

Doing Research for Official Statistics, Some Experiences and Some Suggestions

Beat Hulliger

University of Applied Sciences Northwestern Switzerland¹,
School of Business, e-mail: beat.hulliger@fhnw.ch

Abstract

Research for official statistics will be discussed from the perspective of a methodologist of the Swiss Federal Statistical Office (SFSO) who collaborated in international research projects and represented SFSO in the EPROS working group, and from the perspective of an academic at a University of Applied Sciences who actively participates in research projects for official statistics. While collaborating in a research project of official statistics is extremely interesting the rewards for doing it are limited and the community of researchers in official statistics inside and outside of national statistical institutes is small. Opportunities and incentives to do research for official statistics from within an NSI and from outside are discussed. A key problem is that it is difficult to publish research for official statistics in the traditional scientific journals. A further issue is the lack of allies. Nowadays the banks, the market research companies and the social sciences work on similar research problems but there seem to be few collaborations or joint initiatives among these areas and official statistics. Transfer of the knowledge into practice is an important issue and a challenge.

Keywords: Incentives for research, scientific publishing, innovation, research collaboration, research alliance

1 Introduction

The community doing research for official statistics has been able to produce solutions for important and difficult problems of official statistics. This is not so evident and natural as it might seem. Certainly, there were always researchers and institutions curious enough to go ahead on their own and there were always efforts of coordination by UN institutions and other international bodies. Nevertheless a sustained coordinated effort to do research for official statistics is relatively recent. For example the International Association of Official Statistics, the section of the International Statistical Institute which is devoted to Official Statistics, was founded in 1985 only. The formulation of a chapter on research for official statistics within the the EU Framework Programmes for research and development started, to my knowledge, in the late 80s. The ESSnet of the European Statistical System is a recent, and very welcome, initiative. Many National Statistical Institutes have set up a research programme to foster their innovation process. However, many NSIs do not have such

¹The views expressed in this article are those of the author and do not necessarily reflect the views or policies of University of Applied Sciences Northwestern Switzerland or of any other institution mentioned in the article

a programme and even those who actively pursue a research programme do have problems of financing them.

The objective of this paper is to give some personal experiences and to formulate some suggestions which should help to develop research for official statistics. Inevitably I do not have all the information that the community of researchers and readers have and I apologize for any omission or misinterpretation of facts that I should know. And, of course, I am not speaking in the name of the institutions which I have been working at. This paper reflects my own opinion and my personal experiences and is solely my responsibility. The views expressed in this paper are mine and do not necessarily reflect the views or policies of the Swiss Federal Statistical Office or the University of Applied Sciences Northwestern Switzerland. I count on the feedback of the readers to the discussion.

Since the paper will show my personal view the reader has the right to know a bit more about the author in order to judge his point of view and to disagree freely when he/she deems it convenient.

I graduated in mathematics at ETH Zürich and did postgraduate studies at Universidad Autónoma de Madrid. Back at the Seminar für Statistik of ETH Zürich I was a statistical consultant, teaching and research assistant. I wrote a Ph.D. thesis on the robust estimation of finite population means. I started work as a methodologist at Swiss Federal Statistical Office (SFSO) in 1991. During my time at SFSO I have been collaborating in several research projects of the European Union Framework Programmes for Research and Development. In 2007 I became Professor of Economic and Social Research at the University of Applied Sciences Northwestern Switzerland (FHNW). I have had several roles concerning research for official statistics:

- Delegate and coordinator: Collaboration in the development of research programmes.
- Project Officer: Formulation of calls for tender and supervising project execution.
- Research manager: Formulation of proposals and project leader.
- Researcher on methodology (outliers, editing and imputation).
- Evaluator and referee

In every role other aspects of research for official statistics become important and you never stop learning. Therefore, in spite of the many roles I had, I do not think that I understand research for official statistics completely.

Publications in scientific journals are the most visible part of research in any scientific field. Browsing through some of the important scientific journals in statistics I encountered a few articles on official statistics, too. For example, in the latest issue of the International Statistical Review (ISR) there is an article by Yves Tillé and coworkers on "Sampling Procedures for Coordinating Stratified Samples: Methods Based on Microstrata" [Nedyalkova, Pea, and Tillé, 2008]. This work originated from the collaboration of his group with the Swiss Federal Statistical Office. Now among the 6 articles of this issue it is the only one on official statistics and when I looked for the next article on official statistics I had to go back to the August 2007 issue of ISR.

Table 1: Articles on and from official statistics

Journal	Articels	On Official Statistics	From Official Statistics
ISR	47	6	1
JASA	87	1	0
JRSS	70	2	1
SJES	33	3	0
SRM	12	2	1
JOS	28	25	4
SM	34	34	8
Total	311	73	15

Overall I would count about 6 articles on official statistics of about 30 articles in the last two years.

A rather unsystematic overview over the publishing activity in some of the journals that I would suspect to publish on research for official statistics revealed the picture of Table 1. The choice and classification in the table is subjective, I admit. For example I considered nearly all articles on sampling as of relevance for official statistics. There are certainly many more articles from official statistics in scientific journals which I did not look at, in particular subject-matter oriented journals. And there are many articles which is not declared as research for official statistics and is obviously of interest for official statistics. Thus my very partial view of recent publication in official statistics is as follows.

The Journal of Official Statistics (JOS) and Survey Methodology (SM), which are edited by the National Statistical Institutes of Sweden and Canada, publish the large majority of the articles on official statistics. Journals from economics like the Swiss Journal on Economics and Statistics (SJES) or in more sociologically oriented areas like the Survey Research Methods Journal (SRM) publish a few articles on official statistics, too. Official statistics is a marginal topic in general statistical journals like the Journal of the American Statistical Association (JASA) or the Journal of the Royal Statistical Society, Series A (JRSS-A). ISR, in fact, has a relatively high share of official statistics. The second interesting point is that of the 73 articles on official statistics only 15 were written by authors working in statistical agencies.

Phrasing it optimistically: there is a potential to increase the publication on official statistics in general and a potential to increase publications from inside official statistics.

A first impression thus shows two aspects: There is research for official statistics, but it is not a large field with a lot of publications. There is room for improvement.

2 Some Experiences

The first research project on official statistics I got involved with was about the living conditions of elderly people in Switzerland. This was a project funded by Swiss National Research Fund. In a secondary analysis I studied the sample design and the weighting scheme as part of my diploma thesis at ETH Zürich. The study suffered from a mistake in the weighting scheme of households. A household was selected into the sample when one of the persons living in it was selected and while per-

sons were included with approximately equal probability clearly the households did have unequal inclusion probabilities which were neglected in the weighting. Now the study itself was not conducted by a national statistical institute but obviously of public interest. The debate on the mistake was involving sociologists mainly. Is it research for official statistics when a new weighting scheme is developed, including a variance estimator which tries to take the sampling scheme into account? How narrow or wide do we want to define research for official statistics?

At SFSO I got involved in a research project on distributed access to statistical data in 1994. The project, called RAINBOW, was, as many of the EU-funded projects in official statistics at that time (FP4) driven by information technology. The focus on information technology was somewhat unfortunate because the main problems were on the metadata level and the push towards a demonstration software hindered the reflection on the concepts. This lesson has been learnt and the later programmes were more focused on statistical methodology than on information technology.

A project within SFSO in 1994 on a nomenclature server proved to be more challenging than foreseen in the beginning. There were several nomenclature servers around at that time and the EDIFACT Working Group on statistical data exchange was developing a model for nomenclatures (I was representing SFSO in that Working Group). However, there was no formal description of the revision of a nomenclature though the basic principles were clear. Subsequently nomenclature revisions were investigated more closely by the Statistical Methods Unit of SFSO ([Salamin , 1998], [Hulliger , 1998]). However, the nomenclature project remained in a conceptual state because it became clear that in order to be useful a nomenclature server had to be embedded into an information system architecture and this was not sufficiently developed at that time. The nomenclature research was never officially declared as research and less it was part of a research agenda. When the systems architecture was finished methodology research had been reoriented to other topics. This is part of the fate of research for official statistics. It is very difficult to keep the focus on certain topics during several years. However, often the fruits of research can be expected to arrive after five or more years. In the mean time the production machine is running and many more urgent needs have to be covered.

In 1998 a group of researchers started working on a proposal for a research project on the development and evaluation of new methods for editing and imputation, EUREDIT. The proposal was accepted under FP5 and EUREDIT was executed in 2000-2003. A very interesting group of people came together and it was a very fruitful time of collaboration on the project. A wealth of methods on outliers, automatic imputation and an evaluation methodology was developed. Some of this work was documented in a special issue of JRSS ([Charlton , 2004],[Béguin and Hulliger , 2004]). But the main output, a book which was nearly finished, remained unpublished for several reasons (The last version may be downloaded here: [Charlton et al. , 2003]). The problem is well known: after research is finished there is a need for resources to bring out the fruits. Of course, this should be foreseen right from the beginning. But honestly, we were not sure at the beginning of EUREDIT whether we would be able to gather and synthesise our findings well enough to be able to publish a book. In the end several problems added up to prevent the publication of the book. Nevertheless the EUREDIT project led to a number of articles in scientific journals and was an important input for the development of the EDIMBUS manual on editing and imputation in cross-sectional business surveys [Luzi et al. , 2007].

Research for official statistics is applied research which has its ultimate goal to lead to innovation of official statistics. How much of the research for official statistics has been implemented in practice? There have been projects to deal with how the transfer into practice can be enhanced. Typically these projects find out, that the way from the research findings to the actual implementation is very long and sometimes the research cannot even be recognised anymore. This lack of tracing back innovation to research is common in many areas, not only official statistics. Nevertheless it could help the research community if more reference would be made when implementations use their results.

In this respect the EDIMBUS project mentioned above [Luži et al. , 2007] was and is a success, at least for SFSO. The EDIMBUS project was a forerunner in some sense to the CENEX of the European Statistical System. The initial grant from Eurostat was multiplied by the investments from the three participating NSIs (ISTAT, CBS Netherlands and SFSO) at least fourfold. The funding of Eurostat had the merit to trigger this investment. The development of the manual was tough work. Though writing a manual may not sound like research I personally think that the change of focus during the project from particular methods to an overall design of statistical data preparation was typical of research. The implementation at SFSO was again an investment for SFSO but it has led to a change of the perception of data preparation. The development of data preparation and its place in the production processes at SFSO got an important impulse from EDIMBUS.

The Statistical Methods Unit of SFSO was involved in a large project on variance estimation from 2001 to 2004, DACSEIS, as a data provider [Münnich et al. , 2003]. This involvement proved useful for SFSO because some of the methods could actually be implemented afterwards. However, the small amount of funding by no way covered the expenses for the project. In fact the overhead for the management of the project showed that it is sometimes preferable to obtain no money at all than to obtain too little money.

In 2002 a group of researchers at NSIs and universities started work on a proposal to be submitted to FP6: Network for Improving Public Statistics (NIPS, or eNIPS because it had to be submitted to the Information Society Programme). It was a network of excellence proposal covering a large part of the research for official statistics, gathering some 70 partner institutions and worth 20 Million Euro. The FP6 officials at that time said that networks of excellence should be large and all encompassing. The proposal, submitted in April 2004, was not accepted. Among the reasons was that the management of such a big network is something non-trivial. Maybe we let jump in too many institutions in the end. But as an afterthought NIPS was probably just too big and it was difficult to show that a research community, which had not yet established a network, would be capable to form itself into an efficient network. A lesson learned was that research for official statistics must organise itself and not rely on the Framework Programmes alone.

As the delegate of SFSO in the EPROS meetings I also had the occasion to follow and support as much as possible the programming of research for official statistics on a European level. One of the problems we were facing always was the multitude of topics that should have been handled. Everybody working in official statistics is aware of the wide range of areas to be covered and the complexity of the processes involved in statistical production. Sometimes there was a feeling that official statistics is alone with its problems. It is true, for many persons outside official statistics it

may sound quite strange to spend millions of Euros in the development of nomenclature servers. However, I think that there are more people working on similar problems than we believe. As an example take editing and imputation, or statistical data preparation how I prefer to call it. Many market research institutes do practically no data preparation on an individual level. However, they certainly do what we call macro-editing, say just proper statistical data analysis. And micro-editing? When I talked to a colleague of a large market research company specialised in analysing scanner data, I realised that they have a huge data preparation programme going on and it takes them a lot of money and time to get the masses of data into a usable shape. They had the same problems as official statistics! Also risk evaluation of banks and insurances suffer from the problems of data preparation. We probably did not work enough on these parallels to really grasp the opportunities of collaboration on research topics of interest outside official statistics.

3 Some Suggestions

My suggestions for research for official statistics draw on the above experiences and on my personal view of how research for official statistics could be strengthened. Some of these suggestions may reveal unrealistic or wrong in the view of the readers and I am looking forward to an interesting discussion. The important thing is that we try to think about how research should be done because there is room to enhance things and good research is necessary for the future of official statistics.

3.1 Raise the number of papers on research for official statistics in scientific journals:

This is most important for the visibility outside the relatively small circle of official statistics. And this is most important to obtain its share in the funding process. The journals JOS, SM and ISR do all what they can to bring out interesting articles on research in topics related to official statistics. However, I dare to ask whether we need more possibilities to publish research for official statistics is needed. Is there space and need for a journal on research for official statistics? There was a journal called ROS, for research for official statistics, edited by Eurostat, which could be re-animated. This could give a further incentive for researchers from academia to go into the field of research for official statistics. For researchers from outside the statistical agencies, in particular from the business sector, it may be necessary to include the publication of papers in tenders. Obviously this would increase costs of research in the short. However, writing and publishing papers is also a quality assurance tool and the increased visibility of the research will have fruits in a larger share on research programmes on the long term.

3.2 Raise the number of papers written by authors from inside official statistics:

Too often, the daily business of official statistics does not leave time for writing up important work. Sometimes this is due to the feeling that the work at a particular agency is too specific to be usefully exposed in a scientific journal. Certainly an effort towards generalisation must be made. However, we should not be too afraid of publishing research for official statistics with a rather special topic. Other disciplines

do this, too. Therefore, statistical agencies should encourage publication of papers by their researchers more. Maybe an award for the best paper from authors of national statistical agencies could work as an incentive.

3.3 Raise the number of technical reports written in English and make them available electronically:

One of the problems why research is not transferred into the practice of other National Statistical Institutes is the language barrier. It may be necessary to translate much more or to write from scratch in English in order to foster exchange of experiences. This is sometimes not compatible with national rules about publications or it needs extra financing to translate reports. Nevertheless I am convinced that this is a key problem in European research for official statistics. Once there are more reports in English there is the urgent need for an electronic repository accessible freely by the community of researchers for official statistics.

3.4 Find allies in fields related to official statistics:

For outsiders official statistics is a closed world which has little connection to their own research. However, there are many researchers in banks, market research companies, retail organisations, information technology companies and, of course in universities, which treat similar problems. There is a potential for a fruitful collaboration. Maybe we should abandon the denomination “research in official statistics” or “research for official statistics” in favor of a more general term. For example it may be advantageous to put “macro statistics” in the title of a new journal to attract researchers which work in areas related to official statistics.

3.5 Bring innovation of official statistics to the foreground:

The funding through the FP programmes is very welcome and indeed has the potential of leveraging much more research by the NSIs. However, the influence of official statistics on the programmes can never be perfect and compromises in topics, execution, results, timing etc. have to be made. Innovation of official statistics itself should set the goals and priorities of research for official statistics. And therefore the transfer of the research results into practice is so important. The ESSnet initiative is an utmost important step in the collaboration on a European level and for fostering transfer into practice. It will be also important to involve the above mentioned allies in shaping a research agenda for macro statistics and in sharing experience in innovation.

4 A final word

As a part of the experiences in researching for official statistics I have to admit that I had a lot of fun and the work was exciting. The fun and excitement refers to the discussions at the project meetings with researchers devoted to the project and to the heated discussions on the best way forward. These are the highlights and I am very grateful for these occasions. There was also boring and annoying administrative work. Fortunately it did not prevail.

The collaborators in the statistical agencies are very keen on attaining high quality and, at the same time, are very realistic in what concerns time, resources and problems. And there are researchers in academia and in the private economy who are willing to collaborate in research for official statistics. This is a good basis for challenging, exciting and rewarding research projects.

References

- Charlton, J. and the EUREDIT project team (2003). Towards Effective Statistical Editing and Imputation Strategies - Findings of the Euredit project, unpublished manuscript, 2 volumes. Available at <http://www.cs.york.ac.uk/euredit/results/results.html>.
- Charlton, J. (2004) Editorial: Evaluating automatic edit and imputation methods, and the EUREDIT project, *Journal of the Royal Statistical Society Series A*, 167/2, 199-207.
- Luzi, O., De Waal, T., Hulliger, B., Di Zio, M., Pannekoek, J., Kilchmann, D., Guarnera, U., Hoogland, J., Manzari, A., and Tempelman, C. (2007). *Recommended Practices for Editing and Imputation in Cross-Sectional Business Surveys*. ISTAT.
- Hulliger, B. (1998). Linking of Classifications by Linear Mappings. *Journal of Official Statistics* 14, 255–266.
- Béguin, C. and Hulliger, B. (2004). Multivariate Outlier Detection in Incomplete Survey Data: the Epidemic Algorithm and Transformed Rank Correlations. *Journal of the Royal Statistical Society Series A* 167, 275–294.
- Münnich, R., Schürle, J., Bihler, W., Boonstra, H.- J., Knottnerus, P., Nieuwenbroek, N., Haslinger, A., Laaksonen, S., Wiegert, R., Eckmair, D., Quatember, A., Wagner, H., Renfer, J.-P. and Oetliker, U., (2003), *Monte Carlo Simulation Study of European Surveys*, DACSEIS deliverables. <http://www.dacseis.de>.
- Nedyalkova, D., Pea, J. and Tillé, Y. (2008). Sampling Procedures for Coordinating Stratified Samples: Methods Based on Microstrata. *International Statistical Review* 76 (3), 368–386.
- Salamin, P.A. (1998). *Etablissement d'une clef de passage pondérée entre l'ancienne (NGAE 85) et la nouvelle nomenclature (NOGA 95) générale des activités économiques*. Swiss Federal Statistical Office.
- Schweizer, W. (1980). *Die wirtschaftliche Lage der Rentner in der Schweiz*. Vol. 1 and 2. Bern.