We see how social composition and, correspondingly, mentality of population are changed and censuses become litmus paper which clearly shows these changes. At present CIS-STAT is preparing jointly with Canadian statisticians a special seminar in the framework of the Conference of European Statisticians devoted to the results and lessons of the last round of population censuses and preparation for the next round.

Big work, which CIS-STAT is dealing with at present, is the translation into Russian and scientific editing of the text of the last version of the SNA 2008, which was approved by the UN Statistical Commission. By the end of this year the translation and scientific editing will be completed.

In 2010 CIS-STAT began preparatory works on the conduct of the second 2011 global round of comparisons of GDP and purchasing power parities of currencies. The first round was realized in 2005 and was considered to be successful. This very complicated and interesting work is an example of statistics of future. In the current global round 186 countries participate and there will be carried out practically one stage unified methodology observation of 3 000 commodities selected on the basis of strict specification. All this work is concentrated in one centre (the World Bank) and during 2012 it is envisaged to receive the first data.

In 2011 the World Bank allocated grants for two very important projects which were already started by CIS-STAT. The first is the project on training

personnel of the CIS statistical services. At present priority themes and partners, who will prepare training courses, are determined. The second big project is devoted to complex study of labor migration in the CIS countries.

\*

\*

\*

During the past years much was done and today we can positively say that thanks to joint efforts of international organizations and national statistical services modern system of statistics, meeting international standards has been created in the Commonwealth region.

## **STATISTICAL PARAINFORMATION PLATFORM – FOUNDATION OF INFORMATION TRANSPARENCY IN HETEROGENEOUS INTERNATIONAL ENVIRONMENT** (with special reference to the Eastern Partnership initiative)

**Józef Oleński<sup>1</sup>**

### **ABSTRACT**

1. Transparency and coherence of social and economic information between countries is the prerequisite and of international cooperation and strengthening relations based on partnership and mutual trust. The closer are institutionalized and informal relations between countries, their governments, NGO's, businesses and citizens, the stronger is the need for transparency in all spheres of social and economic information.
2. The mission of official statistics in globalized, more open economy is to (a) providing methods and tools of quality control of public information, (b) producing information relevant to the needs of all users and (c) widening and facilitating the accessibility and international interchange of information.
3. In modern ICT environment the international parainformation platform embedded in official statistical systems of cooperating countries is the optimal, efficient, easy to implement and relatively cheap instrument to provide information transparency in international, heterogeneous environment.
4. General conceptual foundations and function of the platform are presented. It is stressed that the participation in information interchange *via* the platform does not need to introduce any changes in existing information systems of cooperating countries. The value added of the platform for all stakeholders would be the homogeneous retrieving of data from accessible systems and data bases both statistical and administrative systems via the platform. Important value added would be the integration of statistical parainformation resources of countries in one database, i.e. the base storing verified and updated information on accessible systems, information resources and the institutions managing those systems and resources of all cooperating countries and organizations.
5. The role and contribution of the statistical parainformation platform in strengthening the integrity of statistical information systems and relevant administrative and scientific information systems on international scale, with special reference to the needs of the Eastern Partnership, are discussed.

---

<sup>1</sup> International Institute of Transborder Studies, State High School of Technology and Economics, Jaroslaw, Poland. E-mail: j.olenksi@onet.pl.

## 1. Introduction

Statistical observation, measuring, analyzing and monitoring of social, economic and ecological processes in globalized, more open, market – driven, knowledge - based economies and in information societies need methodological approaches adjusted to their specific features and attributes.

Main theses of this paper are following:

- (1) information transparency, integrity and interchange between countries is the prerequisite of political, social and economic cooperation and partnership;
- (2) national statistical offices are predestined to developing, maintenance and operational management of information systems supporting the coherence, quality control and information transparency between countries;
- (3) *international statistical parainformation platform (ISPP)* embedded in and supported by national official statistical systems of countries is optimal, most efficient, easy to implement and relatively cheap instrument of supporting information interchange and transparency in international, politically and economically heterogeneous environment.

The needs for increasing of international transparency and interchange of information and growing information requirement of cooperating subjects based in different countries are the consequence of tightening international and supranational interdependencies of economic and social processes, i.a.:

- a) On international level legal and institutional systems are decisive for economic and investment decisions of businesses and individuals; national and regional governments use the procedures of deep institutional interventionism to regulate the processes of economic cooperation;
- b) Technological and economic processes in many economic branches are realized on the territory of two or more countries;
- c) Growing number of enterprises based and registered in different countries interlinked and realizing interdependent processes of production, trade and financing with enterprises of other countries;
- d) Foreign direct investments are important factor of economic development for all countries; foreign financial investments have strong impact on financial stability of national businesses;
- e) Financial institutions constitute integrated international financial market, especially in the field of banking and insurance;
- f) International remittances and transfers play important role for economic and social stability and development of national economy and regions;

- g) Growing importance of transborder cooperation for regional development of neighboring countries;
- h) For small - scale national economies and for specific branches of economy the transborder processes may have the impact on macroeconomic situation;

Information infrastructure of the countries cooperating within the frames of the Eastern Partnership has the following specific features and attributes important - from statistical point of view – for information transparency and interchange:

- a) Particular countries have developed their own, specific information infrastructures; because of that for social and economic subjects of one country the access to information from other country is often difficult, troublesome and expensive;
- b) Terms, concepts and definitions used in governmental administrative records and systems, primary records of legal persons and entrepreneurs, are specific for each country;
- c) Concepts and definitions used in social and ecological statistics (in broad sense) are adjusted to legal and administrative specificity of particular country;
- d) National systems of official statistics are integral parts of those infrastructures;
- e) In open economy statistical concept of “national economy” is fuzzy; economic processes in national economies cannot be explained and monitored without information from other countries;
- f) National systems of official statistics have implemented or are on the way of implementing international statistical principles and standards (mainly ISI, UN and specialized organizations of the UN, EU).

Because of wide use of international statistical standards national official statistical systems are more coherent and integrated than other administrative, economic and social national information systems. Higher level of integration, coherence and compatibility of statistical systems could and should be achieved by building international platform providing access and interchanging metainformation and parainformation between national statistical systems.

## **2. Basic concepts and definitions**

The definitions of terms and concepts presented below are commonly used in theory and practice of statistical systems modeling. The objective of presenting those definitions is to pay the attention on specific, in some sense new, functions of metainformation and parainformation in national official statistical information systems (NOSIS) in modern ICT environment.



Statistical metainformation is the information on statistical information realizing in the NOSIS the following functions related to statistical information and data:

- Standardization
- Identification
- Consolidation
- Integration
- Interpretation
- Evaluation of data and information

Statistical parainformation is the information on statistical systems, processes and resources realizing in the NOSIS the following functions:

- Functional specification of NOSIS
  - Standardization of NOSIS
  - Documenting of NOSIS
  - Organization of NOSIS
  - Co-ordination of NOSIS
  - Retrieval of information and metainformation
  - Management of statistical resources, processes and systems
- Coherence of NOSIS's is the *isomorphism* or *equimorphism* of functions, organization and procedures of NOSIS's represented in national statistical parainformation systems. **Coherence is the category referred to the level of statistical systems, processes and resources of NOSIS's.**
  - Integrity is the *isomorphism* or *equimorphism* of relevant statistical data and information in different NOSIS. **Integrity is the category referred to the items of data, metadata, information, metainformation, paradata and parainformation.**
  - Transparency between NOSIS' is the unabridged availability and easy access to complete resources of statistical metainformation and parainformation for all stakeholders, both national and foreign. **Transparency is the category that is referred to the stakeholders of the NOSIS's.**
  - Interchange between NOSIS's is the availability, accessibility and transfer of data, information, metadata, metainformation, paradata and parainformation between NOSIS's and all stakeholders of official statistics. **Interchange is the category describing relations between different NOSIS's as well as between NOSIS's and stakeholders of official statistics.**
  - Statistical parainformation platform is the warehouse realizing the functions of collecting, statistical metainformation and parainformation from all NOSIS's, integrating, organizing those metainformation and parainformation in warehouse structures, providing the access for all stakeholders, and links between meta- and parainformation stored in the platform with relevant statistical data stored in particular NOSIS's.

It should be stressed that the terms defined above should be understood as categories relevant not only to statistical data, metadata and paradata, to statistical information, metainformation and parainformation, to statistical systems processes and resources but also to data, information, metainformation and parainformation and systems participating in statistical processes, i.e. administrative information systems and administrative records used for statistical purposes and primary records and information management systems of respondents supplying data to statistics.

### **3. Information transparency and its role in international cooperation**

In globalized world, in information society and in knowledge – based economy, all political, social and economic subjects are acting in the environment of large and growing information gaps. The reducing of those gaps could be achieved by implementing and maintaining metainformation and parainformation systems oriented for the needs of specific groups of users, i.a. public administration, social and cultural organizations, political organizations and politicians, businessmen, researchers, teachers and students etc.

Providing methods and tools of quality control, accessibility and international interchange of information for specific, professional groups of users is the mission of official statistics on national, regional, international and trans – boarder scale. In modern IC environment official statistics should not only publish statistical data in the form “ready to use” by individual users, but also to create effective data retrieval systems supplying reliable, of good, verified quality, accompanied by relevant metainformation necessary for proper interpretation, understanding and use data. In modern ICT environment the metainformation platform for common use all stakeholders from the countries of partnership seems to be most effective tool for information interchange.

In the international environment of the countries cooperating within the initiative of Eastern Partnership the creating and maintenance of effective information retrieval systems enabling access to information stored in systems and resources managed by different organizations and subjects based in different countries, working in specific conditions of particular countries, is of special importance. In the community of Eastern Partnership there are collaborating numerous organizations, central, regional and local governments and subjects that are acting in different legal, cultural, social, economic, technological and environmental conditions of their own countries. Those countries have achieved different level of economic development; they have different legal systems, different competences and organization of governments, different regulations, procedures and practices of managing information.

Each country has developed its own information infrastructure. Legal regulations and practices of information production, storage and dissemination may not be compatible. Those differences are often the reason of the lack of

transparency of foreign information infrastructure and often cause difficulties in the exchange of information between managers of information systems, especially official information systems. Harmonization of these rules, procedures and practices is the complex process of harmonizing the laws of information management by governments and other subjects responsible for national information infrastructures.

Official statistics is – happily and unfortunately – the only official infrastructural information system that has achieved most high level of harmonization, standardization and coherence on global scale. Thanks to that, national official statistical systems are also relatively well harmonized on international and national levels. That is why national statistical offices and their regional units seem to be predestined – better than other infrastructural information systems - to creating methods and tools of information transparency between different countries. For information transparency of the countries of Eastern Partnership high level of compatibility and transparency of NOSIS's is extremely important.

As it was mentioned above, international cooperation in all fields of social life and in economic activities is based on trust of cooperating political, social and economic subjects. This trust is determined by information transparency of all parties on national, regional, local levels as well as on information transparency of particular political, social and economic subjects. The trust between organizations and individuals is based on open interchange of reliable information. Any gaps or difficulties in availability and accessibility of reliable, relevant, pertinent, complete political, social and economic information and its interchange between social and economic subjects cause disturbances of collaboration and may lead to blocking of mutually effective cooperation.

Except typical, simple, everyday situations, those subjects need more information that they have for their permanent disposal in their own resources. They have to get easy, transparent access to information that is relevant and pertinent for their specific usage situation. Usually the subjects are creating their own information environment, however it is limited to national administrative, statistical and commercial information. Important direct source of those information are mass media and in specialized public or commercial information services (now often portals in internet).

Economic and social information in mass media and in specialized services is extremely redundant. In the situation of extreme redundancy of social and economic information the retrieval of relevant and pertinent information for specific usage situations is difficult and expensive for decisions and practical activities of subjects. It is often not possible for the users to evaluate the quality of retrieved information.

Modern ICT technologies are providing effective tools for browsing very large information files and retrieving information in different forms of numerical data, texts, pictures, as well as in multimedia forms. However ICT tools are not solving growing problems of control of quality, completeness, timeliness,

comparability and pertinence of accessible information for particular users in their concrete usage situations.

The problem of control and evaluation of quality and pertinence of retrieved data is extremely important in global, international and supranational scale. For end – users that need information on other country the access to pertinent information is very difficult, sometimes not possible in due time.

Transparency and coherence of social and economic information on the level of countries is the prerequisite and of international cooperation and strengthening relations based on partnership and trust. This fact is well understood by international statistical community. Because of that, statistical services of global and regional international organizations put a lot of efforts to develop and maintain global and specific information standards and principles of information interchange within the frames of official information systems maintained or coordinated by those organizations.

Global system of official statistics coordinated by the United Nations in cooperation with regional and branch – oriented international organizations, is the largest information system of generating and disseminating information covering all areas of social, economic and ecological processes. It should be also remembered that very important function of the global system of official statistics is the elaborating and proclaiming information standards for statistical information as well as for many other administrative and economic information. Those standards refer mainly to *metainformation* (information about information) and *parainformation* (information about information resources, processes and systems) (see p. 4 below). Information availability and accessibility refers both to information describing phenomena and processes of real world as well as the *metainformation* and *parainformation*.

Common metainformation and parainformation platform seems to be most effective methodological and ICT solution for interchange of official and verifiable information between countries of the Partnership.

#### **4. Transborder statistics – new challenge of official statistics in integrated international European environment**

Important attribute of political and economic cooperation, especially in the European region, is the change of social and economic functions of political borders. Political borders between countries in large part of the world have played the function of geographic delimitation of national economies and societies. Political borders separated different political, legal, social, economic and monetary systems. Global system of statistics was based on the concept of national economies.

Political changes accelerated by democratization, integration processes and transition processes in many regions of the world have created new situation for official statistics. Basic statistical entity of global statistical system (coordinated

by the UNSC) – *national economy* – became fuzzy. Moreover, some sectors of national economy are operating on international and global scale, e.g. banking and finances, transport, communication, information sector (mass media), research and development etc. The fuzziness of national economy as a statistical category has the impact on the system of national accounts and on the interpretability of basic macroeconomic statistical categories. The macroeconomic indicators of the SNA have to be estimated on the basis of data from collected within the frames of countries, while economic and social processes are crossing the borders, and economic subjects are operating on many national markets. The estimates of GDP and related categories are based to large extent on conventions, less on precise algorithms and on complete source data. In transition countries the precise estimation of statistical aggregates is more difficult because of dynamic changes of information environment of statistics generating gaps and holes in reliable information sources.

Specific information needs of users representing regional governments and entrepreneurs were generated by dynamic social and economic processes that are taking place in transborder regions. In small scale national economies statistics the transborder processes is also of interest of central governments.

The dynamics of transborder processes extremely high along the borders between economic systems that differ much in the areas of laws, economic regulations and of the level of economic and social development (laws regulating economic activity, taxes, wages and salaries, laws on labor, social insurance, health insurance, environmental regulations, formal administrative procedures and their informal, practical implementation, like corruption and safety).

The “differences of potentials” between neighboring regions are stimulating and accelerating the activity and cooperation between enterprises. Good political relations between countries and regional governments are facilitating the cooperation of local governments along the border is also stimulating economic and social cooperation. Statistical identification and measuring of the “differences of potentials” in transborder regions may help the governments to support positive processes and phenomena of transborder cooperation as well as to eliminate or reduce negative social and economic processes.

Main statistical indicators characterizing the “differences of potentials” are the indexes or variables characterizing the differences in:

- Prices of comparable goods and services,
- Wages and salaries,
- Access to labor markets,
- Laws regulating labor markets,
- Supply of goods and services,
- Access to social services (health, education),
- Laws regulating economic activity (taxes, social insurance, reliability of financial system, risks of economic activity etc.),
- Ecological laws and practices,
- Safety and security of economic activity (laws, transparency, anti-corruption measures)

- Quality of infrastructure (transport, energy, social infrastructure),
- Access to the markets of other regions (e.g. to the EU single market, to free trade zone of the CES as a whole),
- Policy in the field of non-registered economic activity and shadow economy.

The experiences of Polish statistics (Regional Statistical Office of Rzeszow and the Center of Transborder Statistics in Krosno) have elaborated specific methods of delimitation of transborder areas, methodology of monitoring and evaluating the differences of potentials and synthetic indicators of transborder cooperation. Those methods were verified, and implemented in cooperation with regional statistical services of Slovakia, Ukraine, Belarus and Russian Federation (Kaliningrad region) in the transborder regions along east, south and north boarder of Poland with those countries. The results of the transborder surveys and analyses have proven that transborder statistics should be considered as the specific domain of official statistics. It was also proven high importance and usefulness of specific transborder statistics methods for monitoring and explanation of economic and social processes, especially on transnational regional level.

It seems that in progressing liberalization of international trade, more free transfer of goods, services, money (investments, remittances) and more free migration of people, growing scale of international infrastructural projects, the transborder statistics should become standards segment of official statistics.

Information produced by transborder surveys may help governments to identify and evaluate different forms of transborder cooperation, to program and plan joint actions and projects in for optimizing transborder infrastructure, establishing proper rules of competition stimulating the development on both sides of boarders.

The Krosno - based Center of Transborder Statistics of the Regional Statistical Office in Rzeszow in cooperation with universities and local governments is preparing the project of monitoring and surveying the external Easter border of the European Union "*From Barents' Sea to Black Sea*". Preliminary interest in joining this project was expresses by most of the statistical services of the countries along this borderline. It seems that this project would be good opportunity for more extensive development of specific methods of transborder statistics and for verification of those methods in statistical practice. It is expected, that the experiences of statistical identification, monitoring and analysis of transborder processes of the countries along the borderline from Barents Sea to Black Sea, that represent highly developed economies, post-transition countries and the countries in different advancement of transition processes in society, economy and in official statistics, would contribute to the progress of official statistics on global scale.

## 5. Information coherence and transparency as the prerequisite of international cooperation

As it has been mentioned in the introduction, international information coherence and transparency is the prerequisite of social and economic cooperation of governments, businesses, organizations and individuals. The transparency and coherence refers to all kinds of information: legal, administrative, statistical, to all areas of life: economic, social, cultural, ecological as well as to any kinds of alert information.

Particular classes of political, social and economic subjects need different kinds of information, presented in forms and technologies, adjusted to the possibilities and specific conditions users and other stakeholders of information systems.

The governments and other organizations managing infrastructural information systems of cooperating countries of Eastern Partnership are expected by the stakeholders to develop and maintain information systems producing and providing to the users pertinent, verified, accessible information. Any information gap may cause of lack of trust, generating uncertainty and hesitations of economic and social subjects in undertaking economic and social activities, joint projects or investments in other countries, strengthening international cooperation.

Information transparency can be achieved if and only if in all collaborating countries or regions there is the reciprocal *information equilibrium* i.e. the compatible quality, availability and accessibility of information in all domains of life, in which all subjects are acting.

The reciprocal *information equilibrium* applies to all classes of infrastructural information systems, i.e.

- official juristic information and services
- organizational information on governments and institutions,
- information on administrative procedures,
- public alert information systems,
- information systems managed by public governments on central and regional level (administrative registers, tax systems, social insurance registers, registers of local and regional governments),
- information systems of businesses, especially information systems of enterprises offering services for the public and businesses (supply of energy, telecommunication, consulting firms, safety, security and insurance etc.),
- mass media and specialized public information services,
- official statistics.

Most important role in building trust thank to information transparency is played by public information systems managed by governments, official juristic and procedural information systems, mass media and official statistics.

## 6. Role of official statistics in information coherence and transparency

Official statistics, because of its special position and functions in the information infrastructure of the country, is responsible not only for providing qualitative, verified statistical information as the public good for any stakeholders. It is also responsible for supplying all metainformation needed for proper absorption and interpretation of users of information.

Official statistical agencies are (or should) be developing complex *metainformation* systems and metadata warehouses supporting the realization of all functions of metadata mentioned above, i.e.

- Standardization
- Identification
- Consolidation
- Integration
- Interpretation
- Evaluation of data and information
- Documenting of methodology

Official statistical agencies are also developing (or should develop and maintain) official public parainformation and parainformation warehouses supporting the realization of the functions of the *parainformation* in the information infrastructure of the country, i.e. (as mentioned above, the functions supporting:

- Specification of official infrastructural systems
- Standardization of official information systems and resources
- Organization
- Co-ordination
- Retrieval of information and metainformation
- Management of statistical resources, processes and systems
- Documenting of information processes, systems and resources

In modern ICT environment official statistics may effectively support other official, public and administrative information systems by providing them good information standards, professionally elaborated *metadata* (classifications, nomenclatures, typologies, coding systems, registers), and professionally elaborated *paradata* (standards and tools for organizing, documenting, identifying and retrieval of systems, databases, data files, stakeholders of information processes and systems, procedures of management of information systems etc.).

Nowadays the responsibility of official statistics for information equilibrium should not be limited to official statistics only, but the duty of NOSIS as an important, specific layer of information infrastructure of the country and as the segment of international (e.g. EU, ECE) and global (e.g. UN) information system, is to help other infrastructural information systems of countries to obeying general information standards, integrity and quality rules and requirements on national and international level. It seems that in global ICT



environment and in more active international cooperation the integrating and standardizing function of information infrastructure is an important mission of official statistical agencies.

## **7. Practical problems of statistical information transparency in varied international environment of Eastern Partnership**

From the point of view of statistics, the NOSIS's of countries taking part in the Eastern Partnership initiative could be classified to three classes:

- a) developed market – driven economies,
- b) post - transition countries,
- c) countries in transition.

Ad (a) The NOSIS's of developed market - driven economies were developed in long lasting, slow, evolutionary process of adjustment of statistics to the changes of economic and social life, to new needs of users, to new international statistical standards introduced UN and other organizations and new ICT. International standards were elaborated with active, often leading contributions of experts form statistical agencies of developed market economies.

Ad (b) The NOSIS's of post – transition countries are the systems created on the basis of statistics of centrally planned economies. They have passed accelerated (*revolutionary*) process of deep reconstruction of legal and methodological foundations of economic statistics, rapid transition from MPS – based economic statistics to SNA standards. Main changes consisted in replacing MPS based metainformation – classifications, nomenclatures, terms and definitions, by SNA classification standards, often in their versions adopted by EU. The NOSIS's in many post-transition countries have also passed the process of reorganization of relations and cooperation between the NSI's, ministries, central banks and regional governments. Usually the ministries take more active part in statistical production processes, collecting and processing statistical questionnaires and modernizing their administrative information systems. The decentralization of production of statistics was not always accompanied by strengthening the coordinating role of central statistical institutes in the field of methodology. Because of that in some fields of statistics the excessive methodological diversification is observed. The NOSIS's of post-transition countries have also passed very deep technological changes of statistical processes. This technological reconstruction, modernization and upgrading is continued.

Ad (c) The NOSIS's of the countries - in - transition are in different phases of accelerated, dynamic changes in all fields of statistical activities: statistical laws, programs of surveys, statistical categories, metadata, organization and technology of statistical processes, statistical capacity building adjusted to dynamically changing needs and requirements of all stakeholders of statistical processes and systems. Those processes of transforming the NOSIS's are advanced in different degree, according to the strategies of transition of economies and of social life.

The models of the NOSIS's in the countries of all three classes characterized above are different. Main reasons of the differences result from the specificity of political, legal, social and economic model of particular country. For the post-transition countries (b) and countries-in-transition (c) the differences are also the result of the adopted in practice strategy of transition of official statistics from centrally planned to market driven economy. Those differences are one of the most important reasons of gaps in transparency, coherence and compatibility of statistical systems. They also cause practical problems of accessibility and interchange of information between countries and stakeholders of statistical processes.

Practical procedures of dissemination of statistical data and metadata and of user's access are regulated by national statistical offices. In practice the laws, organization, procedures, economic conditions and practical information services offered by statistical offices are different in each country. For incidental and casual users, especially for foreign users from other countries, this situation is rather uncomfortable. In special cases it is possible in some countries to hire the *infobrokers* (firms specialized in providing information from different sources on individual request of users, on commercial basis), however the costs of information services offered by *infobrokers* are high and the quality of services is rather uncertain.

Often the institutional end – users (governments, businesses, social organizations) and the scientists, researchers and advisors need information from many countries and regions. They are looking for coherent, uniformed, interpretable, comparable data. Many of them need statistical information matched with administrative, scientific and commercial information. In such cases retrieval of complete, relevant, pertinent information from many countries and from many sources is extremely complicated for regular users, and for casual users it's a real pest.

The duty of official statistics is to develop end – user friendly information retrieval systems facilitating easy access of end – users to complete and pertinent information. International partnership of statistical agencies is the opportunity for developing statistical and related information retrieval systems on international scale.

## 8. Official statistical metainformation and parainformation standards as prerequisites of international information coherence

The foundation of transparency of social, economic and ecological processes on international scale is the implementing of relevant international statistical standards by all partners. Extremely important are metainformation and parainformation standards.

The process of implementing international metainformation standards in official statistics is rather advanced and the use of them in statistical surveys is rather common. Statistical agencies that are using national classifications, nomenclatures and definitions of terms, have elaborated, are maintaining and updating the classifications, nomenclatures, code lists, glossaries of terms, correspondence tables and methodological comments to definitions of concepts, algorithms of computing indexes and derived indicators.

In transition countries as well as in post-transition countries the processes of implementing new standards and methods in official statistics are rather well documented. This documentation is helpful for external end users for retrieval and interpretation of data. However often those detailed documentation is not available for the public, on the website of statistical agency. The access to full metainformation resources, including detailed documenting of methodology, is the task of statistical agencies. This obligation was directly was expressed in the UN Fundamental Principles of Official Statistics.

Official statistical and administrative information systems are very complicated and non-transparent for external users, even for regular users. Usually the retrieval of relevant data is the process of several stages of identification:

- 1) country or region
- 2) information system
- 3) survey
- 4) data base
- 5) data file
- 6) pertinent data
- 7) metainformation relevant to retrieved data

The multi-level process of retrieval and access to information is realized with the help of *parainformation*. As it was mentioned above, the information on information systems, processes, resources and stakeholders is called *parainformation*<sup>1</sup>. The development of *statistical parainformation systems* is in rather early stage of development. Methods of designing and managing parainformation have been developed by information scientists and widely adopted in librarianship and in scientific and technical information systems. It

---

<sup>1</sup> The term *parainformation* was proposed by ICT experts for information on information systems, processes, resources, and stakeholders of information systems and processes. This term is not correct etymologically, but it occurred very useful in practice.

seems that statisticians should study the methods and practical experiences of librarians and adapt them creatively to the specificity of statistics.

For effective, user – friendly retrieval of statistical data there are necessary coherent metainformation standards and - what is still in the phase of research and experimental implementations – *statistical parainformation standards*. Harmonization of parainformation standards on international scale is still the future, hopefully not very distant future.

Practical information retrieval in heterogeneous information systems environment requires complex parainformation bases and metainformation bases. End – users should be given the tools for full identifying the existence of pertinent information in information systems, databases and publications. They should be navigated, how to access pertinent data, what are legal, economic administrative and technical constrains and conditions of access and use of required information. The answer to such questions shall be given by statistical parainformation bases. After getting positive answer from parainformation base, the users should be navigated to next phases of retrieval, accessing metadata base, formulating detailed queries in metadata-based retrieval language. The end - users should get final information together with all relevant metainformation.

However it would be a *wishful thinking* to expect that standards harmonizing parainformation on international scale will be commonly used in short time. Official statisticians are in the beginning of developing harmonized parainformation systems. What seems to be realistic is the designing of **common platforms for storage of structured parainformation**. The idea of such platform is presented below.

## **9. Common platform of statistical metainformation and parainformation – the tool of international information transparency and interchange**

In many information systems (librarianship, scientific and technical information, business information etc.) effective tools supporting transparency and interchange of information on international scale are metainformation and parainformation platforms. In librarianship and in scientific information systems the interchange of information via metainformation and parainformation platforms is rather common. However in administrative and statistical information systems those methods and experiences are known by IT researchers. Statisticians have not paid the attention to the achievements of their colleagues from libraries and scientific information management centers. It seems that main problem of rather conservative approach of statisticians to information retrieval methods and techniques is the monopolistic position of official statisticians on their segment of information market, the monopoly for production and dissemination of official statistical data man metadata. However this monopolistic position has come to the end in the field of dissemination. Dissemination of official statistical information

is in hands of specialized portals, professional mass media as well as other intermediaries on information markets (national and international press agencies, *infobrokers* etc.).

Dissemination of statistical information and metainformation by the mass media and other intermediaries for the public, for non-professional, casual users could be accepted by statistical offices, if the mass media and intermediaries obey the rules of precise representation and interpretation of real content of statistical data. In case of dissemination of erroneous data, erroneous interpretation and incorrect presentation of statistical variables, indicators and indexes, statisticians should actively react explaining the errors (see UN Fundamental Principles of Official Statistics).

In modern ICT environment official statistical institutes have got new, exceptional opportunity of direct dissemination of statistical information to all professional users and to the public using Internet. The modern ICT enables also to define individually profiled information services for regular professional users and to provide direct information services for “VIP-users”. Those users could also be offered direct access to statistical data and metadata stored in database system or data warehouses.

Problem that needs improvement is the lack of simple, reliable, end - user friendly query languages for data and metadata retrieval. Usually each survey has its own metadata. Each database system is using specific procedures for accessing data and retrieving relevant information. On international level those problems are much more complicated. The postulate - often met in statistical ICT literature - standardization *ex ante* of all catalogues of statistical variables in the NSS, harmonization of names of variables and developing on this basis one query language for all surveys and data sets generated by surveys seems to be pure wishful thinking.

Much more realistic is the developing of *tailored metadata and paradata bases* realizing the functions of the gateways between end – users and statistical data stored and maintained in existing forms and structures. Those retrieval gateways – *metadata and paradata platforms* – are scalable according to the possibilities and need of statistical systems and end-users. They could be developed also for heterogeneous complex of many NOSIS’ in the form of common meta- and parainformation platform. The meta- and parainformation platform can also work in multilinguistic environment, storing metadata in many national languages and maintaining multi-linguistic correspondence tables .

The concept of the statistical common meta- and parainformation platform for retrieval and dissemination of statistical and related data is the adoption of similar platforms that are constructed in many other (but not in official statistics) information retrieval systems for multi-linguistic hypertexts. From technological point of view the parainformation platform is the data warehouse storing weak - structured information describing in harmonized form the objects of NOSIS’ and

related information systems and resources. Basic metainformation and parainformation objects stored on the platform are following:

- Statistical offices (institute, office, regional and local units) and its organizational structure
- Metadata bases: classifications, nomenclatures, code lists, registers, frames, glossaries)
- Statistical surveys
- Administrative records
- Administrative data sources
- Primary records used as statistical data sources
- Statistical microdata bases
- Statistical output data bases and warehouses
- Publications containing official statistical information
- Archived statistical files
- Stakeholders of statistical processes: managers of source records, respondents, intermediaries, users (all types)

The parainformation platform should also store the descriptions of similar objects of statistical systems of ministries and other institutions realizing official statistical processes. It is recommend to store on the statistical parainformation platform not only statistical paradata but also the structured descriptions of objects belonging to other infrastructural information systems of the country, e.g. national information systems of taxes, social insurance, health insurance, registers of population, business registers, territorial registers, registers of infrastructural objects etc.

General model of structured description of objects stored in the platform is the documentation format in library or in scientific information system adapted to the specificity of each type of information system and process. It seems that the list of objects can easily be reduced to limited number of types of objects.

The parainformation platform should be opened for all interested statistical agencies that are ready to share their *parainformational descriptions* of statistical objects listed above in standardized structures and form, and – reciprocally – to get free and full access to equivalent parainformation supplied by other stakeholders of the platform, i.e. the parainformation on other official statistical and administrative systems.

The parainformation platform as the common tool for navigation in numerous national statistical information systems is simple, cheap and effectively supporting the statistical information retrieval in heterogeneous international environment. The parainformation requirements do not interfere in existing laws, procedures and structures of statistical systems of participating countries and statistical offices. Each national statistical agency may take the decision of the scope of parainformation that is willing to deliver to the platform for dissemination and interchange both on national and international level. The parainformation platform is adjusted to the specificity and constrains of heterogeneous international environment.

## 10. Conclusions

Political, social and economic cooperation between developed market economies, post-transition and transition countries of Eastern Partnership needs high level of information transparency, coherence and wide interchange of data between countries, governments, businesses, social organizations and researchers.

The mission of official statistical agencies of Eastern Partnership countries is the creating of legal, organizational, technological and informational conditions and tools of interchange of official statistical information of good quality, obeying methodological standards, accessible in simple and harmonized way by all stakeholders of statistical processes.

*Common parainformation platform* seems to be most effective way to achieve high level of information transparency between countries cooperating within the frames of Eastern Partnership.

The implementation of statistical parainformation platform for transborder regions shall benefit to more dynamic and complex cooperation on political, economic and social spheres in transborder regions of all countries.

## REFERENCES

ISI Declaration on Statistical Ethics, Amsterdam 1985.

UN Fundamental Principles of Official Statistics, Geneva 1992.

OLEŃSKI J., *The citizens` right to be provided with information and the duties of democratic state in modern IT environment in a democratic state in the light of the UN fundamental principles of official statistics and the ISI declaration on statistical ethics*, ISI Special Topic Session, ISI Committee on Statistical Ethics, Proceedings of 53<sup>rd</sup> Session of the Statistical Institute, Seoul 2001.

OLEŃSKI J., *The framework of standards in social statistics - with special reference to transition countries in the ECE region - non-candidates to the European Union*, Economic Commission for Europe, Statistics Division. Invited paper. Geneva, June 2001.

OLEŃSKI J., *Information crises in new economy*, Proceedings of the International Interdisciplinary Conference Crises and Sustainability, Warsaw, 9 – 10 October 2004.

STATISTICS IN TRANSITION-new series, December 2011  
Vol. 12, No. 3, pp. 511–518

## QUALITY - KEY ELEMENT OF DEVELOPMENT OF OFFICIAL STATISTICS

Lucia Spoiala<sup>1</sup>

### ABSTRACT

The assessment of the activity of National Bureau of Statistics (NBS) regarding compliance with the European Statistics Code of Practice reveals that NBS observes to a large extent the examined CoP principles.

At the same time, there are problems related to the implementation of a performing system of quality management.

One of the key problems is that of adequacy of resources, especially at human ones.

The improvement of the structure and of the functionality of NBS, the larger access to administrative data, the implementation of the new IT concept, focused on re-designing and modernization of the statistical data production and dissemination architecture will contribute to the increase of the efficiency of statistical activity and the alleviation of resources constraints.

Quality has become a central issue in the European Statistical System (ESS), but also a key aspect in the integration of the national statistical systems into ESS. A systematic approach to quality was adopted in many institutions. This is based on some common and basic principles of quality management.

At the European level, an important step to this respect was the adoption of the European Code of Practice (CoP) by the Statistical Program Committee on 24 February 2005. Code of Practice, ensuring the coverage of such aspects as the institutional environment, statistical processes and of statistical output within the European Statistical System has an ambitious approach of quality in statistics. The Code describes the common European principles of quality and is based on the experience obtained in the previous period. With the adoption of the Code of Practice, the EU Member States committed themselves to adhere to its principles. For monitoring the compliance of the NSI of Member States with the 15 principles of the Code, a monitoring procedure has been set up. The self-assessments of the countries are combined with elements of peer review, the compliance to CoP being evaluated and monitored based on the explanatory indicators contained by every principle of the Code.

---

<sup>1</sup> General Director, National Bureau of Statistics of the Republic of Moldova.



## **Evaluation of the compliance of the Moldovan official statistics with the selected principles of the Code of Practice**

For evaluating the compliance of the NBS activity with the quality principles stipulated in the Code of Practice, the self-assessment questionnaire had been used. The ESS national statistical institutes and Eurostat had completed such questionnaire, during December 2005-January 2006. Thereafter, this exercise was followed by some external reviews (peer review). The peer-review results had been introduced in the Report on the implementation of CoP, presented by Commission to the European Parliament and Council in 2008. The peer reviews had been limited only to those parts of the Code that refers to the institutional environment (principles 1-6 of CoP) and to the dissemination (principle 15).

The evaluation presented below is also focused on the compliance of the NBS in its activity with these principles. As NBS is the most important producer of official statistics of the Republic of Moldova (the statistical system can be characterised as centralized), the evaluation of the activity of NBS according to the CoP criteria largely characterizes the situation regarding the compliance of the NSS as a whole.

### *Principle 1: Professional independence*

The statistical activity in the country is advantaged by the existence of a good legal basis - the Law on official statistics (hereinafter called "statistical law") adopted by the Parliament through the Law no. 412-XV of December 9, 2004. The non-interference of policy, political, regulatory and other interest groups is clearly stated in Article 5 of the Law. The Central Statistical Body (CSB) is authorized to establish - in an impartial and independent manner, free from any pressure from the Government, political parties, ethnic groups, trade unions and other organisations or natural persons - systems of indicators, nomenclatures, classifications, methodologies, data sources, registration and processing techniques, to disseminate statistical information. In practice, this mandate is fully observed by the CSB.

As regards the mandate of SCB to decide on the statistical methods, standards and procedures, this is clearly stipulated in the statistical law as well. The statistical authority is under obligation to make available official statistics to all categories of users, in terms of equality, simultaneity and non-discrimination regarding the quality and timing of statistical releases, which are strictly observed.

All the statistical activities of NBS are part of the Annual Program of Statistical Works (PSW), performed independently and transparently. These responsibilities are observed in a professional and impartial manner, free from the influence of a third party.

### Principle 2: Mandate for data collection

The mandate for data collection by NBS is stipulated in the statistical law, in the Annual Program of Statistical Works and is sufficient for the performing of statistical surveys. All statistical surveys included in PSW are compulsory for the respondents. Sanctions for non-compliance by respondents are provided in the law, but they are only used if other measures are not efficient to obtain primary data from the reporting units.

The statistical law (art.19) allows the access of NBS to the administrative data sources (ADS), necessary for production of official statistics. These activities are also included in PSW. When the access to ADS is performed, it is based on protocols signed by NBS and institution-source, reciprocally negotiated. Sometimes, the data transmission is limited to aggregated data.

The potential of using the ADS are not fully explored yet. In some cases, especially when certain provisions from other laws contradict Article 19 of statistical law, the access of NBS to administrative data is limited.

### Principle 3' Adequacy of resources

The personnel of NBS consist of professionals with extensive experience in statistical activity, high responsibility, creativity and initiative regarding the development of methods, processes and statistical products. However, the high turnover of the staff, especially among highly skilled statisticians and well trained young people, is a problem that decreases the capacity of the Bureau to meet the ever growing information requirements at the internal and international level (especially in the perspective of European integration). One of the main reasons for such a high staff turnover is the relatively low level of salaries. Another reason is the unattractiveness of the civil servants' promotion system for highly qualified people.

It is crucial the need to rationalise NBS's Information Technologies (IT) applications and adopt solutions which have wider usage. This issue is addressed in the IT development Concept, approved recently by Government, whose implementation will provide the opportunity to re-examine and modernize the information flow within the NSS.

### Principle 4: Quality commitment

Assuring the high quality of statistical data, including micro-data constitutes a permanent concern of statisticians at all stages of statistical activity. At present, timeliness and punctuality are basic criteria in the management of statistical processes.

From institutional and organisational point of view, the existence of two consultative bodies, Statistical Council (SC) and NBS Board represents a guarantee for promoting surveys coherent with users' needs (SC) and adopting appropriate methodologies (Board).

External review of NBS statistical products is practiced, being very beneficial for SSN development. IMF, the ROSC framework did the review in several statistical fields, related to indicators that are part of SDDS in 2005. In 2007, global evaluation of Moldovan national statistical system according to EU standards was carried out.

Although the reference to quality issues is regularly done in the documentation on current activity and on future strategy of NBS, quality commitment is not yet fully integrated in the management and implementation of statistical products and processes. Thus, quality monitoring does not belong to an integrated system of quality management. The lack of skilled human resources in the field and their high mobility, as well as inadequate level of IT development represent critical factors for integration and standardization of the processes among the production units and to develop a quality management system within the organisation in general.

From the viewpoint of legal framework and implementation of legal stipulations in the NBS activity, this principle is strictly respected. Chapter 4 of statistical law regulates data confidentiality, addressing the interdiction of disclosure aspect of confidentiality, as well as the strict limitation of confidential data to statistical use. There are strict procedures applied for external users that access statistical micro-data for research purposes.

All staff, including temporary personnel, has to sign a Confidentiality Undertaking. The obligation to keep data confidential is valid beyond the termination of the contract.

#### Principle 6: Impartiality and objectivity

NBS commitment related to scientific independence, objectivity, transparency and impartiality is based on the adequate legal framework and executed de-facto.

Objectivity and neutrality are key characteristics of all statistical publications and press releases. The sources and production methods of official statistics chosen by NBS are solely driven by statistical considerations. There are two important statutory structures, which guarantee that statistical considerations prevail over any other reasons: two high level professional advisory bodies: (i) Statistical Council and (ii) NBS Board, the last one providing a *permanent* methodological support *and* statistical expertise.

NBS treats all users in an absolute impartial way. All users have equal and simultaneous access to statistical releases and publications.

Principle 15: Accessibility and clarity

According to legal stipulations, NSO should produce and disseminate objective and reliable statistical information and provide all users with equal access to statistical data. Dissemination is a key policy for the institution. Taking into account that NBS is a public institution and official statistical data is „a public good“, all the information, including statistical publications, is put at disposal of users by means of NBS site, free of charge.

Generally, users of various categories are largely satisfied by the accessibility to information provided by the NBS, information considered to be clear and easily understandable. The results of user satisfaction survey carried out in 2008 shown that for 54 % of respondents, statistical data is easily understandable and for 11 %, even very easy to understand.

Information on methodologies and data sources of statistical products is widely documented and it is exposed on the site. Taking into account the results of user satisfaction survey carried out in 2008, most of respondents (77%) consider that methodological explanations provided by NBS are sufficient and adequate. Description of metadata is standardized according to Euro SDMX Metadata Structure (ESMS), applied at EU level.

**Comparative analysis of assessment results**

Self-assessment of compliance of Moldovan statistics (based on NBS example) with principles stipulated in the Code of Practices, based on the analysis and evaluation of respective indicators according to CoP, in comparison with the state of art in ESS states, points out the following:

**Table 1.** Assessment of compliance of ESS (NSIs) and Moldovan statistics (NBS) with CoP

CP principles	Assesment results of CoP indicators, %			
	Fully met	Largely met	Partly met	Not met
<b>ESS (NSIs) [3]:</b>				
1. Professional independence	76	18	5	1
2. Mandate for data collection	83	14	3	0
3. Adequate resources	19	43	38	0
4. Quality commitment	20	37	41	2
5. Statistical confidentiality	86	11	3	0
6. Impartiality and objectivity	71	24	5	0
15. Accessibility and clarity	45	41	14	0
<b>Total</b>	<b>62</b>	<b>25</b>	<b>12</b>	<b>1</b>

**Table 1.** Assessment of compliance of ESS (NSIs) and Moldovan statistics (NBS) with CoP (cont.)

CP principles	Assesment results of CoP indicators, %			
	Fully met	Largely met	Partly met	Not met
<b>Republic of Moldova (NBS):</b>				
1. Professional independence	57	43	0	0
2. Mandate for data collection	67	33	0	0
3. Adequate resources	0	0	100	0
4. Quality commitment	0	40	60	0
5. Statistical confidentiality	80	20	0	0
6. Impartiality and objectivity	71	29	0	0
15. Accessibility and clarity	33	50	17	0
<b>Total</b>	<b>51</b>	<b>34</b>	<b>15</b>	<b>0</b>

In the Republic of Moldova, the situation is almost similar to the one existing in most of the countries within ESS in terms of legal provisions, commitments and practices of the national statistical institute regarding: statistical confidentiality, impartiality and objectivity, the mandate for data collection, professional independence. According to the results of self-assessment, based on the use of quality indicators, the CoP principles are "fully" met in the Republic of Moldova for 80-57 % of the indicators and "largely met" for 20-43 % of the indicators (in comparison with the average level for ESS countries, respectively 86 - 71 % and 11-24 %). Despite the fact that ensuring transparency, accessibility and clarity is a very important dimension of the NBS activity, supplementary efforts are still necessary in view to improve the situation in this respect, in order to achieve the European states level.

At the same time, there are great problems related to the implementation of quality management and resources adequacy. In the Republic of Moldova, these CoP principles are only "partly" met. Although the situation related to the compliance with the mentioned principles is much better at European level than in Moldova, the improvement of resources and of quality management constitute major challenge for many states from the ESS as well.

It is to be mentioned the analysis limits, since the evaluation of NBS activity was done based on the assessment carried out by author in October 2010, while the data related to the situation existing in the ESS are based on the results of external assessments, carried out during 2006-2007, presented in the Commission Report of 2008 [3].

## **Improvement activities**

The analysis of the results of self-assessment on the level of compliance of NBS with the principles stipulated in the Code of Practice pointed some major gaps. Below some activities for further improvement of the compliance with the CoP are mentioned.

### *Principle 1: Professional independence*

Thorough analysis of the use of official statistical data, in view to identify and to speed up the reactions to erroneous interpretation and misuse of statistics.

### *Principle 2: Mandate for data collection .*

Carrying out the NBS' assessment of administrative data sources, aiming at their possible use for official statistics purposes and further collaboration with the owners of ADS, in view to get the access to additional data that are relevant for statistics.

### *Principle 3: Adequacy of resources*

Improving the structure and the functionality of units within the NBS system.

Collaborating with decision-makers in view to set up the NBS staff earnings, at the same level with other central administration authorities.

Implementing the new IT concept, envisaging the redesign of statistics production and dissemination architecture, etc.

### *Principle 4: Quality commitment*

Drawing up a strategic plan in view to implement the quality management system in accordance with the European best practices.

### *Principle 6: Impartiality and objectivity*

Regularly organising press conferences to communicate factual results and increase the number of products with impartial analytical results and series with analytical components (such as seasonally adjusted time series, etc.).

### *Principle 15: Accessibility and clarity*

Increasing the number of series in the data bank. Extending the meta-data described in the standard approved by BNS.

Promoting the statistical culture among the users.

Regularly carrying out user satisfaction surveys.

## **Conclusions**

The assessment of the activity of National Bureau of Statistics (NBS) regarding compliance with the European Statistics Code of Practice reveals that NBS observes to a large extent the examined CoP principles. In the Republic of Moldova, the situation is almost similar to the one existing in most of the countries within ESS in terms of legal provisions, commitments and practices of the national statistical institute regarding: statistical confidentiality, impartiality and objectivity, the mandate for data collection, professional independence.

At the same time, there are problems related to the implementation of a performing system of quality management. The need for a systematic approach of quality measuring and reporting, based on the tools recommended according to the European best practices is obvious.

One of the key problems is that of adequacy of resources, especially at human ones. The improvement of the structure and of the functionality of NBS, the larger access to administrative data, the implementation of the new IT concept, focused on re-designing and modernisation of the statistical data production and dissemination architecture will contribute to the increase of the efficiency of statistical activity and the alleviation of resources constraints.

## **EXPERIENCE AND EXPECTATIONS OF THE EUROPEAN NEIGHBOURHOOD COUNTRIES: THE EXAMPLE OF UKRAINE**

**Oleksandr Osaulenko<sup>1</sup>**

### **ABSTRACT**

Main Components of the National Statistical System Transformation Process consists of: regulatory support, organization of activity (planning and coordination), methodological work, technical and tooling support, staffing.

We can distinguish several priorities of further development of the National Statistical System

- Further improvement of the regulatory foundations of statistical activity towards bringing them in line with the European Statistical Compendium.
- Implementation of a process approach in the statistical activity organization and improvement of the state statistical bodies structure and management on this basis.
- More complete satisfaction of the current needs and identification of future priority expectations of data users as a basis for prospective and current planning of state statistical observations.
- Improvement of the statistical data quality, implementation of the quality measurement, using criteria of the European statistical system, and preparation of the quality reports for users.
- Reduction of the reporting burden on respondents by expanding the use of administrative data, introducing sample observation methods, integrating state statistical reporting forms and promoting the electronic reporting methods.
- Development of statistical infrastructure, use of the methodology and implementation procedures that are, to the highest extent possible, compatible with the European and other international standards.
- Introduction of the metadata driven integrated statistical data processing system.
- Preparation of the national Code of Ethics for employees of state statistics bodies.
- Implementation of measures of the Concept of Confidential Statistical Information Protection.
- Development and maintenance of knowledge and competence of employees of statistical bodies, including, besides Russian, fluency in English.

---

<sup>1</sup> State Statistics Service of Ukraine.



From the outset of an independent state one of the priority objectives of Ukraine's foreign policy was integration into the European Union as a common economic, cultural and information environment. The consistency of the European choice evokes the need to develop an effectively functioning national statistical system that is able both to satisfy all domestic needs of the country in statistical information and meet the requirements of the European community.

The statistical system of Ukraine (slide 1), as in most countries, consists of the main national statistical service and several other organizations that are given adequate powers to collect, compile and disseminate statistical data. For historical reasons Ukraine has a centralized statistical system, so a major place in the national statistical system is given to the State Statistics Service (hereinafter – SSSU) that is composed of the central office, the territorial state statistical bodies (27 regional, and 536 rayon (district) and city bodies), working at the regional level, and functional state statistical bodies that comprise research institution “Scientific and Technical Complex for Statistical Research” (the core task of which is to develop the methodological foundations of statistical observations), educational establishment, the National Academy of Statistics, Accounting and Audit (provides training and retraining of specialists for statistical bodies), and state enterprise “Information and Analytical Agency” (satisfies user needs in information and analytical services on a fee basis).

According to the Regulation on the SSSU approved by the President of Ukraine, the SSSU is an independent state body of the executive power the activity of which is directed and coordinated by the Cabinet of Ministers of Ukraine through the offices of the First Deputy Prime Minister of Ukraine - Minister of Economic Development and Trade of Ukraine.

Besides the SSSU, the statistical activities are carried out by the National Bank, that in charge of such statistical areas as monetary statistics, banking statistics and balance of payments statistics, by the Ministry of Finance, that collects and analyzes statistical data on government finance, and the State Customs Service, that exercises the function of compiling customs statistics.

Effective functioning of the national statistical system is not possible without a special well-balanced regulatory statistical framework (slide 3).

With this in mind, over a number of years the SSSU carries out a task-specific work to create national statistical legislation that would ensure the state statistical activity in the new market economy conditions and at the same time would be adapted to the widely recognized international, principally European, legal regulations. This work resulted in the adoption by the Parliament of Ukraine of the new Law of Ukraine “On State Statistics”, “On Population Census” and “On Amendments to the Code of Ukraine on Administrative Offences” in 2000, the Law of Ukraine “On Amendments to Legislative Acts of Ukraine in Connection with the Adoption of the Law of Ukraine “On State Statistics”, in 2002, the Law of Ukraine “On Agricultural Census”, in 2008. In 2009 the Law of Ukraine “On State Statistics” was emended, in particular, to settle the relations of the statistical

bodies with respondents, co-producers of statistical information, administrative data providers, and data users, as well as the procedure for the statistical bodies activity financing.

Moreover, in 1999, the Law of Ukraine "On National Bank of Ukraine", and in 2002, the revised Customs Code of Ukraine, that regulate the statistical activities in the relevant areas were adopted.

As a result of the conducted work an entire system of legislative statistical acts has almost been formed in Ukraine. They have determined the structure of the national statistical system, tasks and powers of the central statistical service, regulated the relations with respondents, statistical organizations and data users, established the obligation of respondents to provide the primary data, and ensured the protection of confidential information, etc.

It should be noted that the created national statistical legal framework is fully harmonized with international and European regulations, and meets their standards.

As part of the systematic work to adapt the regulatory statistical framework with the EU requirements over the years of 2008-2010 the SSSU has developed national principles governing the activity of state statistical bodies.

It stands to mention that the successful preparation of the national principles have been preceded by a range of effective activities that formed a reliable basis for this work. In particular, in the period of 2008-2009 the SSSU conducted a functional survey of state statistical bodies, carried out the self-assessment against the principles of the European Statistics Code of Practice on the national questionnaire that is based on CoP Questionnaire, and developed the quality management system for state statistical bodies, etc.

The Principles of Activity are designed to promote the introduction of the best European statistics experience into the national statistical practice, to ensure a high-quality statistical information and on this basis to increase the user confidence in the statistical bodies activity as well as placing of Ukrainian statistics to the European statistical system.

The National Principles of Activity include all 15 principles of the European Code of Practice, but at the same time additionally comprise three following principles: "management and professional ethics", "effectiveness of activity", "on-going improvements". Besides, principle 9 "non-excessive burden on respondents" of the European Statistics Code of Practice was supplemented by a number of indicators that characterize the cooperation not only with the respondents, but with all other entities involved in the statistical activities. This is reflected in the name of the principle "partnership relations in the state statistical activity" of the National Principles of Activity. The range of indicators in principle 10 of the European Statistics Code, "cost effectiveness", was also extended. The amendments mentioned above were caused by the necessity of creating an appropriate environment and a precise mechanism to ensure continuous improvement and development of statistical bodies based on the

introduction of approaches defined in TQM, EFQM, ISO 9001:2000 into statistical activities.

With regard to the organization of statistical activities, the key requirement for any statistical system, especially with a high degree of centralization, is the availability of effective procedures for planning and coordination (slide 4). Effective planning is needed to define the strategy and coordinate the objectives, to ensure the system sensitivity to user requirements, to mobilize the financial and other resources, to maintain a favourable external environment, to assure consistency in actions, etc.

In order to identify perspective targets and strategic directions for development of national statistics, ensure continuous improvement of the quality of statistical information the SSSU jointly with the National Bank of Ukraine and other compilers of statistical and administrative data elaborates the long-term program for statistics development by the results of the user demand observed and the European experience.

This program is usually prepared for five years. Currently, we implement the fourth long-term program, the Strategy of State Statistics Development for the Period till 2012.

To provide state and society with the required statistical information the SSSU with all statistical and administrative data compilers annually develops a plan of state statistical observations.

The long-term program for statistical development and the plan of state statistical observations are approved by the Cabinet of Ministers of Ukraine.

In 2010 the new approaches to the current planning of state statistical activity were introduced. Thus, when forming the draft plan of statistical surveys for 2011 we significantly changed its structure and information content. The draft plan is formed by the results of the statistical observation descriptions made with the purpose to create a system of feedback with respondents and develop a statistical metadata base. With that, the plan has the listing of statistical surveys that are to be conducted by state statistical bodies, and the descriptions define the observation objective, its type, organizational form and working method, as well as specify the relevant statistical tools.

One of the main and most difficult aspect of the activity for state statistical bodies is to coordinate the activities of co-producers of statistical information and providers of administrative data within the national statistical system.

To solve the problem the Law of Ukraine "On State Statistics" stipulates that the relations of state statistics bodies with the users and respondents, including those that carry out the activity related to the collection and use of administrative data, are based on the principles of ensuring the reliability, impartiality and confidentiality of statistical information, the effective use of costs for its preparation and transparency of statistical methodology that envisage:

- approval by state statistical bodies of methodologies and reporting documents associated with collection and use of administrative data as well as

methodology for compilation of banking and financial statistics, balance of payments statistics etc.

- free-of-charge provision to state statistical bodies at their request with administrative data collected as well as data of banking and financial statistics, balance of payments statistics etc.

The SSSU carried out the extensive work to adapt statistical methodology to European requirements and standards (slide 5). For this purpose, principally, the SSSU developed and implemented the classifications harmonized with the relevant international and European counterparts, namely Classification of Types of Economic Activities (harmonized with NACE), Classification of Products by Economic Activity (in accordance with the CPA), Nomenclature of Collecting Production Data (PRODCOM), Classification of Individual Consumption by Purpose (COICOP), Combined Goods Nomenclature (CN), Classification of Foreign Economic Activity Commodities (HS and CN), etc.

Currently, a new revision of the Classification of Economic Activities (KVED 2010) based on NACE (rev.2) is approved, the SSSU carries out the works to introduce it into the state statistical activity. We are finalizing the draft Statistical Classification of Products (SKP 2011) developed on the basis of CPA-2008.

The Regulation on the Register of Statistical Units and forming the sampling frame of statistical observation units, that will ensure the wide use of sample survey methods in the statistical practice, is endorsed.

On the basis of the European standards the SSSU has developed and constantly improves the methodology for basic statistical observations, for example, recently the SSSU has prepared the methodological provisions for compiling satellite accounts on education and social protection, economic accounts of agriculture, for structural survey of non-financial sector enterprises, for compilation of the industrial production and turnover indices, for the formation of energy balance, etc.

In order to ensure the legislative provisions on confidentiality the SSSU developed the Concept of Protection of Confidential Statistical Information. The implementation of the Concept will guarantee the confidentiality of statistical information according to the European standards, enhance the confidence of respondents in the state statistical bodies activity, and facilitate user access to statistical data that will positively affect the quality of statistical information as a whole.

An essential prerequisite for an effective statistical system formation is the upgrading of information and communication technologies (slide 6). For this purpose within recent years the SSSU has carried out a number of large-scale activities:

- the state statistical bodies, primarily, the territorial ones, are reequipped with modern computers and communication facilities;

- an uniform centralized IT infrastructure based on Microsoft Windows Active Directory is created that allowed to obtain an information system of state statistical bodies of qualitatively new level;
- a set of works to provide respondents with the broad options to electronically submit the primary statistical information to the state statistics bodies is executed;
- the Sybase and SPSS software for processing and analyzing statistical data are introduced;
- currently, the work on development of a metadata driven integrated system of statistical data processing in the state statistical bodies has almost been finalized by the SSSU.

An effective and comprehensive cooperation with international organizations, including the World Bank, Eurostat, UNECE, EFTA, ILO and others, aimed at improving the national statistics, increasing its efficiency and, ultimately, the quality of statistical information (Slide 7) plays a key role in the development of the Ukrainian statistics.

Thus, in recent years, under the World Bank support the SSSU jointly with the Ministry of Economic Development and Trade, Ministry of Finance and National Bank of Ukraine has successfully implemented loan project "Development of the State Statistics System for Monitoring Social and Economic Transformations", the main objective of which was the development of the sustainable state statistical system of Ukraine through its comprehensive and systematic reforming to better meet the needs of users in objective statistical data;

Within the framework of cooperation with the European Union the TACIS-8 and TACIS-10 programs had great importance. In addition, in the years of 2009-2010 in order to prepare for the implementation of the Statistical Requirements Compendium that is annexed to the EU-Ukraine Association Agreement, the SSSU effectively accomplished the Twinning Light project "Harmonizing the Ukraine's official statistics with the EU statistical standards" resulting in:

- the action plan to achieve harmonization of national statistics with the EU standards in priority areas developed;
- the system of indicators that will allow to assess the approximation of the national statistics to the EU standards developed;
- tools and methods for the Ukraine's statistical bodies self-assessment against the principles of the European Statistics Code of Practice to be conducted on a regular basis defined;
- the tools developed and test calculations by the full set of accounts of SNA 1993 international standard conducted.

As for now, the contract for implementation of Twinning project "Development of new statistical methodologies and indicators in specific areas of statistics in line with EU statistical standards" that is expected to start in November has been signed. Our partner for this project, as in previous one, is Statistical Office of Denmark.

In September this year the Global Assessment of the national statistical system of Ukraine was launched. According to the preliminary results the EU expert mission expressed positive opinions regarding the status and development trends of the Ukrainian official statistics.

An important event for us became the proposal of international organizations to hold the prestigious event, the Conference of the International Association of Official Statistics 2012 in Kyiv. We take it as the recognition of the achievements of the Ukrainian statistics in harmonizing with the European requirements.

Finally, a few words about the prospective tasks facing us in terms of further transformation of the national statistical system (slide 8):

1. Further improvement of the regulatory foundations of statistical activity towards bringing them in line with the European Statistical Compendium.
2. Implementation of a process approach in the statistical activity organization and improvement of the state statistical bodies structure and management on this basis.
3. More complete satisfaction of the current needs and identification of future priority expectations of data users as a basis for prospective and current planning of state statistical observations.
4. Improvement of the quality of statistical data, implementation of the quality measurement, using criteria of the European statistical system, and preparation of the quality reports for users.
5. Reduction of the reporting burden on respondents by expanding the use of administrative data, introducing sample observation methods, integrating state statistical reporting forms and promoting the electronic reporting methods.
6. Development of statistical infrastructure, use of the methodology and implementation procedures that are, to the highest extent possible, compatible with the European and other international standards.
7. Introduction of the metadata driven integrated statistical data processing system.
8. Preparation of the national Code of Ethics for employees of state statistics bodies.

9. Implementation of measures of the Concept of Confidential Statistical Information Protection.
10. Development and maintenance of knowledge and competence of employees of statistical bodies, including, besides Russian, fluency in English.

*STATISTICS IN TRANSITION*-new series, December 2011  
Vol. 12, No. 3, pp. 527–536

## **TRANSFORMATION OF ESTONIAN STATISTICAL SYSTEM TO THE EUROPEAN STATISTICAL SYSTEM – MAIN ASPECTS AND EXPERIENCES\***

**Priit Potisepp, Kaja Sõstra, Allan Randlepp<sup>1</sup>**

### **ABSTRACT**

A lot of similar features can be observed in the development of the Eastern and Western European statistical systems during the past 20 years, but there are also a lot of differences. While after the collapse of the Communist Block the CIS countries started to co-ordinate statistical activities, the other group of countries followed the goal of acceding to the European Union. For this reason the development of statistical systems in these countries was almost completely focused on taking into use the shaped and continuously developing practice of the European Statistical System. The transition experience of these countries was of great value. For its small size and compactness, Estonia can be looked upon as a laboratory where new undertakings can be implemented more easily compared to large systems. Transition experiences can be observed from different aspects. A new legal-institutional framework had to be established, sample-based survey methodologies and new classifications had to be put into use, collection of entirely new type of data (e.g. the intra-Community foreign trade statistics) had to be prepared, and statistical registers had to be established. As the resources of transition society are scarce, the organisation and processes had to be organised as effectively as possible in order to cope with a large-scope statistical programme of the EU. Estonia has the IT-state image but swift implementation of opportunities arising from the IT development in the production of statistics has not proved easy. Nevertheless, during the last decade, the data collection, processing and publication have undergone great development. One of the most important aspects ensuring successful transition is a change in the mentality of statistical organisation. The transition period has proved that encouraging the younger generation specialists to stay at Statistics Estonia has been a real challenge and in several spheres we have managed to fulfil time- and energy-consuming tasks due to the great devotion of our specialists who are known for their long-term experience.

---

<sup>1</sup> Statistics Estonia.

\* Acknowledgement: English Editing by Heli Taaraste.



## **The current system of official statistics in Estonia – brief introductory overview**

Estonia as a small country has a centralised official statistical system. According to the law, official statistics are produced by Statistics Estonia (SE) and the Bank of Estonia. In addition, a remarkable quantity of administrative statistics is produced by a large number of state agencies. The official statistical programme is approved by the Government of the Republic and it is in firm correlation with the European Union Statistical Programme. It can be said that, at the beginning of the 90ies, Estonia anew built up the official statistical system, which had come to an end in 1940, and step by step, started to pursue the European Statistical System (ESS) membership status from the mid-decade and officially joined the system in 2004. The milestones reached during the two decades include laying a European-type foundation for official statistics, implementation of the principles applied by the statistical systems of the today's developed world, introduction of a European-type work programme and ongoing search and testing of modern technology-based opportunities. The use of statistical data in the public information space is clearly revealed in Estonia. Fact-based analysis and discussions are conducted on an ever wider scale while developing and monitoring policies and strategies. SE has been making purposeful and systematic efforts in this direction. For statisticians, a statistical action implies above all the use of data and, through it, benefits gained by the whole society.

SE operates in the area of the Ministry of Finance, but is professionally independent. Statistical Council consisting of representatives of academic world and other users of statistics and respondents provides advice to the statistical system.

### **I. Development of legal-institutional framework**

1) In Estonia, a statistics act typical of the transition time was passed in 1990. Its regulatory effect was weak and in a few years there emerged extensive controversies with other new laws and also with the Constitution. The next statistics act was passed in 1997 and it was, in general lines in the same wording, in effect until 1 August 2010 when the current statistics act was adopted.

2) One of the most powerful stimuli encouraging drafting of the new act was the Peer Review on the Implementation of the European Statistics Code of Practice at the end of 2006. The results thereof elicited a certain need for better regulation in order to reinforce the professional independence of SE. A need to revise the Agricultural Census Act and Population and Housing Census Act which had been functioning separately so far was also a topical impetus. The three above-referred acts were later combined into one uniform act.

3) In 2009, Regulation (EC) No 223/2009 of the European Parliament and of the Council on European statistics was adopted. This coincided with preparations for the new Official Statistics Act in Estonia and, due to that in this national legislative act a strong and visible correlation was achieved with the respective EU act. This was done, first and foremost, in order to emphasize validity of the main principles underlying official statistics, although duplication of regulatory matters laid down in an EU regulation is not common practice.

4) The Statistical Council set up pursuant to the new Official Statistics Act gave additional solidity to the legal-institutional structure of official statistics. Despite the fact that there was a statistical council in Estonia already in the 20ies and 30ies, it can be stated that the Statistical Council should have been founded earlier in the present-day Estonia.

## **II. Infrastructure of official statistics – IT, classifications, statistical registers and methodology**

### **A. Metadata driven development of statistical IT systems**

5) Implementation of a coherent corporate statistical system requires a systematic approach to software development, project management and testing. Over the last decade, Estonia has considerably increased outsourcing in the field of software development, but relied on in-house resources in the field of project management and system analysis. Since the beginning of financial crisis in 2008 which led to sharp budget cuts, the European Union Structural Funds have had an important role in the development of statistical information systems.

6) The possibility of outsourcing IT services in the statistical system has become a recurrent theme on the international forum of IT specialists. For sure, it should be explicitly defined what is to be outsourced and what remains the responsibility of a statistical institution's IT department. The situation of ESS members is different, but still a growing trend of outsourcing can be noticed. In Estonia, limited knowhow and the IT sector's high salary level constitute a major obstacle to supplying a statistical institution's IT unit with a full-functional in-house staff. Therefore, focus has been set on cross-agency cooperation networks. Traditional IT support services and IT infrastructure have been consolidated for all institutions under the government of the Ministry of Finance.

7) Within the international statistical system, efforts have been made to share the statistics production software. At the UNECE level, the Sharing Advisory Board has been set up and Eurostat has carried out two ESSnet projects for standardising the IT system architecture (CORA, CORE). Whenever possible, SE has used software developed by others (for example PC-Axis and Blaise) and is willing to share any in-house developed software.

8) In the statistical system, at the national as well as ESS level, metadata which function as integrating elements in the statistical information systems are of central importance. Development of metadata driven information systems has been the keyword of the last decade. SE has been developing the metasystem for over ten years and, in mid-2011, the new central metasystem called iMeta was put into operation. The previous metasystem developed in 2000–2003 is gradually being replaced with the new one. In iMeta, the basic code is open and the software can be used on the basis of EUPL licence. The documentation and software are in English.

9) Designing of technical metadata application is not the only important task, standardisation of metadata embracing the whole organisation is essential, too. SE started to implement the latter task from describing statistical actions, from classifications and output tables; currently the work with statistical questionnaires and definitions is under way.

## **B. Statistical registers**

10) In Estonia the register of enterprises, agencies and organisations was set up in 1990 and SE used this register for 2 years in producing economic statistics. Since statistical needs are wider than those of the register of enterprises, agencies and organisations, thus in 1994 SE decided to establish the Business Register for Statistical Purposes. The aim thereof was to set up a register which complies with the EU legislation and contains as up-to-date information as possible. A separate farms register was created in 2002 in Estonia.

11) At the beginning, the Business Register for Statistical Purposes was based on the FoxPro platform, later on it was developed further on the basis of Oracle platform. The register which is in use also today has been in operation since June 2002 and contains data on economic units starting from the year 1995.

12) SE participates in an expert group which deals with developing methodology for the ESSnet register of multinational enterprise groups (EuroGroups Register) in collaboration with colleagues from Italy, the Netherlands and the United Kingdom.

13) In 2011, SE launched an extremely large-scale and ambitious project for setting up a system of statistical registers. The purpose is to essentially enhance the functionality of the Business Register for Statistical Purposes, reform the farms register and create two new statistical registers – the population register and the buildings and dwellings register. A need for new registers proceeds from the future objective to conduct the 2021 Population and Housing Census as a register-based Census.

### C. Development of classifications and methodology

14) The use of international classifications in statistics became especially important in connection with joining the ESS. SE managed to fully bring in international classifications by the end of 1990ies and, a general assessment of the Estonian statistical system carried out at the end of 1999 confirmed its full compliance with requirements.

15) Administration of national classifications is a legal responsibility of SE. Therefore, all amendments that state agencies wish to make to their databases must be first coordinated with SE.

16) Major classification-related projects which started after accession to the EU include introduction of the Classification of Economic Activities (NACE vol 2.0) and of the International Standard Classification of Occupations (ISCO 2008) at SE. Classifications belong to the meta-information system, where respondents and users of statistics can conveniently use and download classifications and correspondence tables as well as search for necessary additional information.

17) Transition to the market economy statistical system entailed fundamental changes in methodology. Step by step, surveys based on probability sampling methods were introduced. The first sample-based household survey was organised in cooperation with the Norwegian Research Foundation (FAFO) in 1994. The ESS regular and main Labour Force Survey (LFS) monitoring the labour market was organised for the first time in 1995 by the Estonian Institute for Population Studies whereas the sample was taken from the Soviet Union last Population Census (1989) database. From 1997 onwards the LFS has been conducted by SE. Implementation of the European Union Statistics on Income and Living Conditions (EU-SILC) was a high objective as the methodology of this survey had been prescribed in an extremely detailed way. Estonia was the only one of the countries having acceded to the EU in 2004 which managed to implement EU-SILC in 2004. Allowing for some concessions, the Household Budget Survey can also be included among the continuous surveys mentioned above. Besides continuous surveys, SE has also conducted several ESS ad-hoc surveys, e.g. the Time Use Survey (1999–2000 and 2009–2010), the Continuing Vocational Training Survey, the Work Life Survey, the Immigrant Population Survey, etc.

18) Introduction of the probability-based Structural Business Surveys started in 1994. Larger enterprises (with at least 20 employees) were observed by total survey and a stratified random sample was drawn from among smaller enterprises. In 2001, the used probability-based sample surveys were further elaborated. Following the example of Statistics Sweden, the sample coordination methodology was taken into use in order to ensure a more even distribution of enterprises' response burden.

19) In parallel with the introduction of probability-based sample surveys, computation and publication of quality indicators of surveys were started. The first indicators were response rates and the standard errors or confidence intervals of estimates. Confidence intervals and the response rate were first published for the 1994 wage statistics. At that time, quality indicators were published in printed

publications, after the creation of statistical database also in the database tables or as attached metadata. In 2010, the publication of quality indicators was thoroughly revised and dates were set for publication of the quality indicators of all sample surveys. During 2011, publication of quality indicators has remarkably increased.

20) A legal requirement to compile and publish detailed quality reports has been established in the ESS with respect to several statistical domains. The first such reports were compiled by SE four years before accession to the ESS in structural business statistics (SBS). Thereby some quality indicators were even calculated by individual variables (e.g. coefficient of variation, item non-response rate, unit non-response rate). At the same time, first reports were prepared also on short-term business statistics.

21) Later on, the structure of quality reports was updated according to the quality aspects (e.g. timeliness and punctuality, etc.) agreed on within the ESS.

22) Improvement of quality in statistics involves an international effort. For example, the statistical institutions (incl. SE) which had joined the ESS in 2004 participated in an international project "Quality in statistics". The existing quality reports were revised, the quality assessment questionnaire was tested on six surveys of different domains and trainings were conducted. This project gave a lot of people systematic knowledge about the quality of statistics.

23) Quality of statistics is influenced not only by the quality of data processing processes. The European Statistics Code of Practice encompasses 15 quality criteria. All members of the ESS commit themselves to adhering to these principles in their activities at the national as well as international level. In order to introduce the Code, the ESS members conducted self-assessments in 2005 and, in 2006 Estonia was among the first countries where also a Peer Review was carried out. SE has been continuously involved in the implementation of EFQM (European Framework of Quality Management). For the given purpose, the first self-assessment took place in 2007, followed by a Peer Review. SE got the EFQM certificate "Committed to Excellence". Certificate of the same level was awarded to SE also in 2009.

#### **D. Development of processes**

24) The way how official statistics had developed in the past, created a situation where different statistical domains developed rather in isolation from one another. That is why, in lots of countries, fragmented pictures can often be seen in data processing. This is also characteristic of SE. Differences in processes have been caused by variability in methodology and technical applications used in data processing. New initiatives have been taken at the international level with a view to standardising statistical data processing, which should open fresh prospective in the ESS for the use of standardised processes and data processing applications.

25) In the last decade, SE has moved towards bigger centralisation: the so-called horizontal units in SE are methodology, IT, data collection and dissemination.

Generalised statistical business process model (concerning metadata and statistical processes) forms a basis for IT architecture and also for software development.

26) Centralisation of the dissemination of statistics on 01.09.1993 can be pointed out as the first case of withdrawal from work organisation by the so-called stove-pipe pattern. Nowadays, there is nothing new in that for statistical institutions. But nearly 20 years after a thorough reorganisation of the Estonian statistical system it should be noted that namely centralised dissemination of data created preconditions for developing a consolidated client perspective and similar services irrespective of the statistical domain. Furthermore – centralised dissemination allowed for an easier start in following the fundamental principles of ESS data dissemination. In this way, it is easier for a centralised dissemination unit to apply the equality of treatment principle with respect to consumers than in decentralised statistical systems or in organisations where the dissemination function is scattered.

27) In addition to acting as a watchdog the principle of equal access to statistical releases, the central dissemination unit functions as a lawyer for users of statistics since he stands for the timeliness and relevance of statistical data, develops modern services on a constant basis and develops client relationships. In the ESS, dissemination heads closely cooperate in order to

- a) ensure smooth publication of official statistics of European countries;
- b) consistently learn the application of best practices from one another and get new product-related ideas.

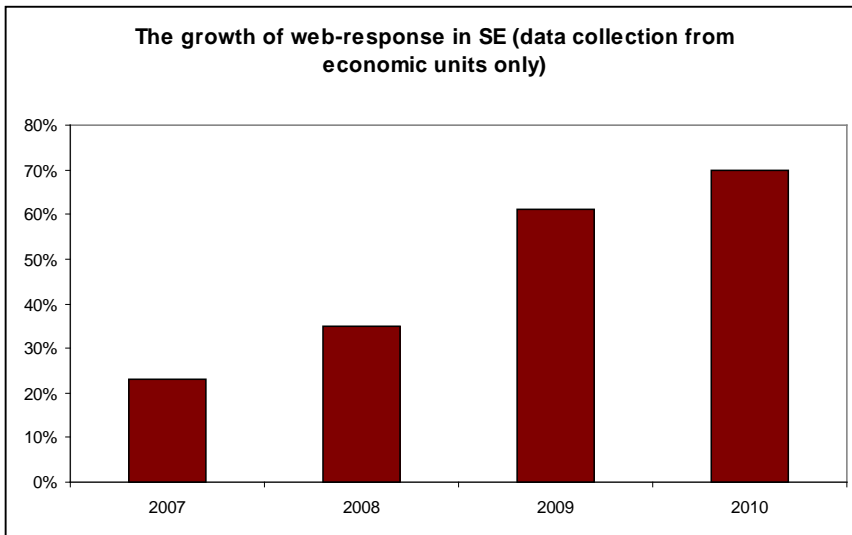
28) The dissemination unit could make use of fast developing IT possibilities in the publication of all statistical data. Application of the PX-Web software while the public database was created can be considered a specific example of eastward collaboration between Estonian and Swedish statistical systems. The ESS strategy for the coming decade attaches a remarkably bigger importance than earlier to statistical information processing systems which are being developed in common interests. Relying on a common wallet, the most capable ones are invited to develop services and processes for the benefit of the whole statistical system. Census-Hub – an environment for disseminating the 2011 Population and Housing Census data – could be highlighted here as an example.

29) Web technology and the Internet have essentially facilitated the use of microdata for scientific communities. Instead of coming in person to SE or taking microdata saved on a CD-ROM along from SE, researchers can download public use files and have even more detailed data at their disposal at their workplace via remote access.

30) Data collection is the sphere where extremely great changes have taken place in the modernisation of processes during the last 7–8 years. Estonia as a small country has to work with relatively large samples in order to publish as detailed data as agreed on in the EU. This fact lays a relatively large response burden on respondents. On the other hand, every Member State is committed to fulfilling Principle 9 of the European Statistics Code of Practice (non-excessive burden of

respondents). Thus, strong development of electronic data collection and an effort to expand the use of administrative databases have been self-evident.

31) Several solutions were considered during the initial years of electronic data collection (file upload, web-questionnaire, file per e-mail, etc.). To get data processing, which follows data collection, standardised and unified, data collection had to be grounded on a single technical solution – a web-questionnaire for economic units and laptop applications for face-to-face interviews. SE has systematically and consistently worked towards preparing respondents for giving up paper-based reports. This work has proved successful but was not an easy task. In 2009, we unilaterally dropped the send-out of blank report forms to respondents who had been included in the sample by informing them that they could use the possibility of downloading the questionnaire form from the SE web site and submit the report on paper, but we strongly advised them to submit data over the web. By now, all monthly, quarterly and annual reports can be submitted over the web. In 2010, 70% of all reports received from economic units were submitted via the web. Nevertheless, achieving of such a result has taken several years.



32) Timeliness of data depends on timely receipt of input data. Here, no higher mark than ‘satisfactory’ can be given to the situation. In short-term statistics, data are received by a fixed date in 50–60% of cases, in annual statistics – one third of cases. This means that the data collection unit (centralised at SE in 2004) should spend lots of resources on a fast delivery of delayed data in order not to detain fulfilment of the time schedule fixed for further processing.

In a short-term perspective, the solution is oriented to using enterprises’ annual reports submitted to the Commercial Register in structural business statistics and, to using monthly reports submitted to the Tax and Customs Board in enterprises’ short-term business and wages statistics.

33) SE has taken interest in developing cooperation possibilities with representatives of several accounting software companies for the purpose of creating solutions for importing statistical data directly from companies accounting systems. The referred attempts have unfortunately proved futile so far. During 2009–2010, main attention has been paid to other aspects: developing of the common taxonomy part of annual accounts of enterprises and structural business statistics, to implementation of compulsory electronic annual accounts and to the importation of data from there for structural business statistics.

34) By now, laptops are made use of in all household surveys. In single cases SE has tested completion of web-based questionnaires. The 2011 Population and Housing Census will be carried out by a combined method encompassing e-census and face-to-face interviews. Several countries in the world (South Korea, Bulgaria, Australia, Latvia and Lithuania) have recently achieved extremely encouraging results in conducting the e-census. The advantages of transition of interviewers' work from paper-based questionnaires to e-questionnaires lie in instant data quality check up and in central monitoring of interviewers' compliance with their work schedule.

35) During last decades, the development of statistical data processing has been most of all influenced by IT development. Production of a data unit has become cheaper namely due to technological development. A sharp increase in the volume of data would not have been possible without IT investments. Statistical institution of every country should take into account the development of e-services in other public sector or private sector institutions and, proceeding from that, the expectations of cooperation partners. Since the end of the past century, submission of tax returns through the e-Tax Board/e-Customs has been pursuing a fast course of victory. In Estonia the relevant e-forms are in general simpler than the statistical questionnaires intended for economic units. But the habit of enterprises of submitting value added tax returns, income and social tax returns via the web has created favourable preconditions for web-based collection of statistical data.

36) Upon joining the ESS, Estonia had to adjust its release calendar to that of the ESS, which meant accurate planning of time schedule for every survey. Considering rather widespread criticism of the timeliness of data, various work stages had to be accelerated and this has been possible only due to elaboration of information technology solutions.

### **III. Effect of ESS-collaboration on the development of statistical institution**

37) In a small country like ours, hiring and keeping of statisticians on the payroll of our institution is a really complicated task due to a small number of statistics specialists graduating from universities. Joining the ESS has had an obvious effect on the development our statistical organisation's capabilities. In 2010, the



public servants of SE spent 905 days at international meetings which totals approximately 4.5 man-years. At least half of this time was spent at the ESS meetings.

38) The ESS director groups as well as working groups and task forces of top specialists provide valuable opportunities for exchange of experience. The ESS is unique in the sense that a statistical action is in detail and predominantly governed by legal acts. Legal acts are compiled in collaboration with the Member States; implementation of new statistical actions is in a lot of cases financed from the EU budget funds, from where the Member States get support. Estonian statistical system became eligible for the ESS member status as a result of several EU transition facilities.

39) The ESS strategic development trends initiate strategic choices at the national statistical institution level. Similarly, the ESS annual and multi-annual work programmes constitute the most important input for the national official statistical programme.

40) The ESS members are committed to the national as well as European needs. In the past, there were several discussions on international forums about the controversies between the statistical needs of Europe and those of Member States (e.g. at the DGINS conference in Cracow in 2006). According to Mr Tim Holt, former national statistician of the UK, who made a presentation in Cracow, there definitely exist European statistics and there cannot exist large divergences between the statistical needs of Member States and the respective needs at the European Union level. During recent years, the ESS with its members has been developing into an ever more harmonious family who is strongly committed to common principles, work plans concerning the whole statistical system and forthcoming strategic tasks.

## **Summary**

In the framework of the Eastern Partnership Program of the European Commission, incl. Eurostat, the experience gained by the Member States which acceded to the EU in 2004 is of great value. Pursuing the ESS member status has not been an easy task, but the results achieved are extremely positive. It has taken the referred countries a decade to build up their statistical systems which are no worse than the respective statistical systems of the so-called old Member States. It is true that sometimes set-up of a new system proves easier and takes less time than reorganisation of a complicated old system. Like in other small countries, Estonian official statistical system has been centralised, the legal-institutional basis has been updated and Estonia is successfully implementing the statistical programme. In development rather the stove-pipe approach with its merits and drawbacks has been followed. The challenges for prospective development are in general similar to those faced by the entire ESS.

*STATISTICS IN TRANSITION*-new series, December 2011  
Vol. 12, No. 3, pp. 537–545

## **DIRECTIONS FOR DEVELOPMENT OF TRANSBORDER AREAS – STATE AND PROSPECTS**

**Marek Cierpiał-Wolan**

### **ABSTRACT**

The processes of integration in Europe cause the need for more and more information concerning transborder areas, thereby growing interest in regional statistics. The reason for this is mainly changes of functions of the borders caused by globalization process. The need for using the results of transborder surveys on the micro-mezo-macroeconomic level gives rise to establishing a consistent research system for these areas. Despite efforts of several international institutions, there still exist problems with lack of information on particular levels of aggregation as well as with data comparability level in individual countries. As a consequence, there is still a need for identification of major research areas and discussion on important methodological aspects relating to transborder areas.

### **Introduction**

Due to globalisation processes the economy becomes a blurred system in many cases, with supranational labour markets as well as goods and services markets being created. On the one hand, we can observe integration processes. The most recognizable one in Europe is the extended Schengen Area. On the other hand, disintegration processes are visible.

In the countries where the Schengen Area is being introduced, open borders facilitate communication, running business, etc., but they generate many information gaps. The removal of border controls and cessation of recording the traffic of people and vehicles cause the information deficit in this regard. This is why there is necessity of replacing the lost source of information through retrieval of data in another way.

What is more important is the fact that opening the borders in some countries usually leads to tightening the rules of crossing the borders in others, in this case at the EU's external border. It is the reason why we can frequently observe higher intensity of non-registered socio-economic phenomena in this area.

Therefore, integration and disintegration processes cause the need for more and more information concerning transborder areas, thereby growing interest in regional statistics.

Regional statistics often comes across different problems of, among other things, 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 and low level of data comparability, especially those pertaining to economic issues.

Therefore it is essential, to create a coherent research system for cross-border areas which would allow for gathering and analyzing comparable data from sample surveys and administrative registers.

In response to a greater demand for information Polish official statistics resumed in 2008 the survey of goods and services turnover in border traffic. Initially, it was carried out on the Polish-Ukrainian border, and since 2010 at the whole European Union's external border on the territory of Poland.

Results of this survey turned out to be unexpected. For example, in the first half of 2011 the expenses incurred on the purchase of goods in Poland by Ukrainians amounted to 277,8 million € whereas by Byelorussians amounted to 173,4 million €, what account for 18,5 % and 23,9 % of the whole Polish export to these countries respectively. Thus, it came out that the scale of unregistered trade is surprisingly high.

## **Towards transborder index**

In order to create the coherent research system for transborder areas, the first think which should be done is delimitation. We have to begin with preliminary delimitation based on, for example, regulations such as Regulation No. 1931/2006 of the European Community, dated 20 December 2006, according to which the border zone covers an area of 30 to 50 km from the border. Another important criterion of delimitation can be results of the questionnaire surveys at the border, journeys to work, etc.

The survey of goods and services turnover in border traffic at the European Union's external border on the territory of Poland, which took into consideration distance from the border to the place of residence of Poles and foreigners, showed that the highest number of foreigners and Poles was making purchase at the distance of up to 50 km from the border – around 71% and 96% respectively, with almost 55% foreigners and about 94% Poles making purchase in the zone of up to 30 km. In terms of the distance from the border to the place of purchase the highest number of foreigners and Poles was making purchase at the distance of up to 50 km from the border – around 71% and 96% respectively, with almost 55% foreigners and about 94% Poles making purchase in the zone of up to 30 km.

Creating a coherent research system we have to remember about dynamic delimitation which means systematic analysis of socio-economic phenomena in

this area and environment (e.g. labour market, entrepreneurship, demographic processes, tourism, differences in the level of prices). For this reason Polish official statistics launched two modules – monitoring of socio-economic phenomena in cross-border areas on the basis of administrative registers and statistical databases as well as sample surveys at the border and in the neighbourhood.

All sorts of information which can be found in statistical databases and administrative registers adjusted to transborder areas fall within the ambit of the monitoring. Most of the findings of our investigations argue that there is higher activity in transborder areas.

The plan of actions which should be taken up, that is the scope of surveys and the schedule of their implementation, was worked out. Firstly, four kinds of surveys should be conducted: questionnaire survey at the border, survey of travelling foreigners in tourist accommodation establishments, household survey (modules concerning international tourism, unregistered work, volume of unregistered purchase), and survey of unregistered economy (market points, retail sales points, etc.), with the use of experience of Gray Economy Research Centre.

Being in possession of great variety of partial indicators based on monitoring and data of surveys on the one hand, and observing ongoing specific changes in transborder areas on the other, a natural consequence seems to be the development of synthetic indicators.

In spatial analysis, where areas of different countries are being surveyed the problem is the choice of the object of a study. We must not forget about the need to compare spatial units which are similar especially in terms of population and area. It is worth stressing that because of these two criteria, the spatial breakdown varies in different countries.

According to the above we should avoid comparing some administrative units in individual countries, although these units fulfil the same administrative functions.

Based on many years of experience in the development of common transborder information system, it should be noted that the best spatial unit is LAU level 1. In many countries administrative units on this level are very comparable in terms of population and area.

### Creation of complex feature

In order to investigate differences in development of territorial units at various levels of aggregation, we can use a synthetic indicator. This index, as it is well known, is constructed based on partial indicators which may come from different fields. They can be used to analyze a specific area of research (e.g. economy, human capital, environment) or the general socio-economic development of a given region.

In the selection of partial indicators both substantial and formal criteria are taken into account. A properly selected set of variables should describe the

phenomenon under study as precisely as possible, and at the same time be not too numerous. The next step is to use various statistical methods including correlation and regression analysis to eliminate the variables that contain the same or very similar information on the investigated phenomenon, or showing little variability. It is important that empirical data should be available for each region and comparable between them.

A set of comparable variables available at NUTS 2 and LAU 1 was prepared for Polish-Slovakian-Ukrainian border area, which includes:

- Population per 1 sq km (NUTS 2)
- Population at age 15-64 in % of total population (NUTS 2)
- Net migration per 1000 population (NUTS 2, LAU 1)
- Infant deaths per 1000 live births (NUTS 2, LAU 1)
- Natural increase per 1000 population (NUTS 2, LAU 1)
- Students per 1000 population (NUTS 2)
- Doctors per 10 thousand population (NUTS 2)
- Average monthly gross wages and salaries (NUTS 2)
- Entities of national economy per 10 thousand population (NUTS 2, LAU 1)
- Gross domestic expenditures on R&D per 1000 population (NUTS 2)
- Number of beds per 1000 population (NUTS 2)
- Number of foreign tourists per 1000 population (NUTS 2)
- Tourist traffic intensity index by Charvat (NUTS 2)
- Percentage of employed persons in market services (NUTS 2)
- Employed persons per 1000 population (NUTS 2)
- Dwellings completed per 1000 population (NUTS 2, LAU 1)
- Unemployment rate (NUTS 2)
- Gross Domestic Product per capita (NUTS 2)
- Area of national and landscape parks in % of grand total area of unit (NUTS 2)
- Emission of particulate pollutants from plants generating substantial air pollution per 1 sq km (NUTS 2)
- Emission of gaseous pollutants from plants generating substantial air pollution per 1sq km (NUTS 2)
- Forest cover in % (NUTS 2).

Domain set of variables - "Society and environment":

- Internal and international net migration for permanent residence per 1000 population (NUTS 2)
- Infant deaths per 1000 live births (NUTS 2, LAU 1)
- Population at non-working age per 100 persons at working age (NUTS 2)
- Natural increase per 1000 population (NUTS 2, LAU 1)
- Median age (LAU 1)
- Industrial and municipal sewerage treated in % of sewerage requiring treatment (NUTS 2)

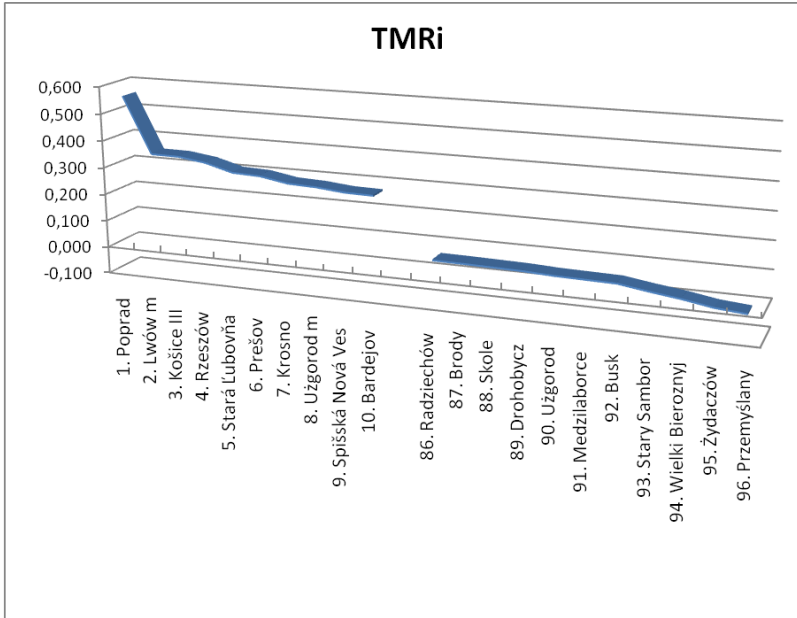
- Emission of particulate and gas pollutants from plants generating substantial air pollution per 1 sq km (NUTS 2)
- Legally protected areas possessing unique environmental value in % of total area (NUTS 2)
- Forest land area per capita in ha (LAU 1)
- Municipal waste collected per capita in kg (NUTS 2).

To obtain the taxonomic measure of development the following steps should be taken. First, a set of variables which contains essential elements undergoes the process of standardization. As a result, a matrix of standardized values of features is obtained. Then, a set of the taxonomic patterns, whose coordinates are given in the form of a vector is being determined. In relation to the pattern, distances are calculated for each tested object (region), using the Euclidean metric. In this way, taxonomic measures of development of objects are being obtained, which undergo the process of standardization. They take values whose level depends on development of an object or region – the more developed the object is, the higher values the measure takes.

Taking into consideration spatial units on NUTS level 2, we receive the following ranking, in which Małopolskie voivodship is placed first and Volyn oblast is ranked last.



As far as the units on LAU level 1 are considered, a similar ranking can be created, in which Poprad from Slovakia is ranked the highest and Przemyślany from Ukraine the lowest.



### Stages of SWOT analysis of administrative units with the application of the Analytic Hierarchy Process (AHP)

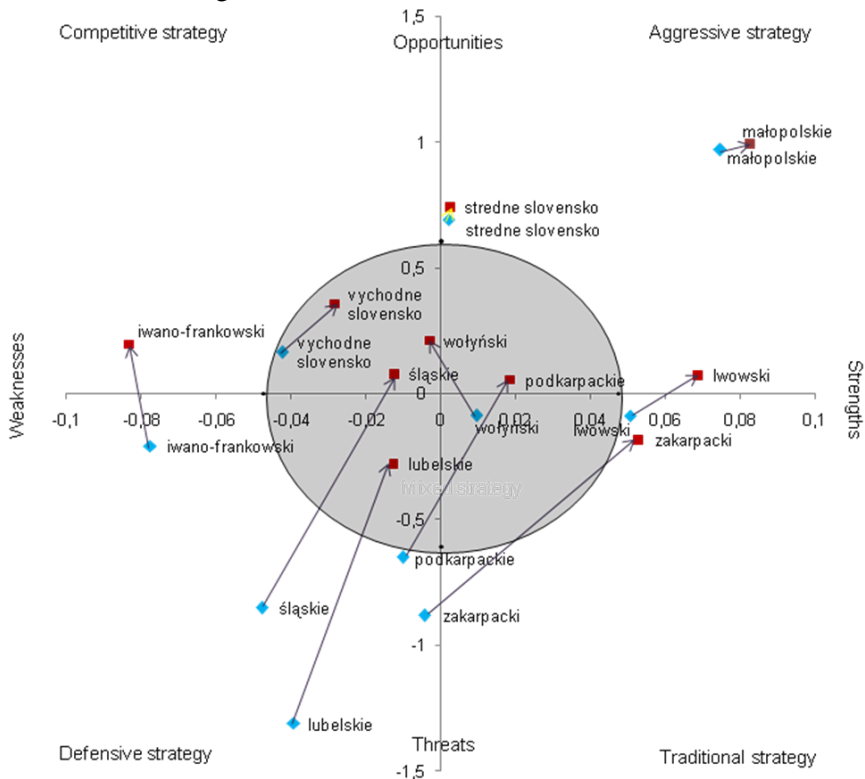
AHP allows one to assess the validity of the factors affecting the development of local units (strengths and weaknesses as well as opportunities and threats), to determine their synthetic assessments of internal and external conditions, and on this basis, to identify the types of developmental units.

Construction of two hierarchical structures related to the external and internal conditions of socio-economic development – specification of the strategic fields and an appropriate choice of sub-indicators for them (selected elements of a SWOT analysis of administrative unit) is the first step in developing this indicator. Then, experts assess the chances and risks as well as strengths and weaknesses of particular administrative unit. The next one important stage consists in determining indicators weights of the SWOT analysis with the application of the AHP method. The AHP method involves pair-wise comparison of SWOT indicators at each level of the hierarchy with the application of the *Saaty's Scale*. *Saaty's Scale* defines the nature of dependency (e.g. an index has the same significance as the second one, an index is absolutely more important than the other). Verification of the correctness of carried out comparisons with the application of the *CR marker*, which determines the extent to which cross-comparisons of the characteristics are consistent, as well as standardization of indicators in cross-administrative units are also significant. The last two steps

include calculation of coordinates value of administrative units position in terms of external and internal development conditions, and presentation of the position of administrative units on in the diagram that is divided into quarters representing four types of growth strategies: aggressive, conservative, defensive and competitive.

Aggressive strategy (maxi-maxi) is created for areas where strengths and opportunities prevail in their environment. This is a strategy of strong expansion and diversified development. Conservative strategy (maxi-mini) is based on the large internal potential but must also attempt to overcome the threats from the outside. Competitive strategy (mini-maxi) is created for areas where strengths overweigh weaknesses, and the layout of external conditions is favourable. Defensive strategy (mini-mini) is created for areas where there are fewer developmental opportunities than in areas belonging to other strategies. These areas exist in less favourable surroundings and their potential for growth is weaker. This strategy is mainly to provide them such a position that is currently in the country and minimize risks and weaknesses occurring within the area. For other areas that do not have clear strengths and weaknesses mix strategy is created (so-called mixed strategy).

As a result of AHP procedure we get assignment of a given spatial unit to one of five kinds of strategies.





The arrows indicate changes of strategies which could be observed during a year. According to this graph many regions are very active aiming at higher level of socio-economic development. The formal evidence is the direction of the arrows that is up and right.

Analyzing spatial units on LAU level 1 we can notice some differences in comparison to analogous figure concerning NUTS level 2, namely not all units on LAU level 1 evolve towards better socio-economic situation.

Spatial models (spatial error models, spatial delay models), allow for tracking demand and supply shocks both in time and space.

Taking into account the location and neighbourhood one can evaluate quite precisely and comprehensively the position of the surveyed territorial unit as compared to others – whether, e.g. it is an island among completely other ones, or a part of a greater territorial structure.

## **Conclusion**

There are some important steps which should be taken to create the coherent research system for transborder areas. The first one is determination of the specificity, and creation of a typology of border crossings. In order to create an effective typology we have to take into account specification of the factors that determine the intensity of border traffic (e.g. type of road which leads to the crossing, the number of lanes, border crossing infrastructure, size of towns near the borders) and use of equipment for automatic traffic measurement (co-operation with the General Directorate for National Roads and Motorways, government and self-government authorities, technical universities).

Another essential factor is the determination of the specificity and creation of a typology of border areas. This process should take into account specification of the most important endogenous and exogenous factors affecting the economy and regional development (entrepreneurship, labour market, border trade, exchange rates, price levels, environmental protection, living conditions, migration, tourism, etc.). The clustering of border areas in terms of similarity is very useful in analysis of transborder areas.

Inventory of information resources of official statistics is another important element. It may happen that we have unknowingly a lot of information concerning transborder areas. Sometimes only a deep insight into statistical databases or a little modification in statistical forms is required so as to adapt survey to our needs.

We have to bear in mind that employing extra-statistical sources of information can be fruitful, that is using mobile telephone operators, data on

passenger air traffic, rail traffic, travel agencies, regional offices - permits to work issued for foreigners, employment agencies assisting in employment of citizens abroad by foreign employers and foreign administrative sources such as Workers Registration Scheme in the UK, Social Insurance System (e.g. Personal Public Service in Ireland).

The next interesting path to follow is combining information from the registers and sample surveys especially in terms of budget constraints.

While analysing social-economic processes in transborder areas, the taxonomic measures of development play a valuable role, in particular, spatial models. They allow, among other things, for detecting whether there was diffusion, exchange or interaction – whether other regions become infected or whether local changes are a response to exogenous shocks. The results of such modelling can be disseminated on the local, regional, national and international levels (micro-mezo-macro-inter), and can be useful not only for analysts but also for policy-makers and decision-makers in business.

What is crucial in creation of a coherent research system is determination of the areas of research and, of course, conducting joint research what in general means creating a uniform information infrastructure, in other words creating knowledge base on transborder areas (individual data, micro-aggregates, aggregates, macro-aggregates, meta-information, para-information) and methodological reports.

## REFERENCES

- CIERPIAŁ-WOLAN M., (2008): *Cross-border surveys – some methodological aspects*. “Statistic in Transition new series”, Vol.9, No. 2, pp. 361-366.
- HARKER P.T., VARGAS L.G. (1990): *The theory of ratio scale estimation: Saaty's Analytic Hierarchy Process*. “Management Science”, Vol. 33, No. 11, pp. 1383-1403.
- CHANG H.-H., HUANG W.-C. (2006): *Application of a quantification SWOT analytical method*. “Mathematical and Computer Modelling”, Vol. 43, Issues 1-2, pp. 158–169.
- LUCZAK A., WYSOCKI F. (2010): *Ocena strategii rozwoju powiatów woj. wielkopolskiego z wykorzystaniem analitycznego procesu hierarchicznego*. „Wiadomości Statystyczne”, No. 3, pp. 54-67.



STATISTICS IN TRANSITION-new series, December 2011  
Vol. 12, No. 3, pp. 547—552

## COOPERATION IN THE FRAMEWORK OF TRANSBORDER AND REGIONAL STATISTICS

**Semen Matkovskyy<sup>1</sup>**

Statistics of cross-border areas plays an increasingly important role. It contributes to regional development and international cooperation. Official statistics satisfies the growing demand for information about cross-border areas. One of the key elements of the research work in border regions is to create a unified information infrastructure.

The specificity of shaping cross-border statistics lies in the fact that the object of research is the areas of two or more states, each of which is subject to national tax, financial and other legislation, as well as is covered by national statistical system.

Data of cross-border statistics are necessary for:

- the regional authorities to provide management in border and cross-border regions, namely for analyses which are carried out by municipalities, counties and provinces on both sides of the border,
- establishing regional development in all fields,
- the needs of regional science,
- the needs of entrepreneurship,
- the needs of society of cross-border regions.

The main emphasis is put on basic statistical information of cross-border cooperation, which is the statistics of regional border areas.

Cross-border cooperation poses some challenges for regional statistics, chiefly the methodological ones.

The experience of regional statistics of the Polish-Ukrainian border area shows that the stage of quite spontaneous cooperation, which consists primarily in the preparation of joint publications, should be developed into a great systematic work to establish a system of cross-border statistics.

The cooperation of the Ukrainian and Polish statisticians in the border area has reached such a level, that what should follow in order to bring a new quality to this cooperation is the establishment of cross-border data bank.

A characteristic feature of the modern-day Europe is the rapid development of international cooperation of regions. Cooperation between regions in Europe contributes significantly to the strengthening of democratic and political stability, in their economic, environmental, cultural and social development.

---

<sup>1</sup> Main Statistical Office in L'viv Region, Ukraine.

Cross-border cooperation has become a factor in promoting social dialogue and social equality, a tool for development of all sectors of the cross-border cooperation. Its place in the regional politics of cross-border regions is determined by the ability to activate and efficiently use the existing capacity of cooperating regions, joining the potential for solving social problems and challenges of territorial development in the transboundary region.

So, what does the cross-border cooperation mean?

Following the Madrid convention, transboundary cooperation is defined as “...a common action aimed at strengthening and intensification of good-neighbourly relations between territorial communities or authorities which are under the jurisdiction of two or more contracting parties”.

Cross-border cooperation clearly defines the cooperation of adjacent areas of neighbouring states, that is the determining factor is the presence of the boundary between the territories.

The cross-border cooperation clearly defines the cooperation of the adjacent areas of neighbouring states as its determining factor in the presence of the boundary between the territories.

The basis for cross-border cooperation is the process of the creation of connections and contractual relations in border regions in order to find solutions to common and sometimes even identical problems.

Cross-border cooperation is aimed at eliminating the negative aspects of the existence of borders, and the consequences that have arisen in border regions due to their location on the national outskirts of states and aims to improve people's lives.

The main objectives of such cooperation are:

- the elimination of existing stereotypes and beliefs on both sides of the border;
- the elimination of political and administrative barriers between the neighbouring nations;
- establishment of economic social and cultural infrastructure.

At present, information, the lion's share of which makes up the statistical information, is one of the key assets of cross-border cooperation.

Statistics of cross-border regions is playing an increasingly important role. It contributes to regional development and cooperation between the countries. Official statistics meets the growing demand for information about cross-border regions. One of the most important elements of the research work in border regions is the creation of a unified information infrastructure.

The specifics of the formation of cross-border statistics lays in the fact that the object of research is the adjoining territories of two or more countries, each of which is a subject of national legislation, taxation, financial and other systems, and is provided by the national statistical information.

Cross-border statistics data are needed:

- for regional authorities for the management of border and cross-border regions, namely, for the analysis, which is carried out by municipalities, districts and regions on both sides of the border;
- for adjusting and planning of regional development in all spheres;
- to ensure the development of regional science;
- for the needs of business;
- to provide information to the population of border regions.

The main accent is given to the regional statistics of cross-border territories, which is the basis of the statistical information related to the cross-border cooperation.

Regional statistics in general, is a part of the official statistics of a particular state. It provides a quantitative description of the socio-economic phenomena within the administrative boundaries of a region.

As practice shows, the regional statistics does three main functions:

- informative function - provides collection, processing, updating of databases, both at local and state level, a hand-over data to users (authorities, government, science, etc.). The source of these data are primarily statistical surveys and censuses, registers, sample surveys, etc.
- analytic function - the conversion of primary information and the following creation of the system of analytical indicators to analyze the effectiveness of socio-economic development.
- the function of providing information and statistical services – both informational services and analytic services, resulting in the assessment of the effectiveness of implemented regional programs.

One can assert that the most important aspect of regional statistics is its informative function. Its effectiveness enables the eventual request for an analytic function, as well as for other information services.

For cross-border regions, especially for those which are at the outer borders of the European Union and those are the regions of Poland, Ukraine and Belarus, there exists an urgent problem of comparability of statistics, especially at the lower level of aggregation than a region / province.

The cooperation of Polish, Ukrainian and Belarusian statisticians began in the 90's of last century as a consequence of the formation of Euroregions and consisted mostly in the preparation of general statistical publications in this framework.

To mention, on the Polish-Ukrainian-Belarusian borderland we have:

- Euroregion "Karpaty" (1993), which includes five counties of Hungary, Podkarpackie Voivodship of Poland, the two krajs (Košice and Prešov) of Slovakia and 4 regions of Ukraine (L'viv, Zakarpatska, Ivano-Frankivs'k and Chernivtsi);

- Euroregion "Bug" (1995), which includes on the Polish side - Lublin province, on the Ukrainian side - Volhyn region and the two districts of L'viv region, on the Belarus side - Brest region.

Also, in the recent years, the bilateral Ukrainian-Polish statistical publications dealing with the assessment of socio-economic development of the border regions has come out.

It should be noted that this cooperation takes place exclusively at the initiative of the statistical offices of the border regions.

Since the beginning of such collaboration, which originated in 1995, more than 30 statistical publications have been published.

Summing up the publication activity of the border departments of Statistics of Ukraine and Poland, we can conclude:

1. Most often the subject of statistical publications is demography, and also data concerning the description of territory, employment, infrastructure and tourism.

2. Virtually all publications are for informational purposes only.

3. There is a lack of publications that would describe the effectiveness of realization of development projects for cross-border regions.

Exchange and analytic processing of statistical data of cross-border regions from different sides of the border has revealed a number of methodological problems:

1. The methodological comparability, comparability of classifications and terminologies.

2. The quality of data, their relevance, completeness and representativeness due to the different types of statistical observations.

3. Access to data, the possibility of disaggregation at the lowest level of administrative and territorial hierarchy. Data confidentiality policy.

The key problem is the first one. It requires a thorough examination and concordance of used concepts and classifications. Here the following three cases can be distinguished:

- full methodological consistency;
- incomplete methodological consistency, but with the possibility of data comparability;
- the impossibility of data harmonization and comparability.

There are significant methodological divergences in environmental statistics, social infrastructure, which are associated primarily with the peculiarities of national classifications.

An important aspect is also comparability, availability and representativeness of data for the administrative-territorial level below NUTS2, that is very important for cross-border statistics.

Addressing these issues requires a meticulous study and analysis of regional statistics of borderland territories and has to become an additional impetus for the development of regional and cross-border statistics.

The cooperation of the Ukrainian and Polish statisticians at the borderland has reached a stage, where the setting up of the cross-border databank furnishing such cooperation a new quality has become apparent.

The accomplishment of the objectives of creation the databank would permit:

- to create an international public Internet databank that meets a common methodology of Eurostat;
- create a coordinated and open to general public metadata database;
- to achieve the harmonization of statistical data;
- to eliminate language barriers when using the databank, since it can be made multi-lingual (at least trilingual).

Such a databank could form a basis for a unified informative infrastructure of the Polish-Ukrainian cross-border cooperation with the possibility of further territorial expansion.

Such work on the databank creation would be useful not only for statisticians of the border regions, but also for Statistical Offices of Ukraine and Poland at large.

Personnel for the job is available on both sides of the Polish-Ukrainian border. However, it is necessary to solve some organizational and financial issues. Here we greatly hope for the support of the main Statistical Offices of Ukraine and Poland.

## **Conclusions**

Borderlands in the integrating Europe is increasingly becoming a subject of interest of regional statistics.

Exchange of information at the transboundary promotes good neighbourly relations on both sides of the border. Regional statistics at borderland is also an informational basis for the preparation and implementation of various types of cross-border cooperation. It is used for the preparation and implementation of various projects, as well as in the process of the working out a strategy of cross-border region development.



Cross-border cooperation, in turn, poses certain challenges for regional statistics, primarily of methodological character.

The experience of regional statistics on the Polish-Ukrainian borderland shows that from a rather spontaneous cooperation, which consisted mainly in the preparation of joint publications, one has to move towards a more systematic work on the developing of cross-border statistics.

For this purpose, it would be desirable to establish a joint centre of transboundary statistics having the objectives:

1. Harmonization of statistical methodology of the regional statistics of cross-border areas.
2. Creation and maintenance of Web-based cross-border databank with an interface to other regional and local databanks, first of all with the regional Eurostat databank, and aimed at a wider range of users (not only of cross-border areas).
3. Coordination of statistical observations of cross-border areas.

STATISTICS IN TRANSITION-new series, December 2011  
Vol. 12, No. 3, pp. 553–558

## NEW CHALLENGES WITHIN CROSS-BORDER STATISTICS

Stanislav Drápal<sup>1</sup>

### ABSTRACT

The Czech Statistical Office has participated in the cross-border statistics since 1990, with first experience in the Euro-region among the Czech Republic, Poland and Germany (Saxony). The CZSO has later participated in several bilateral or multilateral projects focused on cross-border statistics together with Germany, Poland, Austria and Slovakia.

The cross-border statistics is not any one-way road. It gives possibility to exchange experience and knowledge among all participants. It should be an equal partnership of all participating countries and statistical institutions.

The cross-border statistics is a specific part of the official statistics. It is inhibited by identical limits as well and it shares the same statistical tools and methods and it has the same goals and challenges like the statistics in general.

The current statistics must respect two main restrictions: increasing budget and request to decline administrative burden. The consequential demand is to use more administrative data and data from usual statistical surveys.

The new requests on statistics in general have to be applied to cross-border statistics as well. The cross-border statistics should reflect new economic development. The movement of finance, manpower, goods and services has been rapidly changing nowadays. The permanent demand has been described by environmental conditions and factors, but it also should reflect new trends concerning well-being.

The future development of the cross-border statistics demands higher level of collaboration not only between neighbouring countries. It is very important to exchange and to concentrate experience and knowledge from many countries. This is a way to accelerate and to increase efficiency of the cross-border statistics in all countries.

---

<sup>1</sup> Czech Statistical Office, Czech Republic.

## **I. Experience of the Czech Republic**

Up to 1990 statistical experience in the Czech Republic did not involve any encounters with cross-border statistics. While national statistics were well-developed and provided wealth of information for various scopes of detail, from a municipal and district level up to regional statistics, compatibility on the international level was ignored. Year 1990 marked the beginning of profound changes in what used to be Czechoslovakia, political make-over first, followed by economical and social transition. Both the country and its economy started to open up to the world. Related to these developments new demands on statistics emerged. Among others, first requests for cross-border statistics were registered. Czech Statistical Office attempted to accommodate these demands by requesting relevant information from the EU member countries. However, advice received was based on already well-developed practices rooted in different political, social and economic conditions. A significant progress in establishment and development of cross-border statistics in the former Czechoslovakia and after 1993 in the Czech Republic can be accredited to setting up early contacts with the German colleagues in 1990, especially with those in the Federal State of Saxony. It was much more straightforward to apply experience from implementation of the statistical system of the so-called "old" German federal states in Saxony because of its similar conditions to those in Czechoslovakia up to 1989. Thus the first historical lesson learned is the fact that a preferred way is to take advantage of experience of a country which took lead and is either going through a similar transition process or have already accomplished it. Doing this is better than striving for adoption of well-established practices which work in quite a different social system.

The second lesson is in learning an importance of proper sizing of both resources and capabilities of statistical data, while balanced against the needs of cross-border statistics users. At the beginning of the nineties statistics in the Czech Republic has only started its transition to European Statistical System (ESS) methodology. The first task thus was to develop a new statistical methodology which is completely harmonized with the ESS methodology. This meant in the first place changing over statistical classifications and nomenclatures, ensued by changes in the system of statistical surveying. A significant change in the system represented switching from exhaustive surveys to sampling surveys. This step raised the quality level of statistical data and made it more effective on one hand, but on the other hand it reduced the scope of statistical data available at the municipal and district levels. Application of the concepts of confidentiality and individual data protection represented another fundamental change. Up to that point users were used to get data on any individual economical entity and all of a sudden these data either did not exist or it was not possible to provide them any longer. At that time a low quality level of information and communication technology (ICT) equipment represented another

limiting factor. Differences in legislature of the Czech Republic, EU member countries and other non-EU countries presented yet another significant problem.

As statistical data users, foreign investors coming into the Czech Republic were accustomed to information services provided by EU member countries' statistical offices and to nationally comparable data. By contrast a community of domestic users has only started to establish, along with their demands for statistical data. It took a rather lengthy enlightenment process before new users in the Czech Republic realized what they can expect from statistics and how they can utilize statistical data for their decision-making.

At the beginning of the nineties development of the Czech statistics could draw on the experience of Germany. On one hand it was a newly established statistical system in the Federal State of Saxony and on the other hand a mature system in the Federal State of Bavaria. Co-operation was also advancing quickly with our neighbours in Austria and Poland and – after the partition of Czechoslovakia – with the new neighbour Slovakia.

## **II. The current position of cross-border statistics in the Czech Republic**

In the past 20 years we have succeeded in building cross-border statistics into an important component of the national statistical services in the Czech Republic. Its methodology is fully comparable with those of neighbouring countries. Statistical information is provided through state-of-the-art ICT tools and by far most of statistical data are available through the web pages of the Czech Statistical Office in Czech and English. A number of statistical databases are maintained in bi-lateral or multi-lateral co-operation of national statistical offices. Based on demands from the principal users, statistical data sets are custom-processed and publications generated on so-called euro-regions. Apart from securing the functional methodology and technical base for cross-border statistics, we manage to secure operational financing as well, especially for on-demand publishing.

The areas of cross-border statistics methodology, technical support and finances have been provided for by the headquarters of the Czech Statistical Office. However all field activities have been conducted by its regional statistical offices (on the NUTS 3 level).

The main areas of utilization of cross-border statistical data are:

- Provision of data to investors for conducting their research, development of programs and projects supported by EU funds.
- Mutual research and recognition of ways of living in border regions of neighbouring countries.
- Education

- Tourist information

There are, however, new challenges emerging, aside from the successful results achieved so far.

### **III. New challenges**

#### Budget

Global financial and economy problems resulted recently in budget cuts in the area of national statistical services as well. For example, during the last two years Czech Statistical Office had to cut the number of its employees by almost 20 %. Further cuts are probably still ahead in the coming years. Cross-border statistics service must look for ways of securing service continuation for its users on a lower budget, just like all the other parts of the national statistical service, while maintaining its performance.

This means improvement of exploitation of all data acquired during statistical surveys is necessary. Better utilisation of administrative data sources is needed as well. In recent years we have been able to achieve better utilisation of administrative data from the Ministry of Finance databases (tax payers, distribution of tax revenues, subsidies to municipalities and regions, various government subsidies). The Ministry of Interior made available data that were previously obtained from municipalities, such as migration of population, natality and mortality, foreigners. Co-operation with the Ministry of Agriculture and Environment has grown more effective as well. Data are now available on agricultural production, land utilisation, wastes and pollution. All these statistics are also important for cross-border statistics.

#### Statistical information contents

Contents of cross-border statistics corresponds to general contents of statistical data. Fundamental discussions on future developments of statistical data contents must therefore touch cross-border statistics as well. It will be necessary to continue to provide statistical data on population, land or economy on regional level. But there is also an open question whether regional GDP is the best indicator of a level of regions' economical development and also a measure of a standard of living of regions' population. While new approaches have been devised and tested to define a state of well-being on national and international levels, these methods must be applied in regional statistics too. Of course this holds true even under the limiting condition of constrained budget. Undoubtedly, most indicators characterizing well-being on the national level will also apply on the regional level, but each region has its own specifics. When dealing with cross-border statistics, additional specifics between neighbouring countries need to be

considered as well. This implies that a general solution of the problem described above will require specific modifications for regional and cross-border statistics.

#### Provision of statistical information

In the recent years availability of statistical data has shifted from a printed form to electronic one. Getting data on the Internet is a common matter nowadays and there is a new question whether and how to involve official statistics into social networks. Statistical information should be available not only on web pages of statistical bodies, but it is desirable to make it accessible through web pages of municipalities, chambers of commerce, professional associations, regional schools and colleges or tourist information centres. A specific language complexity involves databases with cross-border statistics spanning territories of several countries. E.g. a successful euro-region, which covers three border regions of the Czech Republic, Saxony and Poland, publishes its information in three languages. This requires additional human and financial resources related to creation and maintenance of database data translated in three languages. Streamlining can be achieved by settling on a single language, but this solution is usually not welcomed by users.

### **IV. Recommendations**

Experience of the Czech Statistical Office accumulated so far and new solutions of current issues has laid a base for putting forward several recommendations aimed at further development of cross-border statistics both in the current EU member countries and in other non-EU countries:

1. When establishing and developing a cross-border statistics system, it is advisable to utilise experience of statistical bodies of countries which got through a similar process. These countries can share their knowledge of the process outcome, but also their experience from overcoming various problems while implementing the system.
2. Cross-border statistics cannot be fully successful unless official statistical bodies of the participating countries use internationally comparable statistical methodology, including both classifications and data acquisition and processing methodologies. It is also necessary to consider implications of national legislatures, even though this is beyond competencies of statistical offices.
3. It is important to know well the needs of cross-border statistics data users and to find effective ways of filling these needs. Provision and technical means of presentation of statistical data to users should be based on the same principle.
4. Systematically develop information campaign aimed at potential users of cross-border statistics. Focus on the young generation, school and college

students who represent the future user group on all society levels. Gain support of mass media for promoting utilisation of cross-border statistics to the public.

5. Current conditions require joining forces and sharing best practices to eliminate the impact of budget cuts. Similarly, join efforts to share best practices and look for ways of finding new contents of statistical data information. In particular, develop solution to the problem of expressing state of well-being on a regional level and in the international context. The key role in this respect should be assumed by Eurostat and statistical divisions of other prominent international institutions.

STATISTICS IN TRANSITION-new series, December 2011  
Vol. 12, No. 3, pp. 559–566

## **INTERNATIONAL COOPERATION AND CAPACITY DEVELOPMENT IN GEOSTAT**

**Tengiz Tsekvava<sup>1</sup>**

### **ABSTRACT**

Our paper will consider the processes taking place at the National Statistics Office of Georgia (Geostat) since the beginning of 2010. The focus will be made on the influence of statistical cooperation with the EU statistical offices and international statistical organizations (including Eurostat, in particular) and ENPI members.

Starting from January 2010, a new Law on Statistics elaborated with the support of the Eurostat became effective. The Law guaranteed independence of Geostat as a legal entity of public law (previously being a Department within the Ministry of Economic Development), providing for more flexibility in planning and operating activities.

The change in legal status led to Geostat's more intensive activity in its relations with foreign statistical offices and international organizations with the interest in statistics.

The activities involving international cooperation can be grouped into 3 main directions: 1) evaluation and strategic planning (World Bank support with NSDS, Eurostat and IMF mission assessments), 2) technical assistance aimed at supporting practically all areas of Geostat's activities in the medium term (USDA, Sweden statistics/SIDA, Dutch CBS, EU/TAIEX and other institutions); and 3) support at the operational level (UN institutions, Millennium Challenge Georgia Fund, USAID, World Bank and others).

The immediate and indirect outcomes of the international cooperation have a crucial influence on the main priorities of Geostat: increasing credibility of statistical information among its users, developing professional competence of Geostat's personnel, and strengthening relations with other producers of statistical information in the country.

A short discussion of the most recent example of cooperation with Armenia in the field of International Comparison Program will be given, with specific positive outcomes from such cooperation.

General implications from international cooperation will be summarized.

---

<sup>1</sup> National Statistics Office of Georgia.



## SUMMARY

The present paper attempts at briefly describing the role of international cooperation in the development of the statistical agency in Georgia. The support of international community has led to visible successes in the process of implementing standard statistical methodologies, improving scope of statistical data and observing periodicity and timing of data dissemination.

At the same time, main challenges hindering further development are discussed. The necessity of wider coordination of activities related to statistical production is underlined, with the role of international organizations remaining crucial in this process.

### 1. Background

Collapse of the Soviet Union and abolishment of the central planning agency (Gosplan) led to creation of new states with independent social and economic institutions, including statistical systems. Drastic contraction of the Georgian GDP (estimated to fall as much as 70% relative to the 1990 level) was aggravated by a civil war and political unrest in the country. As a result, the country suffered from an acute economic and social crisis, migration processes accelerated, while hyperinflation of the transitory national currency (“kupon”) exceeded 10,000%. In this situation the conditions for operation of the Georgian statistical office (initially called Social and Economic Information Committee) were far from favorable.

Economic stabilization was achieved in mid-90s with critical assistance of the international organizations. Along with economic and social reforms (currency reform, price liberalization, etc.), the reforms in the area of statistics were also initiated.

### 2. Independent Operation of the Statistical Office

As it was already mentioned, the Department of Statistics (DS) found itself in an absolutely new situation in the beginning of 90s. Difficult economic situation in the country conditioned availability of very limited resources for the DS, sharply narrowing the scope and scale of

statistical production. On the other hand, the functions of the DS’s predecessor, the Central Statistical Division, used to be completely different in many aspects, as hundreds of employees were accustomed to i) full coverage data in different economic and social spheres of statistics, with very high level of formal compliance; ii) focus on data-collection-related activities with less emphasis on data analysis (partly due to the fact that large amount of data was processed centrally in Moscow. The original questionnaires of the 1989 census, which were sent to the central statistical office of the Soviet Union and never

received thereafter, can be cited as an example); and iii) supervision functions of different public agencies, where statistics employees were regularly sent to monitor and audit certain indicators.

In this situation it was obvious that the newly established DS required significant assistance. Support in establishing and improving statistical methodologies and business processes came from different international organizations and through bilateral assistance. Support rendered by the IMF/World Bank, the UN, and the EU was critical in the first years (namely, in business statistics, national accounts, prices), as the assistance from the UK's DFID (particularly active in social and demographic statistics), the Dutch government, and other donors followed.

Active international cooperation of the DS with its partners continued after the Rose revolution in 2004. Apart from technical assistance (regular visits of missions, support in staff training, etc.), significant amounts of money were invested in statistical surveys and building of infrastructure (the UN agencies, the EU, and the World Bank are to be mentioned in particular). Still being significantly underfinanced relative to other government agencies, the DS saw these financial injections as an important source for supporting its statistical operations (e.g. specialized surveys for production of national accounts data supported by the EU/WB/UNDP). The level of donors' financial assistance to the Geostat only in the form of grants with respect to budget financing during 2005-2010 oscillated between 5-50 percent! (See: Table 1).

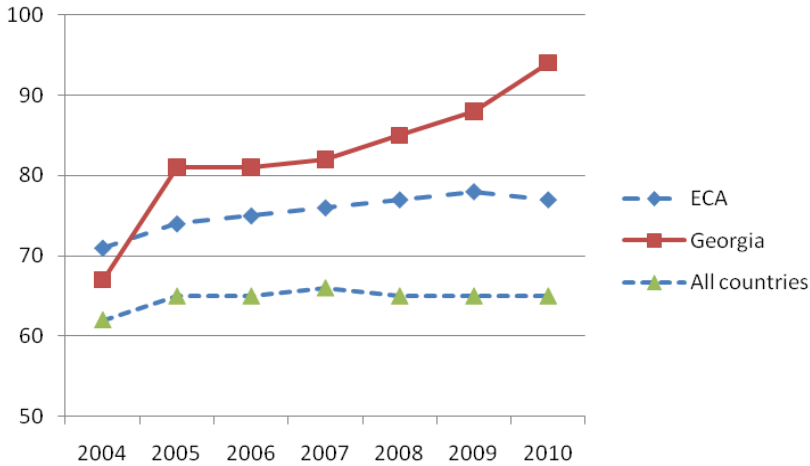
**Table 1.** Grants and Government Financing of Department of Statistics/Geostat in 2005-2010 (GEL thousands)

	2005	2006	2007	2008	2009	2010
Government Financing (1)	2,180	1,947	3,006	4,672	4,332	3,215
Donors' Assistance (2)	318.7	652.8	124.7	955.6	1,167	1,637
Ratio (2):(1)	14.6 %	33.5 %	4.1 %	20.5 %	26.9 %	52.4 %

Positive tendencies (accumulated experience in the DS, increased government financing, closer cooperation with international organizations) clearly translated into tangible improvements in statistical capacity. According to the World Bank's statistical capacity indicator measuring progress in quality of statistics among the developing countries, Georgia's overall capacity score rose from 67 to 94 (out of a maximum of 100), posting one of the fastest growth rates among the developing countries<sup>1</sup>. In 2010 Georgia became member of the IMF's Special Data Dissemination Standard (SDDS), which served as another proof of progress made.

<sup>1</sup> See: Chapter 2 in *Strategy for the Development of Statistics in Georgia (NSDS)*.

**Chart 1.** Statistical Capacity Indicator for Georgia and Other Developing Countries in 2004-2010



*Source: World Bank, taken from NSDS.*

A recent example of a new form of international cooperation represents the Geostat's bilateral comparisons with the Armenian statistical office in the framework of the World Bank's International Comparison Program (ICP). Being no longer part of the CIS, Georgia entered the 2011 round of ICP through bilateral comparison with the neighboring country. Apart from the fact that bilateral cooperation with the Armstat ensured Georgia's inclusion in the Program, bilateral cooperation proves to be very useful in terms of capacity development as well. In contrast to the 2005 round, when the staff's activities were mostly related to price collection and other technical work, the bilateral price validations and comparisons fully performed by personnel of the two agencies (with support of international experts) represent the best practical exercise possible in this area.

### 3. Current Situation and Future Challenges

Along with positive developments which took place in the recent years, the government of Georgia sought further improvement in the area of statistics. In this regard, with the assistance of the Eurostat, a new Law "On Official Statistics" has been elaborated and adopted in December 2009.

The Law stipulates institutional independence of the Geostat, which became a non-governmental legal entity of public law (LEPL). The Geostat (National Statistics Office of Georgia) is governed by an 8-member Board, with 3

representatives from the government (Ministry of Finance, Ministry of Sustainable Development and the National Bank) and 5 non-government members, including the Chairman which simultaneously acts as the Executive Director of the Geostat.

A new institutional form of the Geostat provides more flexibility in the context of international cooperation as well. The Geostat's permanent staff are no longer civil servants – a status that used to limit their involvement in additional projects. According to the new Law, the Geostat is entitled to conduct practically any type of survey or/and research for private companies or international organizations.

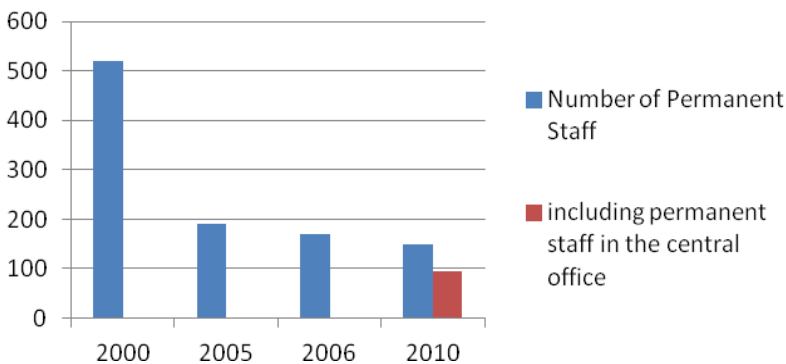
However, the legal independence of Geostat does not automatically resolve all the existing problems which still remain in place and which require more significant efforts, often beyond the immediate area of activities related to statistical production.

The issues which stand in front of the Geostat and not new, as a lot of them have been accumulated over the years:

*a. The number of statistical personnel and physical infrastructure*

The Central Statistical Division of the USSR comprised more than 1000 employees, but their functions became largely irrelevant with respect to the post-Soviet requirements. However, a drastic decrease in permanent staff occurred in the last 10 years saw their number shrinking to a critical level (See Chart 2).

**Chart 2.** Number of Permanent Staff in Geostat



Steady growth of budget revenues as well as efforts to reduce the government sector in the recent years led to a significant increase in salaries in the country's public sector; however, the salary growth was accompanied by a permanent

increase in workload, as new requirements created more demand for public services, including demand for statistical indicators.

Decrease in permanent staff particularly affected regional offices (practically involved only in data collection activities), while resources needed for infrastructure development (e.g. IT development to compensate for a smaller number of regional personnel) were far from being sufficient.

*b. Attracting and keeping skilled statisticians*

To a certain extent, the Geostat is theoretically able to mitigate a lack of personnel by means of hiring non-permanent employees on a contractual basis. However, there are two main reasons as to why the Geostat is still not able to compete on the market.

The most obvious reason was and remains the *relatively low level of salaries* in the Geostat. Recent data <sup>1</sup> showed that the average salaries of the Geostat's permanent staff constituted only 54% of the same indicator for 5 Ministries. This implied that in order to cover the salary gap, the relative increase in Geostat's salaries was to equal 23% annually over a four-year period<sup>2</sup>. The expectations that a semi-autonomous statistical office will generate sufficient own revenues, similar to other LEPLs in the country, failed, as the main products of Geostat have the nature of public goods.

Leaving aside the issue of remuneration, however, there exists another significant problem, which is much more difficult to tackle: the *undergraduate education programs* remain to be incompatible with the standards needed for a modern statistical agency. There are basically three main education areas which could potentially provide graduates for the statistical agency in Georgia. However, two directions of these three are mainly irrelevant at this moment, as i) strong Departments of Mathematics in state universities producing mathematicians with a fairly good theoretical level of competence in line with international standards (legacy of a strong Soviet school) have a relatively small number of graduates, and ii) expensive private universities giving also a fairly good undergraduate education mostly in the areas of economics and business

---

<sup>1</sup> See NSDS, Annex 2.

<sup>2</sup> The outlook for 2012 is more positive, as the government promised at least a 30% increase in salaries.

largely see their graduates joining banks and other types of private businesses upon completion<sup>1</sup>.

Thus, there remain mostly the graduates of so-called “economic statistics” programs (from state universities), many of which joined the Statistics Department over the years. But, as it was mentioned above, these undergraduate programs have been mainly influenced by the Soviet legacy (focus on mathematics/physics departments, full coverage statistics) and do not provide adequate instruction of statistical methods and techniques required for sampled surveys<sup>2</sup>.

#### 4. International Cooperation and Current Challenges

As shown above, the support of international organizations, although as large as it has been, cannot fully solve deeper problems. Advocating for more resources to Geostat has proved to be a difficult task. As it was mentioned, the government shows more willingness to address the issue, but it is likely to take some time until stable financing will allow for longer-term planning. The preparation of NSDS was an important step in this direction since it provides the stakeholders with a full picture of objectives and existing gaps.

Attraction of skilled staff implies that it is *relative* remuneration with respect to other government agencies that matters. However, it is not only competition that affects availability of needed personnel in the statistical area. More support from international organizations and partner countries is needed to tailor the existing curriculum at the universities to the real needs of statistics – a long-term task which required essential efforts from different stakeholders within and outside the country.

It is becoming clear that in order to achieve further progress in the statistical area, actions beyond the immediate scope of Geostat’s activities need to be made. A wider approach to the sphere of statistics is required, where attention will be paid to formation of skills, involvement of different government agencies and ensuring the activities are properly budgeted. And without strong stakeholders –

---

<sup>1</sup> These universities often form cooperation agreements with large businesses (mainly banks), which also provide student loans. Upon completion of the programs students are practically certain of obtaining employment.

<sup>2</sup> Trainings in these areas provided by the international organizations (the IMF and the UN workshops to be mentioned, in particular) are very useful but they cannot replace basic undergraduate education.

first of all, international organizations – the Geostat is not likely to succeed to coordinate the general approach to statistical data production.

## **REFERENCES**

1. Strategy for the Development of Statistics in Georgia, Tbilisi 2011.
2. Law of Georgia “On Official Statistics”, December 2009.

STATISTICS IN TRANSITION-new series, December 2011  
Vol. 12, No. 3, pp. 567–578

## **CHALLENGES TO EUROPEAN STATISTICS' CREDIBILITY - WHY EUROPEAN STATISTICS IS STILL NOT TRANSPARENT AND ITS CREDIBILITY UNSATISFACTORY**

**Vasily M. Simchera<sup>1</sup>, Ali Serhan Koyuncugil<sup>2</sup>**

### **ABSTRACT**

Despite the recognized catholicity and the credibility of European statistics in general (with some exceptions of implemented projects) and in individual European countries still raises number of issues and doubts.

At the same time following the Quételet law, which states the more crimes the less penalties, European statistics today as the most advanced, intellectual and informative one still accumulate doubts and issues towards itself. These doubts, issues and challenges grow in geometric progression and first of all it concerns its credibility, which appears to be its basis and nerve.

Why European statistics credibility is unsatisfactory and tends to go down? Why its transparency and people's confidence are down? At last, why it stirs up the social tensions instead of opposite action in the enlightened European society?

Is it because the European statistics regardless of its virtual image in real life still not transparent, practices double standards, "feeds" society with unnecessary methodologies instead of clear and available information and rather misleading than helping? Or is it because the modern statistics in general and almost everywhere by its nature is dependent and serves authorities and not the truth?

Is it possible to change such situation to the better, and if it is when and on which conditions?

---

<sup>1</sup> Professor, Academician of the Russian Academy of Economic Science, Head of Statistics Division of the Custom Union of the EurAsEC.

<sup>2</sup> The original version of paper can be found in Chapter 3 of the book "Surveillance Technologies and Early Warning Systems: Data Mining Applications for Risk Detection" edited by Ali Serhan Koyuncugil (Capital Markets Board of Turkey, Turkey) and Nermin Ozgulbas (Baskent University, Turkey). Published by Information Science Reference, Hershey - New York, USA (an imprint of IGI Global) <http://www.igi-global.com/chapter/nature-scales-statistical-estimations-divergence/46804>.



General condition of success for any research is the convergence of theoretical assumptions to the facts being observed, and vice versa the facts being observed to the theoretical assumptions. No matter whether we talk about inductive or deductive researches, determined or undetermined facts and their cause – effect connections, natural scientific or socio-economic researches, reliable or less reliable data – the condition always remains the same.

Convergence of theory and practice, forecasts and facts, their adequacy or inadequacy are set by identifications of existing in nature and familiar to science of theoretical and empirical distributions. Convergence at the level of the necessary and sufficient conditions in contrast to abstract ideals is being checked by known criteria of statistical agreement (or in case of its absence) by trial-and-error method and likelihood criteria and common sense. This is the way as many centuries ago the imaginations and theories is tested by facts and practices, and in turn facts and practices is tested by theory and imaginations. And there is nothing else that human mind could invent either in the past nor today. Where a theory relies on the facts and facts relies on theory it is possible to carry out a statistical experiment, which can and indeed gives a significant results with clear sense and paramount scientific and practical importance, and where does not – such an experiment is impossible, and there is no point to initiate this experiment as its results will be false. Unfortunately we should state here that a first case (rather in natural science than in public science) by various reasons, and mainly for a general reason of the limit of knowledge and resources of their realization, is restricted and ultimate, and each successful experiment is interpreted as unique success, while a second case, because of the ignorance of the law of the limited knowledge, is not restricted and infinite. As a result we have domination in a science of simple and mainly false surveys results which are worthless and insignificant, and as a general consequence – depreciation of the efficiency of the science and knowledge's, and total ignorance of them.

The correction of a general situation perhaps requires not only changes in the existing unsatisfactory market treatment, and correction of the negligible treatment towards fundamental researches and labor-intensive experimental results, but raising the systematic level of the knowledge production itself and its application according to the exact form of the identified processes and events in a way their endogen necessity in the world around us. This is the cause of scientific experiment stagnation and further the science as a whole. We should not criticize the external circumstances but think on how to clear a science itself from futile imitation accumulated during centuries, reconsideration of the statistical experiment basis – this is what one should start from and what indeed can help to correct the unsatisfactory situation in the modern science. This means that phenomena and events in the world around us their content, dynamic and structure should not adapt to a format of scientific experiment that is its usual paradigms, algorithms and interpretations but in contrast – the format of scientific experiment itself its set of ideological potential must constantly alter and adapt to a world around us, to catch and produce a future tendencies its fast going and

dominantly differently directed and therefore contradictive changes and give it a shape of essential construction which helps not only for better understanding but also transform our world in efficient and reasonable ways.

Conceiving the situation this way one should start from fundamental basis of the modern scientific experiment, its theoretical hypothesis the basis of which is multivariate statistical distributions and their approximating functions and laws. Depending on how full and certain these functions and laws reflect the structure and dynamic of modern world, the tendencies of its alteration, so this is a degree which define how these functions and laws are applicable today in order to influence the modern events in a constructive way, providing each time the possibilities for effective decision making. The constructive answer to this question demands reconsideration of the whole variety of the existing types of univariate and multivariate distributions their inventory, adaptation and identification applicable to a modern problems solving for production, labor and life. The first step towards obtaining such an answer is typology of present (or in any case most commonly used) distribution functions and laws of observed phenomena and their systematization with regard to problems solving of multivariate distributions as most important and significant ones. There are a lot of distribution functions, but they are separated and can not be united. Part of these functions (linear, normal, power-series, exponential and others distribution functions) acceptable on the level of required and sufficient conditions are approximating observed empirical facts, and have clear substantial interpretation of parameters and sense of the results being received on their basis and these functions are widespread. The other part (the most part) of these functions (logistic, maximum likelihood, and in particular nonlinear and nonparametric distribution functions) are less proved theoretically, and inadequately or even does not reflect the existing empirical distributions and require as a rule robust work while their identification and appliance of them to the problem solving of multivariate analysis. There is also one, some kind of off-balance part of multivariate distribution functions, aimed towards chaotic, partly robust and partly non parametric and further fuzzy distributions, which as a rule lack of any theoretical basis and badly or no way approximating the respective empirical data observed and demand a development of substantially different approaches to their construction. Here we can attribute the synthesis combinatorial problem of distribution function, constructing the unknown hybrid functions on the basis of the existing known ones. In view of exceptional complexity these two tasks are just mentioned here. The overcoming of the existing gaps between theoretical and empirical multivariate distribution functions assumes the typology's ground and representation of these distributions and functions by their resolving powers which are defined by attributes and criteria of necessary and sufficient equivalence. Nowadays one of the most popular ways of using multivariate distributions is data mining. Therefore, data mining has the same gaps mentioned.

It is possible to review data mining under two separate headings as Statistics and Information Technologies. At the beginning of data mining concept discovered in 1990's IT perspective mostly underlined but in 2000's analytical or statistical point of view of data mining has been becoming most integral part of data mining. Hastie et al. (2001) emphasized the statistical view of data mining with statistical learning concept. Rao (2001) linked statistics and data mining with more strong ties and mentioned data mining as a future of statistics. On the other hand, Moss and Atre (2003) mentioned the difference between Classical Statistics and data mining. Data mining mostly interested in big data sets. Therefore, finance is one of the most suitable implementation areas of data mining because of the huge data produces with transactions. Kovalerchuk and Vityaev (2002) inspected the implementation domains of finance in data mining with examples and emphasized that stock prices, currency rates and bankruptcy predictions, claim management, customer profiling and money laundry some of the implementation domains of data mining. In addition, data mining successfully applied to financial performance and distress prediction. Koyuncugil and Ozgulbas (2006a) emphasized the problems of Turkish SMEs and suggested financial profiling as a first step of solutions. Then, the authors determined the financial profiles of 135 SMEs listed in Istanbul Stock Exchange (ISE) according to 2004 ISE data. Koyuncugil and Ozgulbas (2006b) defined a financial performance measure for Turkish SMEs with 2004 ISE data. Koyuncugil and Ozgulbas (2006c) defined the factors which effected financial failers of ISE listed SMEs with 2000-2005 data via CHAID (Chi-Square Automatic Interaction Detector) decision tree algorithm which is one of the most update data mining methods. Ozgulbas and Koyuncugil (2006) and Ozgulbas et al. (2006) determined weak and strenght sides of SMEs in financial mean, financial performance level with 2000-2005 data via data mining. One of the most efficient facilities of data mining is Early Warning Systems because of its definition. Data mining aims to discover hidden relations, covered patterns and then use them for prediction of future behaviours. This definition of data mining makes it the most efficient tool for Early Warning Systems. Therefore, many recent studies in early warning domain has been using data mining. Koyuncugil (2006) developed an early warning system for manipulation and insider trading detection in Stock Exchange Markets and proved that the system works successfully with real data. Simchera (2003a) is given examples of systematic errors, which are made at improper identification of direct and reversed numbers modules, recursive and discursive rates, direct and reversed exchange rates as well as effective interest rates, annuities and bills with recourse, and also numerous examples on collecting and publishing various estimates (such as GDP, life standards etc.) on the basis of the same methods and estimates with the application of different methods. The module increase direct number for example consumer price index with value of 1.25 is 0.25, whilst module of reversed number for example index of consumer inflation would be 0.2 (1/1.25), in everyday practice it is considered to be equal to 0.25 module, which is not right and in turn illustrates as demonstrative as possible

the most spread example of typical routine “misspell” in statistics that is in other words – to judge anything with the same value and present the same thing with different numbers. And from this quite small “misspell” grows fraud in statistics. Keen and Smith (2007) are given with exceptional arguments the examples of fraud in the estimates of British tax calculations; in the working paper of Ruhashyankiko and Yehoue (2006) and in working paper by Mauro (2002) authors argue the corruption in the private sector of economy and its effect on economic growth reduction. The persuasive estimate of debt dynamics and imbalances are given in the working paper by Meredith (2007) and measurement of financial market liquidity is considered in the working paper of Sarr and Lybek (2002). Booth et al. (1999) based on excellent models the authors illustrate the insolvent insurers estimations for portfolio investments, funds and exchange rates, derivatives, pension funds. In the work of Vaitilingam (2007) there are examples of such estimate, in particular examples of divergence of estimates for actuarial indices, based on sample observations (covering 30 companies - Dow Jones Indices up to 500 companies – Standards and Poor’s Index) and mass observations (covering thousands of companies (400 million companies all over the world). In general the methodology of distortion and publishing of unreliable estimates and explanations of cause and sources of fraud in statistics are given in significant work by OECD (2003, 2008).

What we need is to look at the nine types of distributions within two categories (category of linear and category of nonlinear distributions). The adequate choice of which is the first condition for an efficient minimization of fundamental estimation discrepancies in the modern statistics and arising on these basis deep delusions and undisguised lies.

Without such identification all statistical estimations (and further expectations, forecasts etc.) are shifted and have only illustrative meanings. Nowadays, estimation word coincides another concept different from statistics. This new concept is data mining. There are a lot of definitions of data mining. Because data mining is an evolutionary area. One of the most common definition: ‘Data mining is the process of extracting previously unknown, valid and actionable information from large databases and then using the information to make crucial business decisions (Cabena et al., 1997).’ Data mining is not a single step analysis. Data mining is a sequential multi analysis and multi task process. But, mainly the core of the whole data mining process is called data mining.

On the other hand, the process takes place as ‘Knowledge Discovery’ as well. Data mining step of the knowledge discovery process mostly means modified multivariate statistical analysis methods. These methods automatized, scaled to analyze huge data sets via modification. Therefore, it is possible to define data mining as an evolution of statistical methods via Information Technologies and automated processes. Data mining mainly has two different point of view: 1. Statistical, 2. Information Technologies. Data mining from statistical point of view calls ‘Statistical Learning’. Statistical learning let us know what data tells

instead of subjective sayings. Data mining, automatic extraction of predictional strategic knowledge is mostly based on multivariate statistical methods. Discovery process aims to extract valuable knowledge from hidden, covered, unknown patterns or relations. Patterns usually imply two concepts as similarities and dissimilarities. Profiles, specifications, identifications and unique properties can be determined by dissimilarities or divergence. In addition, divergences can be play a key role for determination of early warning signals for anomalies, errors and fraud. Therefore, divergence is not one of the concerns of only statistics but data mining too. Yet there is also more powerful reason for divergence of the existing statistical estimations which are discrepancies in the conceptual scaling reflecting different ideologies of perceptions of the same phenomena. Especially in relief such discrepancies project at the joint of various world view sciences and epochs sometimes containing in the same definitions and classifications and even in symbols and measures completely different indicative view and value sense. The phenomenon in question comes out visually and instructively at their most on the example of widening the gaps between statistical information and socio-economic disinformation which are in temporary crisis circumstances go beyond all sensible apogees, turning even old time lie into the definite value. What has really happened here, what causes, except sheer self-interest which give rise (this process continues) to such terrible condition of socio-economic information? From technological point of view we have to deal today what on the one hand the capacity and flows of economical information for the period of globalization (1991-2008) would increase tenfold. On the other hand – for a mentioned period the credibility and quality of published data and consequently a quality of decisions made of their bases decreased drastically. As a result some countries and world as a whole instead of expected transition from manual control to automatic one have got into manipulative management. In total of published data the volume of primary data has significantly reduced a lot of valuable and demanded information are consumed by commercial classified information and corruption, we have lost many sources and spheres of primary data, the level of data comparability had fallen, virtually the possibilities for testing them for convergence, precision and credibility are taken out of public access. Information is dominantly collecting for information itself that is why it works with reduced efficiency. The Internet is a convincing example which shows that it servicing itself and its providers for 97, 5% and only for the rest 2, 5% it is commercial network and information service. The understanding of that the people and the worlds nations need not information typologies but information itself is substituted and lost. How to turn the contemporary situation for the better, to overcome the accumulated information gaps, misbalances, obstructions and through this to provide freed space and resources for collection, processing and distribution of credible information on the basis of which one only can provide crucial improvement of total socio-economic situation in the countries to organize the real transition from today's reforms to the reforms of forthcoming efficient transformation. Simchera (2003b) is given general answer to this question in his

work. Below with use of additional information we give following concise answer to this question. There are a lot of economical estimations but they are uncoordinated and mainly doubtful and incomparable. The gaps between existing indicators (due to conceptual differences of their understanding and coverage, material divergence in prices, exchange rates etc.) reach sometimes multiple values and strike off any possibility to its wise application. All socio-economic parameters gathered and published today are pretty often characterizing mainly not real situation in the world but represent only calculation effect, and in many cases even less – that is statistical calculations errors. On the basis of such deliberately incorrect and even false data one should not make that conclusions and assumptions which are made in modern statistical science. The science loses even more in its potential due to data manipulating practices which became rather standard than exception, substitution of one parameters for the others, aberration of not only size and meaning but the sense of the phenomena in question, that is what confirm the today's crisis estimations. There is no other brunch of the science where measurements would be more contradictive, estimations divergent and results just useless for a practice, than it is in the economic science. This is why it is logical not accidental that much of published data should be taken out of 'circulation' as useless. Consequently one should not understand all of the above as total result which shows that with this level of divergence, incorrectness there is less point to calculate and publish statistical data than not to. This fully concerns to calculations and publications of market statistical data which not only distort but also misrepresent all normal perceptions about real situation in modern economy, and virtually turn all published estimations into economical phantoms and threaten of credibility loss to all conclusions and values of modern science. Subjective evaluations, interpretations and conclusions generally can be very far away from what data tells us objectively. In addition, the data can be directed, dirty, missing, unstandardized and uncomparable. Therefore, there are some necessary steps for acceptable decision making process: 1. Data must be realiable. 2. Analysis must be suitable for the aim (s) and the data. 3. The evaluation of the results of the analysis must be correctly 4. Conclusions must be in an objective manner In case of missing at least one of the necessary steps given above, then, the gap between estimations and the realizations will be statistically or scientifically unacceptable.

### **The Gap Between Estimations And Realizations With Numerical Examples**

Here are examples illustrating gaps in economical estimations, which directly point to the sources of bad informational situation in the world and from our point of view due to this fundamental reason – situation of modern economical science. According to official statistics GNP growth in the USA in constant prices

of 2000 (with regard to statistical calculation errors) in 2007 in comparison with 2000 accounted for 1,18 times (increase 2,3% per year); (in 1991-1999 – 1,18) for the period 1981-2007 – it would be 1,93 points (2,4% per year); national wealth's estimation of growth are 1,16; 1,18 and 1,72 points respectively. These are actual estimations of growth. And here are the figures – phantoms which substitute the quite humble figures against their background of real rates of the USA economical growth. The growth of the wide spreading Dow Jones Index (share rates of 30 largest companies in the world) for the same period accounted for 1,22; 3,8 and 7,1 times (the points fixed for this index at the end of 1990 were 2892, and at the end of 1999 – 10 991; 2007 – 13 368; June 2009 – 8 501), the Standard and Poor's Index (500 companies) – 1,3; 2,1 and 19,0 times (the points fixed for this index at the end of 2003 – 1 109; 2007 – 1 479; June 2009 – 923) respectively, and the growth for Nasdaq (5 000 technology companies) – only from the middle of 1996 has exceeded 153 times (the points fixed for this index at the end of 2003 – 2 010; 2007 – 2 654; June 2009 – 1 796). In other G-7 countries the gaps between real rates and phantom – rates are approximately the same. In England in particular with the GNP growth for the period 1991-1998 in 1,25 times increase (for the period 1981-1998 - 1,49 times increase) the stock exchange FTSE index showed growth of 2,7 and 9,1 times. The exception was France where while GDP growth for the period 1991-1998 was 1,13 times its stock exchange CAC – 40 index showed growth of “just” 2,7 times. Even if we take estimations of GDP's of G-7 countries in current prices, the gaps between them and stock exchange increases would stay fantastic and destroy representations of any dignity. In addition, here are some others kind of phantoms. The real indicators of the USA's GDP with most favorable calculations – today would be \$14,0 (in 2003 – \$11,0; 1990 – \$5,8) bill. doll. Indicators of national wealth are \$28,3, \$24,8 and \$14,8 bill. doll. respectively; considering real financial assets \$39,5, \$33,2 and \$22,4 bill. Doll (U.S. Census Bureau, 2009). The capitalization value of USA's companies (price multiplied by quantity of issued shares) on financial markets estimated, before crisis 2008, for \$200-\$215 trill., annual turnover of shares – \$100-\$120 trill. The crisis only for a last half of 2008 and first half of 2009 has devaluated these assets to 40-50% and yet it is going to devaluate them to 30-35% in forthcoming months, equalizing them with estimations of real assets, which turns all fund market in the USA into a phantom. The phantoms are also modern indicators for banks assets. The real assets for 1000 largest banks are estimated for \$20 (in 2003 - \$12,7) trill. The derivatives (fictitious capital) of all world's banks exceed \$400 (in 2003 - \$100) trill., while their equity capitals account for only \$30-\$25 and \$19,5-\$18,5 (in the USA -\$14-\$10) trill. respectively. On the assumptions of estimations of U.S. Federal Deposit Insurance Corporation (2007) consolidated real assets of commercial banks (at the end of 2007 – 7 282 banks in the USA), investment institutions or investment banks (total number 1 251 at the end of 2007) and credit unions (8 101 at the end of 2007) excluding mutual benefit societies which would not exceed \$14 till. (to be precise \$13.792,5 bill.) in

the pre-crisis year, including assets of commercial banks - \$11 176,5 bill., and credit unions - \$753,4 bill. (National Credit Union Administration, 2007). And these quite credible estimates in comparison with analogues estimates on assets of some separately taken American banks are also appearing to be phantoms. The allied assets of JP Morgan accounted for \$97,5 trill. (62,5% over the world's GDP), while its real equity capital accounted for \$2,5 trill. (2,6% of the assets). Respectively estimations of Goldman & Sachs are \$50 trill. (83,3% of world's GDP) and \$1,5 trill (3% of assets). Even the assets of largest bank holding HSBC (Hong Kong-Shanghai Bank, Corp.) are phantoms, the fictitious assets of which (in 2000 accounted for \$6,5, and in 2008 - \$108 trill.) exceeds the equity capital for almost 10-50 times more (Bank of Russia, 1999). According to market rates virtually all modern transnational companies including American pride – Ford and biggest world's bank – HSBC which real capitalization in good times would exceed trillion of US dollars – are bankrupts, whilst according to leading rating agencies they still have higher ratings than actually solvent companies in China, Brazil, India or Russia. Against the background of real GDP's volume accounted for about \$60 trill. in 2008, it is absurd how estimates of fictitious capital look like, the guaranteed part of which in the same year would overrun  $6 \cdot 10^{14}$  that is \$600 trill., while unguaranteed part  $3 \cdot 10^{15}$  that is \$3 quadrillion, including the USA with \$175 and \$900 trill. At last the phantoms are the inadmissibly different by module, scope and contents indicators of GDP, inflation, national wealth in various countries. Finally, it is just ridiculous how in the form of phantoms – marginal's are presented today the audition estimations for capitalization of world's leading companies which have fallen drastically at once to 7-10% from the initial value following the world's crisis. The mixing of own capitals with attracted ones, real values with fictitious values, the parameters mentioned above and many other parameters of financial assets cause quite serious concerns and according to specialists' opinion require not only cardinal revision but also international legislative control. The above examples illustrate not only destructive but also demoralizing role of market representations about fundamentals of modern economical life and processes of economic developments of different countries and nations. They cause apprehension in a whole world and persuasively point at necessity of fundamental revision of all modern practices of economical estimations, and a new approach to the principals of economic measurements. Especially loud voices for this decision are coming out from the USA which is a motherland of financial pyramids and financial bubbles the burst of which can destroy not only the US but the rest of contemporary financial and economical system. When Alan Greenspan was speaking at the White House on 5th of April, that were unheard appeals to study statistics of world's financial markets, scrutinize and reveal the consequences of the growth of financial bubbles in the world's economy. The way out from this situation one can see in revision of core structure of a whole variety of market expectations their conventional agreements and standardization of their methods for comparative calculations. For provision of



practical realization of this way out the authorities should preliminary to take concrete solutions on the following basis:

- on the basis of the international standards of calculation and account one should organize the inventory and fundamental revaluation of all international and national assets and on this base one should take all fictitious assets out of balance turnover which dilute real representations in economy and excludes in principle the possibilities to coordinate it on a fair basis;
- instead of floating exchange rates one should introduce fixed ones which reflect real interaction of prices by all line of produced and consumed products, services and capitals;
- one should find corporative indices and fund market ratings as useless and cancel them due to they distort situation with global market estimations;
- on the principally new basis should be constructed and introduced into international turnover the single world's indices, which would representatively reflect the dynamics of world's fund and foreign exchange rates markets;
- on the basis of transparency and free accessibility, public watch and personal responsibility, with compensation of damage and lost profits one should reconsider the principles of IMF and World Bank work, and in case of failure to implement the liabilities – one should develop conditions to seize their activity as the international institutions – regulators of the international financial relationships;
- under the auspices of UN one should establish an International commission on causes qualifications and regulation of world's financial crisis consequences with the functions to determine a size of material loss and moral damage caused by the monetary authorities of the US and other countries – satellites due to failure to take the necessary measures for prevention of their negative consequences and rights to discover in accordance with international law all persons guilty for breaking the rules of financial business leading to a global default, mass bankruptcies which exceed all calculated losses of almost every country and nation in all world wars. The presented examples – are documentary approved facts of the informational distortion. However in today's practice the economy has to deal not with distorted information but with outspoken call to world's economical society what is in legal parlance called fraud, in the form of forgery, which is done by betrayal of trust – the deeds which are qualified as crime and penalties are severe. This could be a decision.

## **Conclusion and Future Studies**

To overcome the differences in modern economic conditions and eliminate the existing unjustified in many cases multiple and even by a factor of ten divergences in economical estimations which distort economical representation

and excluding possibilities of their efficient practical use in the development of the international standards one should organize and implement special work on the correction of fundamental divergences in the economic estimations, restitution of incomparable parameters, standardization of methods of their calculations and publications on comparable basis and degrees of credibility claimed beforehand. In other words one should create a single number of standards of economical information and economical indicators based on total conventional decisions. All information in the world, all economical or rather all other indicators must be built and published on this basis. There is no country which must not and cannot be exception unless it claims to be included into the existing borders of the single world socio-economic space. The conditions and consequences of its activity and life first of all should be transparent and explicit in estimations. A “non-transparent” country cannot and has no moral rights to claim transparency from other countries and first of all from its own citizens. In this connection a total international standardization for all multiplicity of mutually demanded socio-economical parameters shall not mean the automatic unification of every possible other multiplicities, which are characterizing an objectively existing national flavour and specificity of development for every single country. On the contrary the modern standardization should initially take into account the necessity for an existence, maintenance and multiplication of various numbers of national indicators considering them as fundamental origins and solid ground for their own development and existence. That is why what has been done today in the sphere of international standardization of national indicators is only the first level not the top of the enormous work which must be done in every single country and in a whole world in behave of increase of their convergence and guaranteed provision for that level of the mutual subject understanding in the world which is only possible on the basis of the international standardization of the national indicators. Thus it is rather proposed than excluded that all preceding work of the International Statistical Institute (ISI), UN’s League of Nations and other international organizations which has been made during XIX-XX centuries in the field of international standardization with regard to the stated position should be reconsidered and all used in the modern world international standards should be re-standardized. Perhaps against the background of the statistical standards flaws revealed by a crisis the international organizations should initiate and perform this very work. And data mining can be the main tool for this work. The big picture given above shows the need of clean, comparable and standardized definitions instead of directed ones for acceptable estimations and reliable conclusions. Clean, comparable data and standardized definitions lead us the datawarehouse logic or data mining infrastructure or basis. Indicators, proportions, ratios, rates and the other kind of statistics can be the first signals in case of evaluating them in a proper way. We should note that all the raw data, processed data must be clear, transparent and has equal definition for reliable comparison. Therefore, existing data stock can be collect with the same definition and standarts. Actually, UN has standard definitions for indicators, methodologies

and data preparation. Eurostat (European Union Statistical Office) has standards as well for European Union member and candidate countries. Furthermore, Eurostat statistical standards are mostly a subset of UN standards and widely there are no contradiction between UN and Eurostat standarts. Therefore, it is easy to merge UN and Eurostat data with some little justifications. UN country data infrastructure can become the base for unified statistical indicators datawarehouse. Coordinator can be the ISI as mentioned above for much more updated scientifically standards. Furthermore, practice of counteraction against legalization of statistical divergence can be use for occupational objectivity.

Therefore, practice of disclosure, critical assessment and correction of statistical divergence being revealed. The effective tools of counteraction are law on legal responsibility and legal protection for statisticians (currently not in use).

*STATISTICS IN TRANSITION-new series, December 2011*  
*Vol. 12, No. 3, pp. 579—586*

## **A CLASS OF REGRESSION TYPE ESTIMATORS IN SURVEY SAMPLING**

**Govind Charan Misra<sup>1</sup>, Subhash Kumar Yadav<sup>2</sup>, Alok Kumar Shukla<sup>1</sup>**

### **ABSTRACT**

A class of linear regression models has been proposed for the estimation of population mean and total when information regarding auxiliary variate is available in survey sampling using regression method of estimation by introducing a new auxiliary variable  $z$ , which may also be a function of the auxiliary variable  $x$ . The proposed model leads to reduction in mean squared error as compared to ordinary regression method of estimation. The improvement has been demonstrated over ordinary regression estimator and also on ratio estimator with the help of an empirical example.

**Key words:** Auxiliary variable, Mean Squared Error, Ratio estimator, Regression type estimators.

### **1. Introduction**

The intelligent use of auxiliary information for improving precision of the estimates has been done in sampling theory for different purposes. The auxiliary information has been used for the purposes of stratification in stratified sampling. In PPS (Probability Proportional to Size) sampling, the probabilities of selection of units are based on auxiliary information on measures of sizes. In ratio and regression methods of estimation, one uses auxiliary information so as to improve precision of the estimates of population parameters like population mean and population total etc. For detailed discussions on use of auxiliary information in sample surveys, reference can also be made of Sampath (2005) and Cochran (1999).

---

<sup>1</sup>Dept of Statistics, D.A-V College, Kanpur, India, e-mail: drmisragovind@gmail.com.

<sup>2</sup>Dept of Mathematics & Statistics, Dr R M L Avadh University, Faizabad, India, e-mail: drskystats@gmail.com.

In the regression method of estimation, the auxiliary variable  $X$  is correlated with the variable of interest  $Y$  and line of regression of  $Y$  on  $X$  does not pass through origin. The linear regression estimate of population mean of variable  $Y$  is defined as:

$$\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x}) \quad (1)$$

Where  $\bar{y}$  and  $\bar{x}$  are sample mean of variable  $Y$  and  $X$  respectively.  $\bar{X}$  is population mean of auxiliary variable  $X$  and is supposed to be known.  $\bar{y}_{lr}$  is linear regression estimate of population mean and  $b$  is a constant Sukhatme *et al.* (1984) have described in detail, the procedures for deriving estimates of population parameters along with their biases, mean square error etc.

## 2. Proposed model

Motivated by Ekpenyong *et al.* (2008), we are proposing a class of linear regression type estimator by including a linear term in the ordinary linear regression estimator, which includes the estimators proposed by Ekpenyong *et al.* (2008) and Misra *et al.* (2009) as special cases. The proposed model seeks to consider following relationship among variable  $Y$ , auxiliary variables  $X$  and  $Z$ .

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + U \quad (2)$$

Where  $\beta_0, \beta_1$  and  $\beta_2$  are parameters which appears linearly in the model (2).  $Z$  is another auxiliary variable which may also be taken as the function of  $X$ . When  $Z = X^2$ , it assumes the relationship considered by Ekpenyong *et al.* (2008). If  $Z = \frac{1}{X}$ , it takes the form of Misra *et al.* (2009). It has been shown by Misra *et al.* (2009) that their estimators of population mean and total are more efficient as compared to estimators of Ekpenyong *et al.* (2008) and ordinary linear regression estimator.  $U$  is independently and identically distributed random variable with mean zero and fixed variance  $\sigma^2$ .

The estimator of population mean based on (2) is given by:

$$\bar{y}_{gl} = \bar{y} - \hat{\beta}_1(\bar{x} - \bar{X}) - \hat{\beta}_2(\bar{z} - \bar{Z}) + U \quad (3)$$

Where  $\bar{y}_{gl}$  is a general regression type estimator of population mean based on proposed model defined in relation (2),  $\bar{z}$  and  $\bar{Z}$  are sample and population means of the variable  $Z$ .

### 3. Estimation of bias and variance of $\bar{y}_{gl}$

The bias and variance of  $\bar{y}_{gl}$ , the general regression type estimator of population mean can be calculated in following manner.

$$\text{Let } \bar{y} = \bar{Y}(1+e_0), \quad \bar{x} = \bar{X}(1+e_1), \quad \hat{\beta}_1 = \beta_1(1+e_2) \quad , \quad \hat{\beta}_2 = \beta_2(1+e_3) \quad \text{and} \\ \bar{z} = \bar{Z}(1+e_4)$$

Such that  $E(e_i) = 0 \quad \forall i = 0, 1, 2, 3, 4$  putting these values in equation (3), we get

$$\begin{aligned} \bar{y}_{gl} &= \bar{Y}(1+e_0) - \beta_1(1+e_2)[\bar{X}(1+e_1) - \bar{X}] - \beta_2(1+e_3)[\bar{Z}(1+e_4) - \bar{Z}] \\ \bar{y}_{gl} &= \bar{Y}(1+e_0) - \beta_1(1+e_2)[\bar{X}e_1] - \beta_2(1+e_3)[\bar{Z}e_4] \\ \bar{y}_{gl} &= \bar{Y} + \bar{Y}e_0 - \beta_1(\bar{X}e_1 + \bar{X}e_1e_2) - \beta_2(\bar{Z}e_4 + \bar{Z}e_3e_4) \\ \bar{y}_{gl} &= \bar{Y} + \bar{Y}e_0 - \beta_1\bar{X}(e_1 + e_1e_2) - \beta_2\bar{Z}(e_4 + e_3e_4) \end{aligned}$$

Taking expectation on both the sides, we get

$$\begin{aligned} E(\bar{y}_{gl} - \bar{Y}) &= \bar{Y}E(e_0) - \beta_1\bar{X}[E(e_1) + E(e_1e_2)] - \beta_2\bar{Z}[E(e_4) + E(e_3e_4)] \\ &= -\beta_1\bar{X} E(e_1e_2) - \beta_2\bar{Z} E(e_3e_4) \end{aligned}$$

$$\text{Bias}(\bar{y}_{gl}) = -Cov(\bar{x}, \hat{\beta}_1) - Cov(\bar{z}, \hat{\beta}_2)$$

This is negligible for large sample size. For large samples usually  $Cov(\bar{x}, \hat{\beta}_1)$  decreases and it becomes zero if the joint distribution of  $y$  and  $x$  is bivariate normal. Similarly  $Cov(\bar{z}, \hat{\beta}_2)$  vanishes if the joint distribution of  $y$  and  $z$  follows bivariate normal distribution. In this case the proposed regression estimator is exactly unbiased.

To the first order of approximation, by ignoring the terms with  $e_i e_j (i \neq j) = 0, 1, 2, 3, 4$ , we have

$$\bar{y}_{gl} - \bar{Y} = e_0\bar{Y} - e_1\beta_1\bar{X} - e_4\beta_2\bar{Z}$$

Therefore

$$\begin{aligned}
 V(\bar{y}_{gl}) &= V(e_0\bar{Y} - e_1\beta_1\bar{X} - e_4\beta_2\bar{Z}) \\
 &= V(e_0\bar{Y}) + \beta_1^2 V(e_1\bar{X}) + \beta_2^2 V(e_4\bar{Z}) - 2\beta_1 \text{Cov}(e_0\bar{Y}, e_1\bar{X}) \\
 &\quad - 2\beta_2 \text{Cov}(e_0\bar{Y}, e_4\bar{Z}) + 2\beta_1\beta_2 \text{Cov}(e_1\bar{X}, e_4\bar{Z}) \\
 &= V(\bar{y}) + \beta_1^2 V(\bar{x}) + \beta_2^2 V(\bar{z}) - 2\beta_1 \text{Cov}(\bar{y}, \bar{x}) \\
 &\quad - 2\beta_2 \text{Cov}(\bar{y}, \bar{z}) + 2\beta_1\beta_2 \text{Cov}(\bar{x}, \bar{z})
 \end{aligned} \tag{4}$$

Therefore we get,

$$\begin{aligned}
 V(\bar{y}_{gl}) &= \lambda \left[ s_y^2 - 2\beta_1 s_{xy} + \beta_1^2 s_x^2 - 2\beta_2 s_{yz} + \beta_2^2 s_z^2 + 2\beta_1\beta_2 s_{xz} \right] \\
 \text{where } \lambda &= \left( \frac{1}{n} - \frac{1}{N} \right)
 \end{aligned} \tag{5}$$

Here  $s_{xy}$ ,  $s_{yz}$  and  $s_{xz}$  are estimators of the population covariances,  $S_{XY}$ ,  $S_{YZ}$  and  $S_{ZX}$  respectively, while variances  $s_x^2$ ,  $s_y^2$  and  $s_z^2$  are unbiased estimators of population variances  $S_x^2$ ,  $S_y^2$  and  $S_z^2$  respectively.

We need to estimate  $\beta_1$  and  $\beta_2$  such that  $V(\bar{y}_{gl})$  is a minimum. Using the method of ordinary least square, we differentiate partially (4) with respect to  $\hat{\beta}_1$  and  $\hat{\beta}_2$  and obtain following normal equations.

$$\hat{\beta}_2 \text{Cov}(\bar{x}, \bar{z}) + \hat{\beta}_1 V(\bar{x}) = \text{Cov}(\bar{y}, \bar{x}) \tag{6}$$

$$\hat{\beta}_2 V(\bar{z}) + \hat{\beta}_1 \text{Cov}(\bar{x}, \bar{z}) = \text{Cov}(\bar{y}, \bar{z}) \tag{7}$$

Solving (6) and (7) simultaneously, we obtain

$$\hat{\beta}_1 = \frac{\text{Cov}(\bar{y}, \bar{z})\text{Cov}(\bar{x}, \bar{z}) - \text{Cov}(\bar{y}, \bar{x})V(\bar{z})}{[\text{Cov}(\bar{x}, \bar{z})]^2 - V(\bar{x})V(\bar{z})}$$

$$\hat{\beta}_2 = \frac{\text{Cov}(\bar{y}, \bar{x})\text{Cov}(\bar{x}, \bar{z}) - \text{Cov}(\bar{y}, \bar{z})V(\bar{x})}{[\text{Cov}(\bar{x}, \bar{z})]^2 - V(\bar{x})V(\bar{z})}$$

And the expressions for variance of ordinary linear regression estimator and ratio estimator are

$$V(\bar{y}_{lr}) = \lambda \left[ s_y^2 - 2\beta_1 s_{xy} + \beta_1^2 s_x^2 \right] \tag{8}$$

and  $V(\bar{y}_R) = \lambda \left[ s_y^2 + R^2 s_x^2 - 2R s_{xy} \right]$  (9)

where  $R = \frac{\bar{y}}{\bar{x}}$

The estimate of population total ( $y_{gl}$ ) and its variance using proposed estimator  $\bar{y}_{gl}$ , are as follows

$$y_{gl} = N \bar{y}_{gl}$$

$$V(y_{gl}) = N^2 V(\bar{y}_{gl})$$

#### 4. Comparison with ordinary regression estimator

Now we shall compare the variances of ordinary regression estimator and proposed regression estimator and it is found that proposed regression estimator is more efficient as compare to ordinary regression estimator. Using relations (5) and (8), we have

$$V(\bar{y}_{lr}) - V(\bar{y}_{gl}) = \lambda \left[ 2\hat{\beta}_2 s_{yz} - \hat{\beta}_2^2 s_z^2 - 2\hat{\beta}_1 \hat{\beta}_2 s_{zx} \right]$$

Now

$$\begin{aligned} & 2\hat{\beta}_2 s_{yz} - \hat{\beta}_2^2 s_z^2 - 2\hat{\beta}_1 \hat{\beta}_2 s_{zx} \\ &= \hat{\beta}_2 (2s_{yz} - \hat{\beta}_2 s_z^2 - 2\hat{\beta}_1 s_{zx}) \\ &= \frac{(s_{yx} s_{zx} - s_{yz} s_x^2)}{(s_{xz}^2 - s_x^2 s_z^2)} \left[ 2s_{yz} - \frac{(s_{yx} s_{zx} - s_{yz} s_x^2)}{(s_{xz}^2 - s_x^2 s_z^2)} s_z^2 - 2 \frac{(s_{yz} s_{zx} - s_{yx} s_z^2)}{(s_{xz}^2 - s_x^2 s_z^2)} s_{zx} \right] \\ &= \frac{(s_{yx} s_{zx} - s_{yz} s_x^2)}{(s_{xz}^2 - s_x^2 s_z^2)^2} \left[ 2s_{yz} (s_{xz}^2 - s_x^2 s_z^2) - (s_{yx} s_{zx} - s_{yz} s_x^2) s_z^2 - 2(s_{yz} s_{zx} - s_{yx} s_z^2) s_{zx} \right] \\ &= \frac{(s_{yx} s_{zx} - s_{yz} s_x^2)}{(s_{xz}^2 - s_x^2 s_z^2)^2} \left[ 2s_{yz} s_{xz}^2 - 2s_{yz} s_x^2 s_z^2 - s_{yx} s_{zx} s_z^2 + s_{yz} s_x^2 s_z^2 - 2s_{yz} s_{zx} s_{zx} + 2s_{yx} s_z^2 s_{zx} \right] \end{aligned}$$



$$\begin{aligned}
 &= \frac{(s_{yx}s_{zx} - s_{yz}s_x^2)}{(s_{xz}^2 - s_x^2s_z^2)} \left[ -s_{yz}s_x^2s_z^2 + s_{yx}s_{zx}s_z^2 \right] \\
 &= \frac{(s_{yx}s_{zx} - s_{yz}s_x^2)}{(s_{xz}^2 - s_x^2s_z^2)} \left[ s_{yx}s_{zx} - s_{yz}s_x^2 \right] s_z^2 \\
 &= \frac{(s_{yx}s_{zx} - s_{yz}s_x^2)^2 s_z^2}{(s_{xz}^2 - s_x^2s_z^2)^2} > 0 \\
 \Rightarrow V(\bar{y}_{lr}) - V(\bar{y}_{gl}) > 0
 \end{aligned}$$

Which shows that the estimator  $\bar{y}_{gl}$  is more efficient as compared to estimator  $\bar{y}_{lr}$ , as it has lesser mean squared error. We shall demonstrate this result with the help of following example in which  $Z$  has been considered as a function of  $X$ .

### 5. Numerical example

**Example:** The following table describes the estimates population mean and population total and their variances after ignoring f.p.c. using different models, the data given in Des Raj (1972), page 89. The size of the population has been considered as 30 and a random sample without replacement of size 8 has been drawn from it. The observations corresponding to sample numbers 12, 02,22,21,03,08,10,07 gave following results.

Estimates Methods	Mean ( $\bar{y}$ )	$V(\bar{y})$	Total ( $y$ )	$V(y)$
Model ( $z = 1/x^2$ )	4.0308	0.1302	120.9265	117.1428
Model ( $z = 1/x$ )	4.3036	0.1378	129.1102	124.0976
Model ( $z = \sqrt{x}$ )	4.0598	0.1551	121.7963	139.6262
Model ( $z = x^2$ )	4.3451	0.1663	130.3527	149.7388
Linear Regression	4.2133	0.1836	126.4014	165.3044
Ratio Estimator	3.1375	0.9636	94.1250	867.2074

It is observed that estimates of population mean and population total obtained from  $\bar{y}_{gl}$  are more efficient as compared to estimates of population mean and population total obtained from  $\bar{y}_{lr}$ , and also ratio estimator.

## 6. Conclusions

The inclusion of a linear term in ordinary linear regression estimator improves the precision of the estimates of population parameters such as population mean and population total etc. The proposed model is in general form, which includes the models of Ekpenyong *et al.* (2008) and also Misra *et al.* (2009) as particular cases. It has been shown with the help of an example that the proposed model provides more precise estimates of population mean and population total as compared to linear regression estimator as well as ratio estimator of population mean and population total. The inclusion of additional linear term has considered four cases in which  $z$ , a function of  $x$  has been considered as  $\frac{1}{x^2}$ ,  $\frac{1}{x}$ ,  $\sqrt{x}$  and  $x^2$ . The second and fourth values of  $z$  correspond to Misra *et al.* (2009) and Ekpenyong *et al.* (2008) respectively. In all these four cases of  $z$ , the estimates of population mean and total are more efficient as compared to traditional linear regression estimator and traditional ratio estimator.

## Acknowledgement

Authors are very thankful to the referees for their valuable suggestions in improving this manuscript and the financial assistance has been given by University Grant Commission, New Delhi, India under Major Research Project F.No.33-56/2007 (SR).

## REFERENCES

- COCHRAN, W.G. (1999). Sampling Techniques, John Wiley & Sons.
- DES RAJ (1972). The design of sampling surveys, McGraw-Hill, New York.
- EKPENYONG, E. J., OKONNAH, M.I. and JOHN, E.D. (2008), Polynomial (Non-linear) Regression.
- Method for Improved Estimation Based on Sampling, Journal of Applied Sciences, 8(8), 1597-1599.
- MISRA, G.C., SHUKLA, A.K. AND YADAV, S.K. (2009). A Comparison of Regression Methods for Improved Estimation in Sampling, Journal of reliability and statistical studies, Vol. 2, Issue2, 85-90.

SUKHATME, P.V., SUKHATME, B.V., SUKHATME, S. and ASOK, C. (1984). Sampling Theory of Surveys with Applications, Indian Society of Agricultural Statistics. Sampath, S. (2005). Sampling Theory and Methods, Narosa Publishing House, India.

## A GENERAL FAMILY OF DUAL TO RATIO-CUM-PRODUCT ESTIMATOR IN SAMPLE SURVEYS

Rajesh Singh, Mukesh Kumar<sup>1</sup>, Pankaj Chauhan, Nirmala Sawan<sup>2</sup>,  
Florentin Smarandache<sup>3</sup>

### ABSTRACT

This paper presents a family of dual to ratio-cum-product estimators for the finite population mean. Under simple random sampling without replacement (SRSWOR) scheme, expressions of the bias and mean-squared error (MSE) up to the first order of approximation are derived. We show that the proposed family is more efficient than usual unbiased estimator, ratio estimator, product estimator, Singh estimator (1967), Srivenkataramana (1980) and Bandyopadhyaya estimator (1980) and Singh et al. (2005) estimator. An empirical study is carried out to illustrate the performance of the constructed estimator over others.

**Key words:** Family of estimators, auxiliary variables, bias, mean-squared error.

### 1. Introduction

It is common practice to use the auxiliary variable for improving the precision of the estimate of a parameter. Out of many ratio and product methods of estimation are good examples in this context. When the correlation between the study variate and auxiliary variates is positive (high), ratio method of estimation is quite effective. On the other hand, when this correlation is negative (high), product method of estimation can be employed effectively. Let  $U$  be a finite population consisting of  $N$  units  $U_1, U_2, \dots, U_N$ . Let  $y$  and  $(x, z)$  denote the study variate and auxiliary variates taking the values  $y_i$  and  $(x_i, z_i)$ , respectively, on the unit  $U_i$  ( $i=1, 2, \dots, N$ ), where  $x$  is positively correlated with  $y$  and  $z$  is negatively correlated with  $y$ . We wish to estimate the population mean  $\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i$  of  $y$ ,

<sup>1</sup> Department of Statistics, BHU, Varanasi (U.P.), India (rsinghstat@yahoo.com).

<sup>2</sup> School of Statistics, DAVV, Indore (M.P.), India.

<sup>3</sup> Department of Mathematics, University of New Mexico, Gallup, USA, (smarand@unm.edu).

assuming that the population means  $(\bar{X}, \bar{Z})$  of  $(x, z)$  are known. Assume that a simple random sample of size  $n$  is drawn without replacement from  $U$ . The classical ratio and product estimators for estimating  $\bar{Y}$  are:

$$\bar{y}_R = \frac{\bar{y}}{\bar{X}} \bar{X} \quad (1.1)$$

and

$$\bar{y}_P = \bar{y} \frac{\bar{Z}}{\bar{Z}} \quad (1.2)$$

respectively, where  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ ,  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  and  $\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$  are the sample means of  $y$ ,  $x$  and  $z$  respectively.

Using the transformation  $x_i^* = \frac{(N\bar{X} - nx_i)}{(N-n)}$ , and  $z_i^* = \frac{(N\bar{Z} - nz_i)}{(N-n)}$ , ( $i=1,2,\dots,N$ )

Srivenkataramana (1980) and Bandyopadhyaya (1980) suggested a dual to ratio and product estimator as:

$$\bar{y}_R^* = \bar{y} \frac{\bar{x}^*}{\bar{X}} \quad (1.3)$$

and  $\bar{y}_P^* = \bar{y} \frac{\bar{Z}}{\bar{z}^*} \quad (1.4)$

where  $\bar{x}^* = \frac{N\bar{X} - n\bar{x}}{(N-n)}$  and  $\bar{z}^* = \frac{N\bar{Z} - n\bar{z}}{(N-n)}$ .

In some survey situations, information on a secondary auxiliary variate  $z$ , correlated negatively with the study variate  $y$ , is readily available. Let  $\bar{Z}$  be the known population mean of  $z$ . For estimating  $\bar{Y}$ , Singh (1967) considered a ratio-cum-product estimator

$$\bar{y}_{RP} = \bar{y} \left( \frac{\bar{X}}{\bar{x}} \right) \left( \frac{\bar{Z}}{\bar{z}} \right) \quad (1.5)$$

Using a simple transformation  $x_i^* = (1+g)\bar{X} - gx_i$  and  $z_i^* = (1+g)\bar{Z} - gz_i$ ,  $i=1,2,\dots,N$ , with  $g = \frac{n}{(N-n)}$ , Singh et al. (2005) proposed a dual to usual ratio-cum-product estimator

$$\bar{y}_{RP}^* = \bar{y} \left( \frac{\bar{x}^*}{\bar{X}} \right) \left( \frac{\bar{Z}}{\bar{z}^*} \right) \tag{1.6}$$

where  $\bar{x}^* = (1 + g)\bar{X} - g\bar{x}$  and  $\bar{z}^* = (1 + g)\bar{Z} - g\bar{z}$ .

To the first degree of approximation

$$V(\bar{y}) = \theta \bar{Y} C_y^2 \tag{1.7}$$

$$MSE(\bar{y}_R) = \theta \bar{Y}^2 [C_y^2 + C_x^2 (1 - 2K_{yx})] \tag{1.8}$$

$$MSE(\bar{y}_P) = \theta \bar{Y}^2 [C_y^2 + C_x^2 (1 + 2K_{yz})] \tag{1.9}$$

$$MSE(\bar{y}_{RP}) = \theta \bar{Y}^2 [C_y^2 + C_x^2 (1 + 2K_{yz}) + C_x^2 (1 - 2K)] \tag{1.10}$$

$$MSE(\bar{y}_R^*) = \theta \bar{Y}^2 [C_y^2 + gC_x^2 (g - 2K_{yx})] \tag{1.11}$$

$$MSE(\bar{y}_P^*) = \theta \bar{Y}^2 [C_y^2 + gC_z^2 (g + 2K_{yz})] \tag{1.12}$$

$$MSE(\bar{y}_{RP}^*) = \theta \bar{Y}^2 [C_y^2 + gC_z^2 (g + 2K_{yz}) + gC_x^2 (g - 2gK_{zx} - 2K_{yx})] \tag{1.13}$$

where MSE (.) stands for mean square error (MSE) of (.).

$$\theta = \frac{1-f}{n}, f = \frac{n}{N}, C_y = \frac{S_y}{\bar{Y}}, C_x = \frac{S_x}{\bar{X}}, C_z = \frac{S_z}{\bar{Z}}, K_{yx} = \rho_{yx} \frac{C_y}{C_x}, K_{yz} = \rho_{yz} \frac{C_y}{C_z},$$

$$K_{zx} = \rho_{zx} \frac{C_z}{C_x}, K = K_{yx} + K_{zx}, \rho_{yx} = \frac{S_{yx}}{S_y S_x}, \rho_{yz} = \frac{S_{yz}}{S_y S_z}, \rho_{zx} = \frac{S_{zx}}{S_x S_z},$$

$$S_y^2 = \frac{1}{(N-1)} \sum_{i=1}^N (y_i - \bar{Y})^2, S_x^2 = \frac{1}{(N-1)} \sum_{i=1}^N (x_i - \bar{X})^2, S_z^2 = \frac{1}{(N-1)} \sum_{i=1}^N (z_i - \bar{Z})^2,$$

$$S_{yx} = \frac{1}{(N-1)} \sum_{i=1}^N (y_i - \bar{Y})(x_i - \bar{X}), S_{yz} = \frac{1}{(N-1)} \sum_{i=1}^N (y_i - \bar{Y})(z_i - \bar{Z}),$$

$$\text{and } S_{zx} = \frac{1}{(N-1)} \sum_{i=1}^N (x_i - \bar{X})(z_i - \bar{Z}).$$

In this paper, under SRSWOR, we present a family of dual to ratio-cum-product estimator for estimating the population mean  $\bar{Y}$ . We obtain the first order approximation of the bias and the MSE for this family of estimators. Numerical illustrations are given to show the performance of the constructed estimator over other estimators.

## 2. The suggested family of estimators

We define a family of estimators of  $\bar{Y}$  as

$$t = \bar{y} \left( \frac{\bar{x}^*}{\bar{X}} \right)^{\alpha_1} \left( \frac{\bar{Z}}{\bar{z}^*} \right)^{\alpha_2} \quad (2.1)$$

where  $\alpha_i$ 's ( $i=1,2$ ) are unknown constants to be suitably determined.

To obtain the bias and MSE of  $t$ , we write

$$\bar{y} = \bar{Y}(1 + e_0), \quad \bar{x} = \bar{X}(1 + e_1), \quad \bar{z} = \bar{Z}(1 + e_2)$$

such that

$$E(e_0) = E(e_1) = E(e_2) = 0$$

and

$$E(e_0^2) = \theta C_y^2, \quad E(e_1^2) = \theta C_x^2, \quad E(e_2^2) = \theta C_z^2,$$

$$E(e_0 e_1) = \theta \rho_{xy} C_y C_x, \quad E(e_0 e_2) = \theta \rho_{yz} C_y C_z, \quad E(e_1 e_2) = \theta \rho_{xz} C_x C_z.$$

Expressing  $t$  in terms of  $e$ 's, we have

$$t = \bar{Y}(1 + e_0)(1 - ge_1)^{\alpha_1} (1 - ge_2)^{-\alpha_2} \quad (2.2)$$

We assume that  $|ge_1| < 1$ ,  $|ge_2| < 1$ , so that  $(1 - ge_1)^{\alpha_1}$  and  $(1 - ge_2)^{-\alpha_2}$  are expandable. Expanding the right hand side of (2.2) and retaining terms up to second powers of  $e$ 's (up to the first order of approximation), we have

$$\begin{aligned} t = & \bar{Y} [1 + e_0 - \alpha_1 ge_1 - \alpha_1 ge_0 e_1 + \frac{\alpha_1(\alpha_1 - 1)}{2} g^2 e_1^2 + \alpha_2 ge_2 + \alpha_2 ge_0 e_2 \\ & - \alpha_1 \alpha_2 g^2 e_1 e_2 + \frac{\alpha_2(\alpha_2 - 1)}{2} g^2 e_2^2] \end{aligned} \quad (2.3)$$

Taking expectations of both sides in (2.3) and then subtracting  $\bar{Y}$  from both sides, we get the bias of the estimator  $t$ , up to the first order of approximation, as

$$\begin{aligned} B(t) &= E(t - \bar{Y}) \\ &= \bar{Y} g \theta [(\alpha_2 K_{yz} C_z^2 - \alpha_1 K_{yx} C_x^2) + \frac{\alpha_1(\alpha_1 - 1)}{2} g C_x^2 - \alpha_2 g C_z^2 \{ \alpha_1 K_{xz} - \frac{(\alpha_2 - 1)}{2} \}] \end{aligned} \quad (2.4)$$

where  $K_{xz} = \rho_{xz} \frac{C_x}{C_z}$ .

From (2.3), we have

$$(t - \bar{Y}) \cong \bar{Y}[e_0 - g(\alpha_1 e_1 - \alpha_2 e_2)] \tag{2.5}$$

Squaring both sides of (2.5), we have

$$(t - \bar{Y})^2 = \bar{Y}^2[e_0^2 + g^2\{\alpha_1^2 e_1^2 + \alpha_2^2 e_2^2 - 2\alpha_1 \alpha_2 e_1 e_2\} - 2g(\alpha_1 e_0 e_1 - \alpha_2 e_0 e_2)] \tag{2.6}$$

Taking expectations of both sides of (2.6), we get the MSE of t to the first degree of approximation as

$$MSE(t) = \theta \bar{Y}^2 [C_y^2 + g\alpha_2 C_z^2 (g\alpha_2 + 2K_{yz}) + C_x^2 g (g\alpha_1^2 - 2\alpha_1 \alpha_2 gK_{zx} - 2\alpha_1 K_{yx})] \tag{2.7}$$

Minimization of (2.7) with respect to  $\alpha_1$  and  $\alpha_2$  yields their optimum values as

$$\left. \begin{aligned} \alpha_1 &= \frac{K_{yx} - K_{zx} K_{yz}}{g(1 - \rho_{xz}^2)} \\ \alpha_2 &= \frac{-(K_{yz} - K_{xz} K_{yx})}{g(1 - \rho_{xz}^2)} \end{aligned} \right\} \tag{2.8}$$

Substitution of (2.8) in (2.7) yields the minimum value of MSE (t) as

$$\min.MSE(t) = \theta \bar{Y}^2 C_y^2 (1 - \rho_{y.xz}^2) \tag{2.9}$$

where  $\rho_{y.xz}^2 = \frac{\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}}{(1 - \rho_{xz}^2)}$  is the multiple correlation coefficient of y on x and z.

**Remark 2.1:** For  $(\alpha_1, \alpha_2)=(1,0)$ , the estimator t reduces to the ‘dual to ratio’ estimator

$$\bar{y}_R^* = \bar{y} \left( \frac{\bar{X}^*}{\bar{X}} \right)$$

While for  $(\alpha_1, \alpha_2)=(0,1)$  it reduces to the ‘dual to product’ estimator

$$\bar{y}_P^* = \bar{y} \left( \frac{\bar{Z}}{\bar{Z}^*} \right)$$

It coincides with the estimator in Singh et al. (2005) when  $(\alpha_1, \alpha_2)=(1,1)$ .



**Remark 2.2:** The optimum values of  $\alpha_{i_0}$ 's ( $i=1,2$ ) are functions of unknown population parameters such as  $K_{yx}$ ,  $K_{yz}$ ,  $K_{yx}$ . The values of these unknown population parameters can be guessed quite accurately from the past data or experiences gathered in due course of time, for instance, see Srivastava (1967), Reddy (1973), Prasad (1989,p.391), Murthy (1967,pp96-99), Singh et. al. (2007) and Singh et. al. (2009). Also the prior values of  $K_{yx}$ ,  $K_{yz}$ ,  $K_{zx}$  may be either obtained on the basis of the information from the most recent survey or by conducting a pilot survey, see, Lui (1990,p.3805).

From (1.7) to (1.13) and (2.9) it can be shown that the proposed estimator  $t$  at (2.1) is more efficient than usual unbiased estimator  $\bar{y}$ , usual ratio estimator  $\bar{y}_R$ , product estimator  $\bar{y}_P$ , Singh (1967) ratio-cum product estimator  $\bar{y}_{RP}$ , Srivenkataramana (1980) and Bandyopadhyaya (1980) dual to ratio estimator  $\bar{y}_R^*$ , dual to product estimator  $\bar{y}_P^*$  and Singh et al. (2005) dual to ratio-cum-product estimator  $\bar{y}_{RP}^*$  at its optimum conditions.

### 3. Empirical study

In this section we illustrate the performance of the constructed estimator  $t$  over various other estimators  $\bar{y}, \bar{y}_R, \bar{y}_P, \bar{y}_{RP}, \bar{y}_R^*, \bar{y}_P^*, \bar{y}_{RP}^*$  through natural data earlier used by Singh (1969, p.377).

y: number of females employed

x: number of females in service

z: number of educated females

$$\bar{Y} = 7.46, \bar{X} = 5.31, \bar{Z} = 179.00, C_y^2 = 0.5046, C_x^2 = 0.5737, C_z^2 = 0.0633,$$

$$\rho_{yx} = 0.7737, \rho_{yz} = -0.2070, \rho_{xz} = -0.0033, N=61 \text{ and } n=20.$$

**Table 3.1:** Range of  $\alpha_1$  and  $\alpha_2$  for which  $t$  is better than  $\bar{y}_{RP}^*$

$\alpha_2$	$\alpha_1$						
	0.94	1	1.25	$1.39(\alpha_{1(opt)})$	1.5	1.7	1.85
2.7	100.81	104.67	116.46	118.52	117.4	109.7	100.6
3.1	101.01	105.02	116.8	118.9	117.8	110.13	100.9
$3.34(\alpha_{2(opt)})$	101.14	105.08	116.96	119.03	117.8	110.19	101.0
3.4	101.2	105.08	116.95	119.02	117.8	110.19	101.04
3.8	101.02	104.8	116.71	118.76	117.6	109.9	100.8
4.5	100.04	103.8	115.38	117.38	116.2	108.7	99.8
5.0	98.83	102.5	113.77	115.71	114.6	107.3	98.6

The percent relative efficiencies (PRE's) of the different estimators with respect to the proposed estimator  $t$  are computed by the formula

$$PRE(t,.) = \frac{MSE(.)}{MSE(t)} * 100$$

and presented in table 3.2.

**Table 3.2:** PRE's of various estimators of  $\bar{Y}$  with respect to  $\bar{y}$

Estimator	PRE
$\bar{y}$	100
$\bar{y}_R$	203.43
$\bar{y}_P$	123.78
$\bar{y}_{RP}$	213.64
$\bar{y}_R^*$	214.78
$\bar{y}_P^*$	104.35
$\bar{y}_{RP}^*$	235.68
$t_{opt}^*$	278.21

#### 4. Conclusion

(i) Table 3.1 shows that there is a wide scope of choosing  $\alpha_1$  and  $\alpha_2$  for which our proposed estimator 't' performs better than  $\bar{y}_{RP}^*$ .

(ii) Table 3.2 shows clearly that the proposed estimator 't' is more efficient than all other estimators  $\bar{y}, \bar{y}_R, \bar{y}_P, \bar{y}_{RP}, \bar{y}_R^*, \bar{y}_P^*$  and  $\bar{y}_{RP}^*$  with considerable gain in efficiency.

In this paper we have presented a family of dual to ratio-cum-product estimators for the finite population mean. For future research the family suggested here can be adapted to double sampling scheme using Kumar and Bahl (2006) estimator.

## REFERENCES

- BANDYOPADHYAYA, S. (1980): Improved ratio and product estimators, *Sankhya*, Series C 42,45-49.
- KUMAR, M., BAHL, S., 2006, Class of dual to ratio estimators for double sampling, *Statistical Papers*, 47, 319-326.
- LUI, K. J. (1990): Modified product estimators of finite population mean in finite sampling. *Communications in Statistics, Theory and Methods*, 19(10), 3799-3807.
- MURTHY, M. N. (1967): *Sampling Theory and Methods*. Statistical Publishing Society, Calcutta, India.
- PRASAD, B. (1989): Some improved ratio type estimators of population mean and ratio in finite population sample surveys. *Communications in Statistics, Theory and Methods*, 18, 379-392.
- REDDY, V. N. (1973): On ratio and product methods of estimation. *Sankhya*, Series B 35(3), 307-316.
- SINGH, H. P., SINGH, R., ESPEJO, M.R., PINED, D.M., NADARAJAH, S. (2005): On the efficiency of a dual to ratio-cum-product estimator in sample surveys. *Mathematical Proceedings of the Royal Irish Academy*, 105 A(2), 51-56.
- SINGH, M. P. (1967): Ratio-cum-product method of estimation. *Metrika* 12, 34-72.
- SINGH, M. P. (1969): Comparison of some ratio-cum-product estimators. *Sankhya Series B*, 31, 375-378.
- SINGH, R., CAUHAN, P., SAWAN, N., and SMARANDACHE, F. (2007): *Auxiliary Information and a Priory Values in Construction of Improved Estimators*. Renaissance High Press.
- SINGH, R., SINGH, J. and SMARANDACHE, F. (2009): *Some Studies in Statistical Inference , Sampling Techniques and Demography*. ProQuest Information & Learning.
- SRIVASTAVA, S. K. (1967): An estimator using auxiliary information. *Cal. Stat. Assoc. Bull.*, 16, 121-132.
- SRIVENKATARAMANA, T (1980): A dual to ratio estimator in sample surveys. *Biometrika* 67, 199-204.

*STATISTICS IN TRANSITION*-new series, December 2011  
Vol. 12, No. 3, pp. 595—607

## **AN IMPROVEMENT OF QUALITY OF STATISTICAL MATCHING FOR SURVEY DATA USING DYNAMIC CALIPER**

**Pawel Strawiński<sup>1</sup>**

### **ABSTRACT**

Nowadays, matching is a widely used technique to estimate program net effects. The goal of the method is to establish a counterfactual state by choosing from the control pool a group that is similar to those in the treatment group. In this article we propose a modification of the matching with caliper procedure. The novelty in our approach is setting the caliper value as a fraction of estimated propensity score. The simulation results and examples are presented. Using Dehejia and Wahba (1999) data benefits of the proposed approach are stressed. The obtained results indicate that proposed approach is more efficient than the one traditionally used.

**Key words:** matching, propensity score, caliper, evaluation.

### **1. Introduction**

The objective of many empirical works is a replication of controlled experiment via non-experimental techniques. In such circumstances quasi-experimental methods are employed. Despite that they are frequently used in program evaluations, there is no consensus in the econometric literature about its efficiency (Smith and Todd, 2005).

Evaluation of the program impact always includes speculation how one would function without a program. At every moment one could be in only one alternative state, not in both. The one can be a program participant or non-participant. If the data would be experimental data, it would be sufficient to compare the outcome for experimental and control group. Therefore, evaluation problem is a missing data problem (Heckman et al., 1997). Propensity score matching methods are increasingly being used in observational studies to reduce the impact of treatment-selection bias in the estimation of causal treatment effects (Austin, 2009).

---

<sup>1</sup> Warsaw University. Email: pstrawinski@wne.uw.edu.pl.

In the statistical and econometric literature various matching methods are described that allow for reconstruction of control group from survey, that is non-experimental data. The most widely applied technique is a propensity score matching (Rosenbaum & Rubin, 1983). Its popularity arises from the implementations in various statistical packages. The propensity score itself is a probability of being a member of the treated group, for instance a program participant. Dehejia and Wahba (1999, 2002) showed that propensity score could be used to match observations from experimental and non-experimental pool of potential controls. The role of matching procedure is to pick up from control pool a reference control group. In this way one is able to replicate the results of fully controlled experiment, particularly the outcome of the National Support Work Demonstration (NSW) experiment, and the obtained results would be slightly biased in comparison with the controlled experiment. The NSW was a temporary employment program designed to help disadvantaged workers lacking basic jobs skills move into labour market providing them work experience at subsidised positions (LaLonde, 1986). On the other hand, Smith and Todd (2005) also analysed the impact of the NSW on its participants and showed that the matching results may be sensitive to the specification of the propensity score vector and the comparison group.

Our work is located in this empirical research area. We show that with the use of Strawiński (2009) modified matching algorithm we are able to replicate more accurately the experimental results than using standard techniques. The aim of this research is not a result replication, but to point out that modified matching algorithm outperforms others that are well-established in the literature.

The remaining part of the article is organised as follows. In the next section we present the related literature. In the third section we briefly present the matching methods and explain Strawiński (2009) modification. In the fourth we present data and empirical results, while section five summarizes and concludes.

## **2. A brief review of the relevant literature**

The literature describing various approaches to matching is enormous. The seminal paper we refer to is LaLonde (1986). The author evaluated economically the NSW program with experimental and quasi-experimental techniques. In the latter the control groups were chosen from Current Population Survey (CPS) and Panel Study of Income Dynamics (PSID). Various estimation methods were employed from difference in mean values, though regression analysis, difference-in-differences to selection methods. The author showed that various estimation techniques provide different results that fundamentally differ from the experimental ones. LaLonde (1986) concludes that non-experimental methods are not efficient in replication of the experimental results and researchers should be aware of the potential specification errors in similar evaluations.

In similar work based on the same data Heckman and Hotz (1989) used wider range of estimation tools and obtained reverse results. They include in the models only pre-treatment covariates and successfully replicated the experimental results. Dehejia and Wahba (1999) also explored the same dataset. Their aim was to make a judgement about quality of the propensity score matching in the evaluation context. The authors showed that it is sufficient to use simple matching estimators to closely replicate the experimental results. This is possible even when some assumptions of matching method are not satisfied (i.e. random participation).

Smith and Todd (2005) investigated influence of the participation in NSW on beneficent and showed that propensity score based results are sensitive to variable selection for score vector and for the comparison group. In their opinion, there is no well-established methodology and therefore the results are influenced by subjective decisions of the evaluators. They conclude that matching methods are useful, but their capability should not be exaggerated.

Relying on the cited literature one could conclude that there is no consensus between micro evaluators and many scientifically interesting aspects of the empirical evaluation methods require further investigation.

### 3. Matching methods

The goal of matching is estimation of program impact on participants. Let  $P$  be a variable that describes state of the involvement in the program and  $P_i=1$  for person  $i$  being participant (member of treatment group) and  $P_i=0$  for out-of the program person (member of control group). Let  $Y_{1i}$  be a value of the outcome variable when person  $i$  is participant, and let  $Y_{0i}$  be a value of the outcome variable for non-participant. The average treatment effect estimate is

$$Y_{ATE} = \frac{1}{N} \left[ \sum_{i=1}^N (P_i Y_{1i} + (1 - P_i) Y_{0i}) \right] \quad (1)$$

The program effect according to (1) is an effect for the average person from the population regardless his participation status. In practice the difference between the average value of the outcome variable for all treated and all non-treated unit is treated as (1). This may lead to bias, because treated units may systematically differ from the control units.

To circumvent the aforementioned problem another statistics is usually computed. It is the average treatment effect on the treated units (ATT). This is a measure of treatment limited to the treated group

$$Y_{ATT} = E(Y_{1i} - Y_{0i} | P_i = 1) = E(Y_{1i} | P_i = 1) - E(Y_{0i} | P_i = 1) \quad (2)$$

The  $Y_{ATT}$  is averaged over participants and shows quantitative impact of the program to them. Its value is equal to the average change in the outcome variable caused by participation. The value informs about a profit or a loss from participation for program beneficent. It is worth to note, that  $E(Y_{1i}/P_i=1)$  is an expected value of the outcome variable for the program participants and is observed while  $E(Y_{0i}/P_i=1)$  is non-observed counterfactual state.

The idea of matching is to compute similarity measure and use the algorithm to match observations from the treatment group with their closest counterpart from the control group. The aim is a construction adequate comparison group that replaces missing data and allows one to estimate  $E(Y_{0i}/P_i=1)$  without imposing additional a-priori assumptions (Blundell & Costa-Dias, 2000). The role of matching is a replication of experimental conditions from cross-sectional data by selection of the proper comparison units. The successfully constructed quasi-experimental reference group provides bias reduction and more precise results (Rubin, 1973).

Conditioning on all covariates available is not possible due to the course of dimensionality. Rosenbaum and Rubin (1983) proposed balancing scores  $b(x)$ , i.e. functions based on significant covariates, such that the conditional distribution of  $X$  on  $b(X)$  is independent from the state of participation. Among balancing scores is propensity score, i.e. conditional probability of the program participation on observable characteristics  $X$ . Matching procedures that rely on balancing scores are known as propensity score matching (PSM).

Propensity score matching is a general method with different implementations and similarity measures. Firstly, the different functional forms for the propensity score can be chosen. Secondly, various similarity measures can be applied. Typical matching estimator has a form (Smith & Todd, 2005)

$$\frac{1}{N} \sum_{i=1}^N Y_{1i} - E(Y_{0i} | P_i = 1) \quad (3)$$

where

$$E(Y_0 | P_i = 1) = \sum_{i=1}^N W(i, j) Y_{0i} \text{ is an estimator of the counterfactual state,}$$

$W(i, j)$  is a matrix of distance between  $i$  and  $j$ , and  $N$  is a number of matched pairs. Objects are matched according to estimated value of similarity measure. The straightforward algorithm is to choose for each object in treatment group an object with the same or very close value of the similarity measure from the control group. Let us define set  $A_i$  such that only one comparison unit  $i$  belongs to  $A_i$ :

$$A_i = \left\{ j \mid j \in \{1 \dots n\} : \min \|w_i - w_j\| \right\} \quad (4)$$

where  $\|\cdot\|$  is a metric. Then, weight matrix  $W(i,j)$  is a square matrix with zeros and ones as elements. The value one is for the closest neighbour, and zeros for all remaining objects. This type of matching is called one-to-one matching. Each unit from the treatment group is linked with only one element in the control group.

The nearest neighbour matching estimator has good statistical properties if  $w_i$  and  $w_j$  are defined on common set. The role of the evaluator is to decide how to treat poorly matched observations (Lee 2005, pp. 89). The sum of distance, the average distance or the median distance between matched pairs  $w_i-w_j$  may be viewed as quality of matching measure. The lower measure the better fit. For the ideal procedure all quality measures should equal 0. When the nearest neighbour matching uses distance or propensity score as a similarity measure then in the infinite samples procedure has a zero variance (Orazio et al., 2009, p. 46). Relying on all matched pair regardless matching quality may affect the balance. On the other hand, if large number of poorly matched pairs would be left out, the size of the control group would shrink and for certain observations in the treatment group could be no adequate comparison in the control group. As a result, they are dropped from the analysis. This would help with the balance but at the cost of efficiency because some information is not used. The evaluator has to choose among the bias and the variance of the estimator.

One-to-one or one-to-many matching is characterised by the risk having poorly matched pairs that is pairs that are distant in terms of chosen similarity measure. To ensure that only well-matched pairs are compared the caliper is used (Cochran & Rubin 1973). The impact of the caliper may be compared to the focus in the camera. When attention is paid to specific point, other distant points are not visible. The procedure simply drops objects without close match.

$$A_i = \left\{ j \mid j \in \{1 \dots n\} : \min \|w_i - w_j\| < \delta w_i \right\} \quad (5)$$

The set  $A_j$  is made of such objects  $j$  that their distance from the nearest match is not greater than  $\delta$ . Unfortunately, there is no one optimal value for the caliper. The literature suggests small number, such as 0.005 or 0.001 (see Austin (2009) for Monte Carlo results). The caliper reduces the bias of average treatment effect estimator at the cost of increased variance (Heckman et al., 1997). In a special case, when propensity score distribution is the same in the treatment and the control group, the caliper cuts off the worst matched pairs and lowers the bias without significant increase in the variance. The caliper also lowers the value of matching quality measures. The cost is lower number of successfully matched pairs. As a consequence the variance of the average treatment effect may increase. However, this is not a major concern as long as one is interested in precise estimation of ATT (Smith & Todd, 2005). On the other hand, Smith and Todd (2005) pointed out that the potential problem with caliper is a lack of *a-priori* knowledge about its optimal value. It is common practice to set the value by try and error method.



In Strawiński (2009) the modified caliper mechanism is proposed

$$A_j = \{j \mid j \in \{1..n\} : \min \|w_i - w_j\| < \delta w_i\} \quad (6)$$

The caliper value is directly linked with estimated propensity score. For the observations with low treatment probability modified mechanism requires better matches from the control group. In practice, there is a few such observations but it is very likely that there is good counterfactual state in the control group for them. A large number of matched pairs with low treatment probability could cause ATT estimator to be biased. Therefore, their influence should be limited despite that for those observations it is relatively easy to find a match. In a situation where probability of participation approaches 1 dynamic caliper will have no major differences from standard one. As a result, greater number of matched pairs is left aside in the computation, those with low participation probability.

#### 4. Empirical results

Strawiński (2009) performed simulations and showed that dynamic caliper in some circumstances outperforms standard procedure. We will try to confirm these results and apply to the real empirical data. We use data from the National Supported Work Demonstration (NSW) survey. We choose these data because they were examined on many occasions in similar works (i.e. LaLonde, 1986; Dehejia & Wahba, 1999; Smith & Todd, 2005) and their results are often cited. The NSW program took place in mid 1970's in the United States. It focused on providing first working experience. The recipients were disadvantaged economically (young mothers) or socially (with criminal record) on the labour market. Among the applicants program participants were chosen randomly.

The random assignment to the program and the survey after program are important elements from the evaluation process point of view. The original control group were influenced by attrition problem. To reduce attrition aside original sample of two additional group were drawn: one from the Current Population Survey (CSP) and the other from the Panel Study of Income Dynamics (PSID).

**Table 1.** The average values of covariates

Sample	N	age	educ	Black	hispanic	nodg	marr	re74	Re75
NSWre74_T	185	25.82	10.35	0.84	0.06	0.71	0.19	2096	1532
		7.16	2.01	0.36	0.24	0.46	0.39	4887	3219
NSWre74_C	260	25.05	10.09	0.83	0.11	0.83	0.15	2107	1267
		7.06	1.61	0.38	0.31	0.37	0.36	5688	3103
PSID	2490	34.85	12.12	0.25	0.03	0.31	0.87	19429	19063
		10.44	3.08	0.43	0.18	0.46	0.34	13407	13597

**Table 1.** The average values of covariates (cont.)

Sample	N	age	educ	Black	hisp	nodg	marr	re74	Re75
PSID2	253	36.09	10.77	0.39	0.07	0.49	0.74	11027	7569
		12.08	3.18	0.49	0.25	0.50	0.44	10815	9042
PSID3	128	38.26	10.30	0.45	0.12	0.51	0.70	5567	2611
		12.89	3.18	0.50	0.32	0.50	0.46	7255	5572
CPS	15992	33.23	12.03	0.07	0.07	0.30	0.71	14017	13651
		11.05	2.87	0.26	0.26	0.46	0.45	9570	9270
CPS2	2369	28.25	11.24	0.11	0.08	0.45	0.46	8728	7397
		11.70	2.58	0.32	0.28	0.50	0.50	8968	8112
CPS3	429	28.03	10.24	0.20	0.14	0.60	0.51	5619	2466
		10.79	2.86	0.40	0.35	0.49	0.50	6789	3292

Age = age in years; educ = years of education; marr=1 married, else 0; nodg=1 not finished school, else 0; black = 1 for Afro-American, else 0; hisp=1 for Latino, else 0; re74 = real income in 1974, re75 = real income in 1975, u74 =1 unemployed in 1974, else 0; u75=1 unemployed in 1975, else 0.

Source: Own calculations based on Deheija i Wahba (1999)

The NSW sample is an original survey sample. It consists of two subsets: experimental one (NSWre74\_T) and control one (NSWre74\_C). In LaLonde (1986) work only those with available income information for 1974 were considered. The author draws two additional control groups, one from the PSID, second from the CPS. The eye examination of the average values of each variable indicates significant differences between the samples in terms of age, ethnicity variables, marital status and salary received before the program. In order to reduce disparities between program participants and members of the control group LaLonde (1986) drew further two control groups from each survey called PSID2, PSID3 and CPS2, CPS3, respectively, with the structure that is more similar to the treatment group.

In the first step of the analysis the experimental results were replicated. This step is necessary to provide comparability with findings of previous researches. In the next step three matching methods and three specifications for functional form of the propensity score, those proposed by Deheija and Wahba (1999), were used. They are presented in Table 2.

**Table 2.** Propensity score specification

Vector	Covariates
1	Age, age <sup>2</sup> , educ, educ <sup>2</sup> , marr, nodg, black, hisp, re74, re75, re74 <sup>2</sup> , re75 <sup>2</sup> , u74*black
2	Age, age <sup>2</sup> , educ, educ <sup>2</sup> , marr, nodg, black, hisp, re74, re75, re74 <sup>2</sup> , re75 <sup>2</sup> , u74, u75
3	Age, age <sup>2</sup> , educ, educ <sup>2</sup> , marr, nodg, black, hisp, re74, re75, u74, u75, educ*re74, age <sup>3</sup>

Age = age in years; educ = years of education; marr=1 married, else 0; nodg=1 not finished school, else 0; black = 1 for Afro-American, else 0; hisp=1 for Latino, else 0; re74 = real income in 1974, re75 = real income in 1975, u74 =1 unemployed in 1974, else 0; u75=1 unemployed in 1975, else 0.

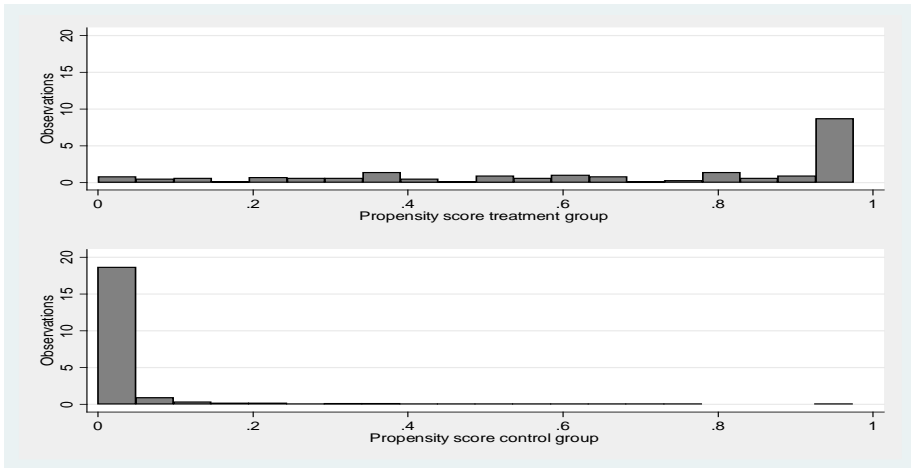
Source: Own calculations based on Deheija i Wahba (1999).

In the next step the propensity scores were estimated for each sample and their distributions were examined. The latter step was to check if there is a potential advantage from dynamic caliper mechanism. The dynamic caliper provides better estimates of the average treatment effect when the distribution of the propensity score values is different in the treatment and the control group. To conserve the space we present graphically the distribution for the first specification only.

Figure 1 presents propensity score distributions for the PSID sample and the first specification of the propensity score vector. For all remaining samples and specifications graphs are similar. There is evident difference in the distribution of the propensity score between treatment and control group. For the treatment group the distribution is skewed to the left and concentrated on (0.95,1) interval, while in the control group it is skewed to the right and concentrated on (0,0.05) interval. This implies major differences between treatment and control group in terms of observed characteristics. In addition, there is a lack of comparison observations for individuals with propensity score value around 0.9. As a result matching without imposing caliper may lead to comparison of incomparable individuals. Therefore, it is advisable to use caliper and dynamic caliper (Strawinski, 2009).

Having in mind the aforementioned properties of the propensity score vector for all specifications three different ATT estimators were used, namely simple one-to-one matching, one-to-one matching with caliper and one-to-one matching with dynamic caliper where a caliper value is a function of estimated propensity score value. The caliper was arbitrary set to 0.005 as this value is usually used in empirical research (Smith & Todd, 2005; Austin, 2009). For models with dynamic caliper mechanism it is assumed that the propensity score for matched comparison unit may differ at most 0.5% from the value of the propensity score for treated unit.

**Figure 1.** Propensity score distribution in treatment and control group for PSID sample



Top panel present distribution of propensity score in treatment group, bottom panel in control group.

Source: Own calculations in Stata 10 using PSMATCH2.

**Table 3.** Panel A. The PSID sample

Method/vector		Linear regression	Matching	Caliper Matching	Dynamic caliper matching
treatment	ATT	-15205			
	SE	1155			
Vector 1	ATT	217	1655	1107	1599
	SE	1106	1864	1871	2254
Vector 2	ATT	275	618	-160	595
	SE	1016	1858	1734	2174
Vector 3	ATT	243	1469	1190	1187
	SE	1023	1614	1424	1533

**Table 3.** Panel B. The CPS sample

Method/vector		Linear regression	Matching	Caliper Matching	Dynamic caliper matching
treatment	ATT	-8498			
	SE	712			
Vector 1	ATT	834	2393	2268	2426
	SE	599	932	923	988
Vector 2	ATT	1199	1991	1603	1355
	SE	555	1013	1009	1066
Vector 3	ATT	1567	1702	1849	1685
	SE	557	1037	1059	1139

ATT stands for the Average Treatment on Treated and SE for standard error.

Source: Own calculations.

The results presented in Table 3 are point estimates of the average treatment effects on the treated. They show a net benefit from taking part in the NSE program. The results include a reference linear regression model and three matching based models. The reference model is a simple regression of earnings in 1978 on treatment dummy. The first column contains the propensity score vector specification. In a seminal LaLonde (1986) article for the experimental sample the estimate of average treatment on treated is 1974 and after correction with linear regression 1962, both numbers statistically significant at 0.01 level. When the original reference group was replaced by the one from another survey the values and signs of the estimates changed dramatically.

When groups are controlled for the distribution of covariates the picture is different (compare results in the first row of each panel with other results). For the PSID control linear regression underestimates the true effect, while matching overestimates. Considering different matching algorithms, the simple matching is closest to the experimental results, and also dynamic caliper results are acceptable, while traditional caliper results are significantly biased. The reason is a vast number weakly matched pairs in the extreme regions of the propensity score distribution. For the CPS control group the picture is slightly different. The linear regression underestimates the effect, while matching overestimates. In case of CPS sample, the use of caliper changes the estimation results. For each propensity score vector specification the result close to experimental one is obtained from caliper or dynamic caliper matching. Therefore, leaving out of the analysis poorly matched pairs is empirically justified. However, one is not able to choose definitely a better method between caliper and dynamic caliper.

In addition, the distance between distributions of the propensity scores for matched pairs were calculated as well as their medians. Usually, when the aim is estimation of the ATT appropriate measure of accuracy is mean squared error or other measures that accounts for bias and variance (D’Orazio et al. 2009, p. 9). In our case we concentrate on reducing the bias. Therefore, we use a median value of difference in propensity scores between treated and control groups. For ideal matching of the treatment and the control group the median of the difference should be equal to zero. The higher median, the worse adjustment in terms of estimated propensity score. The results are presented in Table 4.

**Table 4.** Matching adjustment quality

Method/vector	Matching	Caliper Matching	Dynamic Caliper Matching
<b>PSID</b>			
vector1	2.4239	1.0216	0.8249
vector2	3.7038	1.0459	0.7542
vector3	3.4756	0.9434	0.5722
<b>CPS</b>			
vector1	0.3820	0.1988	0.1196
vector2	0.3859	0.1891	0.1129
vector3	0.3227	0.2348	0.1191

The numbers in the table are medians of differences in propensity scores between treatment and control group multiplied by 1000.

Source: Own calculations.

Independently from chosen sample and functional form of the propensity score, the worst matched are pairs matched by one-to-one without caliper. Setting the caliper improves the results. The median is approximately halved. The dynamic caliper improves matching further. The gain is significant in both samples. After implementation of dynamic caliper the median of unadjustment is a quarter of that for one-to-one matching without caliper.

The results presented in the Table 4 indicate that previously obtained for the ATT should be reinterpreted. Taking into consideration the ATT and the imbalance measures simultaneously we conclude that the proposed mechanism of matching observations with dynamic caliper outperforms one-to-one matching and one-to-one matching with caliper. In dynamic method, the variance of the ATT estimate is reduced and simultaneously the bias is controlled for. Furthermore, we are able to replicate the experimental results in two dimensions. Firstly, in aspect of covariates balance between treatment and control group. Secondly, in terms of the ATT estimate. The results themselves are very close to experimental and the standard deviation of the estimators is only slightly higher than in the standard procedure.

It is important to note that the results are more sensitive to swap of the reference group than to estimation method. In all but one case, where the CPS group is used and the second specification of propensity score vector, the estimates of the average treatment effect on treated are closest to experimental one when caliper or dynamic caliper is used. Therefore, we conclude that dynamic caliper is an effective tool to find unbiased results. The dynamic caliper reduces the variance of the matching procedure at the cost of small increase in the variance of the treatment estimator.

#### 4. Conclusions

The obtained results indicate that in non-experimental methods balancing the treatment and the control group in terms of covariates is at least as important as correct specification of propensity score vector's functional form. Up to now, the process of balancing tedious work and it has been shown that dynamic caliper can ease the problem, because it controls simultaneously for the bias of ATT estimator and the balance of treatment and control group. However, one has to bear in mind that the best estimation tool will fail when there is a lack of candidates for good counterfactuals.

The aim of the current research was a comparison of dynamic caliper with other estimation techniques based on the propensity score. In the literature rather different functional forms for propensity scores were considered or kernel matching. Therefore, we put more emphasis on the process of pairing observations. It has been shown that the procedure modified by Strawiński (2009) is an efficient tool to replicate experimental results.

Relying on the results obtained in this empirical exercise we conclude that matching with dynamic caliper is an efficient method. The variance of matching procedure measures as median of difference in propensity score is reduced (see Table 4), at the cost of the variance of ATT estimator, they are slightly higher (compare standard errors in column matching and dynamic caliper matching in Table 3). But the latter is negligible.

## **Acknowledgements**

The research is co-founded by Polish Ministry of Science and Higher Education grant N111 109335.

I would like to thank you, anonymous referees, for your useful comments and suggestions.

## **REFERENCES**

- AUSTIN P. (2009) „Some methods of Propensity Score Matching Had Superior Performance to Others: Result of an Empirical Investigation and Monte Carlo Simulations.”, *Biometrical Journal*, vol. 5, pp. 171-184.
- BLUNDELL R., COSTA-DIÁS M. (2000) „Evaluation Methods for Non-Experimental Data”, *Fiscal Studies*, vol. 21/4, pp. 427-468.
- COCHRANE, RUBIN (1973) „Controlling Bias in Observational Studies. A Review”, *Sankhya*, vol. 35, pp. 417-466.
- DEHEJIA R., WAHBA S. (1999) „Causal Effects in Nonexperimental studies: Reevaluating the Evaluation of Training Program”, *Journal of American Statistical Association*, vol. 94, no 448.
- DEHEJIA R., WAHBA S. (2002) „Propensity score matching methods for nonexperimental causal studies”, *Journal of the American Statistical Association*, vol. 84, pp. 151-161.
- HECKMAN J., HOTZ J. (1989) „Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training”, *Journal of the American Statistical Association*, vol. 84(408), pp. 862-880.
- HECKMAN J., ICHIMURA H., TODD P. (1997) „Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme”, *The Review of Economic Studies*, vol. 64/4, pp. 605-654.
- LEE M-J. (2005) „Micro-Econometrics for Policy, Program, and Treatment Effects”, Oxford University Press.

- ROSENBAUM P., RUBIN D. (1983) „The Central Role of the Propensity Score in Observational Studies for Causal Effects“, *Biometrika*, vol. 70/1, pp. 41-55.
- RUBIN D. (1973) „Matching to Remove Bias in Observational Studies“, *Biometrics*, vol. 29, pp. 159-183.
- SMITH J., TODD P. (2005) „Does Matching Overcome LaLonde’s Critique of nonexperimental estimators?“, *Journal of Econometrics*, vol. 125, str. 305-353.
- STRAWIŃSKI P. (2007) „Causality, selection and endogenous effects”[Przyczynowość, selekcja i endogeniczne oddziaływanie], *Przegląd Statystyczny* nr 4/2007, pp. 49-61.
- STRAWIŃSKI P. (2009) „Matching with Dynamic Caliper. Preliminary Results”[Łączenie danych z dynamicznym obcięciem. Wyniki wstępne], *Metody Ilościowe w Badaniach Ekonomicznych X*, pp. 232-242.



