DESTATIS
wissen.nutzen.

METHODS – APPROACHES – DEVELOPMENTS
Information of the German Federal Statistical Office

Edition 1/2016

The catchword

Methods of the federal statistics – Further development

Events

The catchword

Provision of European Public Use Files for EU-SILC / Labour Force Survey - Methodological Studies in the framework of a Eurostat project

Eurostat offers access to European microdata for scientific purposes. One type of access is provided in the form of Scientific Use Files (SUFs) as they are called. Those files represent micro-level data that have been anonymised to an extent which permits their exploitation outside the secure premises of an official statistical institute. Since access to such data is however limited, the data are anonymised only to an extent that the re-identification of respondents is made difficult, but not rendered completely impossible. Only organisations recognized as research entities are granted access to those data. Furthermore, the decision whether or not to grant such access is made individually for each project on the basis of a research proposal to be submitted on the planned research project. Since the accreditation process described above may take up to 10 weeks, it would be highly beneficial if there were an additional offer of anonymised microdata accessible without any restrictions. This would allow researchers to previously check whether the microdata files are suitable for their research projects, as well as to carry out initial tests and examine their evaluation scripts in advance. These Public Use Files (PUFs), as they are termed, might also be used for statistical training. Since the relevant microdata were made publicly available in that case, they would however have to be anonymised to a much greater extent in order to completely rule out the identification of respondents. The art in this case is to anonymise the data in such a way that the resulting Public Use Files still have sufficient information content so that they are still largely appropriate for teaching purposes and for preparing the research work with the corresponding Scientific Use Files. Moreover, those files should have the same structure, i.e. the same variables and the same variable values as the associated Scientific Use Files in order to fulfil that purpose.

Hence, Eurostat set up a project[1] at the beginning of 2015, in which suitable procedures are to be examined and a proposal for a harmonised approach is to be elaborated. In this context, the challenge has to be faced that the Public Use Files sometimes have to fulfil quite different requirements in the individual member states. The two sets of EU statistics that are most frequently demanded as Scientific Use Files, i.e. EU-SILC (European Union Statistics on Income and Living Conditions) and LFS (Labour Force Survey), serve as project datasets here. While a traditional approach is tested for generating the LFS Public Use File, the Public Use File for the EU-SILC will be generated completely synthetically.

Traditional Approach

With the traditional approach, the initially anonymised Scientific Use File of the Labour Force Survey is further anonymised using classical statistical disclosure control methods. The methods applied for this purpose are already in use for constructing European Scientific Use Files. Here, the aim is to prevent identification due to unique combinations of variables. For that purpose, some particularly identifying variables are removed first, such as, for instance, all variables which allow to reconstruct the households. On the one hand, this is mandatory in some member states. On the other, it has become apparent that researchers are primarily interested in the personal information when evaluating the Labour Force Survey. To preserve the structure, however, the relevant variables are not removed completely, but their values are merely set to "missing" globally. Also, the regional variable is essentially removed by being globally recoded to the country level. The project partners moreover agreed on a list of thirteen identifying variables such as, for instance, age, gender, degree of urbanisation, nationality, occupation and household size. The information content of the variables is further reduced by building categories. For instance, several age classes are grouped together, and households of the size 6+ are subjected to top coding. Since despite coarsening of the identifying variables unique combinations of them can

---

[1] Partners in the project are the National Statistical Offices of the Member States Austria, Finland, France, Germany, Hungary, the Netherlands, and Slovenia. The National Statistical Office of the Netherlands serves as project co-ordinator.

still be found, some values are removed by local suppression. Two different approaches have been tested to serve this purpose. First, k-anonymity has been performed on a limited set of identifying variables in combination with post-randomisation on the remaining. By this approach, individual values are deleted in such a way that each cross combination of the seven selected identifying variables occurs k = 5 times at minimum. For the remaining identifying variables, their respective values are randomly assigned to another category, with probability of not changing the category set to 80%, and the remaining probability equally distributed over the other categories. Since for some member states only an approach taking all identifying variables into account is viable, the all-m approach as it is called is considered as an alternative. With the all-m approach, each m=4 combination of the identifying variables has to occur at least 10 times in all dimensions. For implementation, the μ-Argus program as well as the R-Package sdcMicro have been tested as alternatives.

Synthetic dataset

Since researchers are particularly interested in the income variables comprised in the EU-SILC dataset, which however must not be included in the Public Use Files in some countries, a traditional approach would not be adequate here. Hence, the decision was made to produce a fully synthetic PUF for the EU-SILC cross-sectional data. Since all values are "fake", it is moreover possible to retain the household structure in the PUF - in contrast to the traditional approach. For reasons of quality, the models for producing the synthetic data are based on the raw data. Those data are used to generate a synthetic population as a start. As a first step, the number of households of each household size in the population is estimated using the Horvitz-Thompson estimator, then the synthetic population is constructed by producing exactly that number of households. Subsequently, the household structure - made up of age and sex distribution within the household - is drawn randomly from the raw data, stratified by household size. This is done to prevent illogical household structures, such as households made up of members of under 5 years of age only, for example. As a next step, some important categorical variables such as highest level of education, for example, are simulated using multinomial logistic regression models. This is performed in a sequential way, so that - in addition to age and gender - all previously simulated categorical variables are integrated in the models. After that, some income totals are generated, such as the total income of a person. For this purpose, the continuous variables are first categorised and subsequently simulated using multinomial logistic regression models again. The continuous variable values are then drawn randomly assuming a uniform distribution within the classes (respectively a Pareto distribution in the highest class). After that, the totals generated in this way are split into their components again. This is performed by randomly drawing the proportions of the components in the totals from the raw data, stratified, for instance, by self-defined labour status. All simulation steps are performed separately for the NUTS level 2 regions. On the basis of the synthetic population generated in that way, a stratified random sample with the region as stratum variable is then drawn in order to generate the Public Use File. All variables not yet simulated up to that point are then generated using less refined methods, such as drawing them from their distribution in the raw data (conditional on the income classes). With the aim to simulate the structure of the corresponding Scientific Use File, the same disclosure control techniques that have been used to produce the Scientific Use File are equally applied to the production of the Public Use File as a last step. This ensures that the structure of the resulting Public Use File exactly corresponds to the structure of the Scientific Use File. For implementation, the R-Package simPop was used.

Outlook

The quality of the contents has not yet been assessed up to now. For that purpose, the relative deviation for some of the main indicators is to be calculated on the basis of the Scientific Use Files and the Public Use Files. For the Labour Force Survey, this is the unemployment rate by age class, for example, and for the EU-SILC, it is the at-risk-of-poverty rate. Upon completion of the project, it is furthermore necessary to obtain the consent of the member states before the Public Use Files that have been produced will be made available. According to the procedure followed in providing the Scientific Use Files, there will also be the opportunity here for individual member

states to adopt further country-specific measures in addition to the proposed disclosure control techniques. Moreover, Eurostat plans to conduct a follow-up project in which Public Use Files are to be produced for further statistics.

Literature

Alfons, A., Kraft, S., Templ, M. and Filzmoser, P. (2011). Simulation of close-to-reality population data for household surveys with application to EU-SILC. Statistical Methods & Applications, 20, 383–407.

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Schulte Nordholt, E., Spicer, K. and de Wolf, P.P. (2012). Statistical Disclosure Control, ISBN 978-1-119-97815-2, Wiley.

de Wolf, P.P. (2015). Public Use Files of EU-SILC and EU-LFS data. Presented at the UNECE/Eurostat work session on statistical data confidentiality, Helsinki, Finland.

Lydia Spies, tel: +49-(0)611 / 75 33 22, e-mail: lydia.spies@destatis.de

Methods of federal statistics - Further development

The electronic diary in the continuous household budget surveys

The continuous household budget surveys (LWR) are an annual voluntary household survey for which households write down, inter alia, their income and expenditures over a period of three months. Since 1 July 2015, a web diary has been offered in addition to the paper diary in the continuous household budget surveys. Hence, the households who take part in the survey have now for the first time the option to use an electronic form for reporting their income and expenditures.

User-friendliness is of course of utmost importance for such a modern survey tool. In fact, only if the respondents fully accept such a tool, and if the complex survey contents are transformed in an adequate manner, will the use of such an instrument be successful. With the aim to ask users only the questions that are actually relevant for them individually, filters have been integrated in the electronic household diary that automatically guide users through the list of questions. Those filters specifically permit to hide those parts of the form that may not apply to the particular household in question. With the aim to make the user interface rather comfortable to use, a variety of web-mode specific functions have been integrated additionally. Among those are, for instance, a calendar function for entering the date and a function for sorting the expenditures that have been entered.

The data generated from the continuous household budget survey serve in particular to establish the basis for determining the detailed weighting patterns for the calculation of the consumer price index. For this purpose, it is necessary to break down the expenditures of households in great detail in accordance with the German Classification of Income and Expenditure of Households (SEA 2013). In the paper diary, the households taking part in the survey are asked to deliver plain text information in as much detail as possible on the majority of expenditures. The information provided by the households in this way is subsequently coded by the Statistical Offices of the Länder in accordance with the SEA Classification. The core of the web diary, however, is the independent coding of expenditures carried out by the households themselves with the help of a search function especially developed for that purpose. The households are asked to enter their expenditures for goods and services in detailed plain text, too, when using the web diary. Subsequently, they allocate their entry to the appropriate category to be chosen from a hit list they are offered. The underlying search

function establishes content-related links between the words entered in plain text and the SEA 2013 classification, which has been modified to suit the purposes of the continuous household budget surveys. To this end, two steering files have been created and incorporated in the electronic diary. The first steering file (list of keywords) comprises more than 50,000 search terms, i.e. keywords in everyday language that households typically use for entering their expenditures, as well as the corresponding SEA 2013 codes. With the help of a second list (SEA hit list), in which the same codes are equally listed, as well as the corresponding categories, the hits which best suit the keywords entered by the user are displayed. The user is then asked to select the suitable category from the displayed hit list and to allocate it to the expenditures he or she has entered. For instance, if the user enters "daily newspaper" as search term in plain text, the following four categories to choose from are offered: Daily newspaper sold by retail, daily newspaper subscribed to in print form, electronic daily newspaper subscribed to on the Internet, or no allocation possible. The user might then select "daily newspaper sold by retail", for example. However, there is always also the option for the user to select the category "no allocation possible" if he or she is of the opinion that no suitable hits have been displayed. Expenditures that have not been allocated to a specific category by the user are subsequently coded by the statistical office, taking the plain text information entered by the user as a basis. In addition, the search function of the electronic household diary allows the user to search for word combinations, too.

While the respondent – when using the paper diary – needs to browse first in order to identify the paragraph required for making entries on income and regular expenditures (e.g. housing expenditure), he or she may select the suitable position for his or her entries immediately in the navigation tree when using the web diary instead. The user may open or close the individual positions in the navigation tree at any time in any way and thus obtain a good overview of the entire survey. When logging on to the electronic diary the next time, the page that was used before for making the last entries automatically reopens. In addition to using the navigation tree, the respondent may also navigate through the web diary by clicking on the buttons "next" or "back", respectively.

Explanations that can be found in the footnotes of the paper diary are, by contrast, offered directly in the context of the respective question in the web diary, i.e. in the form of a context-sensitive help function indicated by an info icon. In practice, that means that if the user moves the mouse pointer over the icon, explanations in the form of text relating to the respective question are displayed on top right. Another function offering support to the user in filling in the electronic household diary is a "help page" which comprises, among other things, a total of six video tutorials (pertaining to each section) which include numerous examples. The electronic diary furthermore comprises a short guide, a user manual as well as "frequently asked questions" as reference sources.

Thanks to the independent coding carried out by the households themselves, the expenditures are categorised already in the web diary. It is hence possible to offer the user summarised overviews of the entries he or she has made. "At the touch of a button", the responding household may generate the following three different types of summary: 1. Income, in detail, 2. Expenditures, in detail, as well as 3. Overview of income and expenditures, in a less detailed breakdown. It is possible to compile each summary for an individual month as well as for the complete reporting quarter. The summaries may be stored as csv-files and then further processed. Additionally, a chart is displayed in each case which shows the shares computed on the basis of the overviews. In addition to compiling overviews, the households may also export their complete entries from the electronic diary in the form of csv-files. This provides them with the opportunity to carry on the household diary for their own purposes independently of the web diary application.

The entries that the households make in the web diary are immediately checked for errors. Error checking is performed at the moment the information is entered, but finally also before the data are transmitted to the statistical offices. With the aim to keep the burden on respondents as low as possible, however, complex combination checks are carried out only

subsequently by the statistical offices of the Länder. By means of such technical support given during the process of entering the information and the initial editing of the data before they are transmitted, the number of queries to be made later to clarify households' entries can be reduced. Nevertheless, as before, re-editing and, if necessary, re-coding of the data are still performed by specially trained staff members in the statistical offices, who also contact the households in cases where the entries need further clarification.

Close to one third of the households taking part in the continuous household budget surveys has opted for using the electronic diary. In fact, the majority of those households use the electronic diary over the complete registration period of three months. At this moment, we cannot predict how many of those households will actually continue to use the web-based diary on a permanent basis. "Online-dropouts" are given the opportunity to switch over to the paper diary at any time. By this, we try to prevent that households drop out of the survey completely. The impact the web-based data collection with the electronic diary has on the data quality of the continuous household budget survey can only be analysed in the year 2016, as soon as the survey data from the third and fourth quarters of 2015 will become available. On the basis of a separate online-user survey, the electronic diary will be assessed by the households, too. In that survey, users are asked about their degree of satisfaction with the software program, any problems they might have encountered as well as potential improvements for the future which might help to further optimise the electronic household diary. The user survey will be evaluated in the year 2016, too.

In the process of further developing the electronic diary, further improvements in the area of information technology must also be taken account of in the medium term. One of the next steps, for instance, should be ensuring the compatibility of the electronic diary with tablet PCs. In addition, the development of an application software for mobile devices (app) would be desirable for the future.

Birgit Lenuweit, tel: +49-(0)228 / 99643-8850, e-mail: birgit.lenuweit@destatis.de

The Integrated System of Tax Statistics - Interim assessment and progress report II

Preliminary note

In the first issue of "Methods – Approaches – Developments" (MAD) of 2012, we made a first interim assessment of the progress of the project: "New concept for the system of tax statistics". The migration of the original sets of tax statistics into the new Integrated System of Tax Statistics (StSys) has almost been finished now, however, new challenges in the context of the system of tax statistics have already arisen. In this contribution, we reflect upon the experiences made in developing a new concept for a statistical system and provide information on the current progress of the project.

In order to fulfill the new requirements to be met by tax statistics[1], the working group „tax statistics", which is composed of the specialised departments of the statistical offices of the Federation and the Länder, first elaborated the fundamental requirements to be met by a new Integrated System of Tax Statistics. In addition to developing a uniform framework application for all tax statistics, the intention was to redesign and optimise the statistical production process for

---

[1] Cf. Ehlert, A.: New concept for the system of tax statistics, in: Methods – Approaches – Developments, 1/2009, p. 11 f.

the individual sets of tax statistics. Furthermore, the physical management of the data had to be restructured. The aim was to process the tax statistical data in line with the principles of Central Production and Data Management (ZPD). Hence, that meant that the decentralised and distributed data management and processing in the statistical offices had to be replaced by central production and data management at a single location.

Project management

The goal was (and still is) to integrate the individual annual sets of tax statistics successively into the framework application to be developed. In line with the underlying aim to subdivide the overall project in different phases, a list of specifications was drafted in the first project which facilitates the programming of a framework application for all tax statistics, and at the same time meets the requirements of the new turnover tax statistics (assessments). Eventually, practice showed that it was not possible to fully implement the original requirement of drafting a list which at the same time specifies the requirements to be met by a framework application for all sets of tax statistics and also for the turnover tax statistics (assessments). Back then, the project participants decided to focus primarily on the requirements for turnover tax statistics (assessments) and to adjust the framework application accordingly, since the first project would have become rather complex otherwise. At a later stage, the framework application was successively enhanced during the process of integrating the remaining tax statistics in order to meet the specific requirements of the individual tax statistics.

The subdivision of the project into different phases exerted a substantial influence on the overall project structure and on the course of the project. According to the rules on coordination of work between the Federal Statistical Office (FSO) and the Land statistical offices (LSOs) in their Statistical Network System, it is the task of the specialised departments of the FSO/LSO network to draw up a coordinated list of specifications if decentralised statistics (such as tax statistics) are concerned. In addition to the programming assignment, an invitation to tender concerning the quality assurance of the program is published within the Statistical Network, and a corresponding contract is awarded. Thus, the step-by-step approach implied a new project and hence a new list of specifications, a new invitation to tender, and a new contract to be awarded for each integration respectively new concept for a set of tax statistics. As a consequence of that approach, seven Land statistical offices have been involved in programming and quality assurance tasks of the overall project up until now. At the same time, three sections of the Federal Statistical Office (project managing responsibility) are in charge of managing the project.

Besides those organisational effects, the phased approach also has repercussions on the timing. The programming work for the first set of tax statistics [turnover tax statistics (assessments)] was finished in February 2010, and programming for the last original tax statistics (corporation tax statistics) was finished in September 2015. The fact that the time spans of the two projects did not coincide, plus the fact that the responsibilities for drawing up lists of specifications and carrying out the programming work were divided, rendered it quite impossible to apply a totally uniform approach to the projects. As a result of those framework conditions, partly different processing methods and procedures were chosen in designing new concepts for the tax statistics, which have to be harmonised in a follow-up exercise now – to the extent that is professionally expedient – in order to fully exploit the potentials for synergies offered by uniform processing procedures. That unification process requires a high degree of regular and intense communication and co-ordination among the project participants and the statistical offices.

As an alternative, one might have developed the new concept for all sets of tax statistics including the framework application in the form of a large-scale project instead. In that case, there would have been a single invitation to tender, and the number of project participants would have been smaller. One might also presume that communication and co-ordination between the parties to the project would have been easier. Moreover, it would have been easier to implement uniform specifications and requirements, processes and procedures and programmings in practice. Even with that approach, however, it is likely that it would have been necessary to perform the programming work for the individual sets of tax statistics in a successive manner,

since the parallel programming for all sets of tax statistics would have required too many capacities in terms of personnel. Hence, there would have been no time gain compared with the approach actually chosen. Additionally, if we had opted for a large-scale project, we would have had to anticipate all the particularities of the respective sets of tax statistics and take account of them in the framework application to be programmed. It would have been impossible to cope with such a task. In some cases, there are large differences between the sets of tax statistics as far as their contents are concerned, which is shown, among other things, by the complexity of tax legislation. For instance, the program requirements of a set of turnover tax statistics markedly differ from those of inheritance and gift tax statistics. In a large-scale project, there would have been a major risk of time delays occuring right at the start of the project with the project design phase (needs analysis, specification of methods, elaboration of lists of features/specifications, etc.), due to the high complexity and number of individual sets of tax statistics. By contrast to that, the step-by-step approach actually chosen has the benefit of enabling the agents to directly focus on the requirements of a set of tax statistics in developing a new concept. Furthermore, it is possible to base the follow-up projects on the experiences made and on the lists of specifications available, which helps to achieve best-practice solutions. Requirements from earlier projects were critically scrutinized and adapted, where appropriate. As a matter of fact, that type of project management is similar to that of prototyping[1]. Prototyping is referred to as a type of project management by which different intermediate results are developed step-by-step and continually adapted until the end product is reached. Prototyping has the advantage that it becomes discernible in the early stages of the product development whether or not the product meets the defined requirements. In contrast to developing a new concept for tax statistics, however, the prototyping approach would require each intermediate product (project) to be a finished statistics program. The continuing adaptations of the existing programs, as would be the case with the prototyping approach, have been performed are still being performed often with each new project.

To sum up, we must admit that the benefits of a step-by-step approach such as an easier project start and an easier project management are partly set off by adjustments required in the follow-up process. However, the project goals, i.e. the production of tax statistics using a uniform framework application at one single location, and the development of a new concept for the statistical production processes and the concomitant use of synergies in data processing, have been achieved without major delays or restrictions.

Current progress of the project and outlook

For the time being, the wage and income tax statistics and the corporation tax statistics have been the last tax statistics for which new concepts have been designed. The development of new programs for wage and income tax statistics was finished in June 2015, and for corporation tax statistics, programming was finished in August the same year. The processing of the data from the 2012 wage and income tax statistics is performed within the Integrated System of Tax Statistics (StSys) now. Hence, in addition to wage and income tax statistics, the two sets of turnover tax statistics (advance tax returns and assessments), the inheritance and gift tax statistics, the statistics of partnerships as well as the trade tax statistics have now been integrated in the overall system of tax statistics (StSys). The corporation tax statistics of the year 2013 will presumably be integrated in the StSys in 2016.

In the wake of integrating the wage and income tax statistics in the StSys, the server performance was enhanced at the same time, since the wage and income tax statistics which comprises close to 40 million datasets and 1 500 variables counts among the largest official statistics. We departed from a data volume of approximately 90 Gigabytes per reporting year in our calculations. The expansion of the server capacities benefitted all sets of tax statistics, and the performance times have improved. During the development of a new concept for wage and income tax statistics as well as corporation tax statistics, the short-term need arose to link data from the net income account (EÜR) with the above-mentioned statistics as well as with data from

---

[1] http://www.gruenderszene.de/lexikon/begriffe/prototyping

the partnership statistics. For business years starting after 31 December 2010, entrepreneurs who are not liable to keep books have to submit the new annex "net income - EÜR", which comprises variables for the determination of income, in connection with their tax declarations filed in electronic form. In the net income account (EÜR), operating expenditure and receipts are balanced against each other in order to determine net income.

We already have the next projects in the starting blocks. As ninth project, the model calculations for determining the share the communities receive from the income tax revenue are to be integrated in the overall tax statistical system StSys, and a standardised program will hence be created for this purpose. As one can tell by the name, that exercise does not represent statistical processing in a stricter sense, but it is rather a calculation procedure. In that regard, that project differs from the preceding projects and poses new challenges to all participants.

Furthermore, the e-balance sheet as "new tax statistics" has entered the project kick-off phase. In that set of statistics, data from enterprises liable to keep accounts are subjected to statistical processing. It is planned to link the data from that set of statistics with other tax statistics, as it is also the case with the EÜR. Upon completion of that project, statistical information on the balance sheet and on the profit and loss account of enterprises will enhance the data offered by official tax statistics in the future.

In addition to those new statistical projects, we have planned IT projects which are aimed to further improve the processing and evaluating possibilities in the Integrated System of Tax Statistics. For that purpose, we have to investigate in the context of the Standardisation of Processes exercise (SteP) which standard IT tools are available and can be integrated in the StSys in an expedient way. Last but not least, it would be desirable to expand the Integrated System of Tax Statistics by an evaluation database which comprises disclosure control applications and which permits to carry out flexible evaluations using the SAS software, among others. Exporting the data from the StSys, as we currently practice it, would be rendered obsolete by this, and the current benefits of the Central Production and Data Management (ZPD) for data processing would be complemented by the benefits provided by centralised evaluation possibilities.

Axel Ehlert, tel: +49-(0)611 / 75 47 56, e-mail: axel.ehlert@destatis.de

Introducing a grossing up procedure in statistics on over-indebtedness

Basic situation

At annual intervals, the statistics on over-indebtedness provide information on the situation of people affected by over-indebtedness in Germany. The basic population comprises people with financial problems who seek support from a debt counselling agency. The data are collected from the roughly 1,400 counselling agencies in Germany. As a rule, the agencies' participation in the survey is voluntary. In the reference year 2014, 396 debt counselling agencies contributed to the statistics. This shows that the statistics' representativity has been a long-standing problem. Even though the participation rate has steadily increased in the past years, the statistics on over-indebtedness have so far only included counselling cases on which data were actually available. This is an unfortunate situation regarding the information value of the statistics, which is to be rectified by grossing up the available data to basic population level.

Grossing up to population level in two stages

In addition to the debt counselling agencies' voluntary participation, the persons counselled need to agree to their data being used for statistical purposes before a specific case of

counselling can be incorporated into the statistics on over-indebtedness. Consequently, even the majority of participating agencies cannot provide information on all the cases they handled in a reference year. Therefore, the data for the over-indebtedness statistics must be grossed up using a two-stage procedure. The first stage serves to compensate for non-participating agencies, the second stage adds information that is missing at the counselled persons' level.

This is clearly not the kind of randomised sample to which grossing up techniques have traditionally been applied. Instead of a random selection procedure, the counselling agencies and the people counselled decide whether they participate in the survey. To gross up the sample under these conditions, it is necessary to assume that the characteristics of the cases reported do not systematically differ from those of the unreported cases. As there is no indication of a systematic pattern behind the cases of nonresponse, it is possible to use simple expansion estimation. The grossing up factor for this is the inverse of the participation rate. A participation rate of, for example, 20 percent translates into a grossing up factor of five. To account for possible regional differences, the first stage of grossing up is carried out individually for each Land. In 2014, the participation rates varied widely across the Länder, from (nearly) 100 percent in Berlin and Schleswig-Holstein to just four percent in Nordrhein-Westfalen. This results in different grossing up factors (from 1 to 25) for the first stage of grossing up. The basis for establishing grossing up factors is provided by a database of the Federal Statistical Office, which contains all known debt counselling agencies that are sponsored by the relevant Land ministries of social affairs. This database is updated every year through enquiries with the Land ministries and represents the population to which the results are grossed up. For the second stage of grossing up, information on the total number of counselling cases in the reference year is gained from the counselling agencies' reports. This figure is provided by the agencies regardless of the number of consents. Using this figure and the number of counselling cases reported, the grossing up factor can be calculated by division for each counselling agency.

Results and outlook

Compared with the ungrossed figures for previous years, there are no significant deviations. This means that grossing up does not cause systematic bias in the results. The error calculation we made supports this conclusion, too. After grossing up, the key figure of the statistics, i.e. the average amount of debt, was 34,504 euros in 2014, in contrast to an ungrossed amount of 33,515 euros. In 2013, the ungrossed amount was 32,996 euros and it had been of similar magnitude in both 2012 (33,749 euros) and 2011 (34,837 euros). More detailed results on the over-indebtedness of private individuals can be found in the German-language publication *Fachserie 15, Reihe 5* on over-indebtedness statistics.

It should be noted, however, that even the grossedup figures do not provide information on the absolute number of over-indebted individuals in Germany. First, not every person counselled is really overindebted. In some cases, there are just minor financial difficulties, so that the number of over-indebted people is overestimated. Second, not every over-indebted person seeks help from a counselling agency, so that the total of cases is underestimated. Nevertheless, the grossedup results of the over-indebtedness statistics for the first time provide information which is representative of all the counselling cases in Germany, instead of just the cases reported. However, this leads to results – at least at a more detailed level – which cannot be released because they are affected by the very narrow data basis in some Länder and the accompanying standard errors. This problem can only be solved by achieving higher participation rates in particular in those Länder where participation is exceptionally low. In some Länder, the provision of public funding to the counselling agencies has already been linked to their participation in the statistics on over-indebtedness. In these Länder, the relevant Land statistics can at least partly be derived from federal statistical data, so that duplicate surveys can be avoided. The statistical burden imposed on the counselling agencies can also be reduced in this way. Let us hope that additional Länder will follow suit.

Walter Joachimiak, tel: +49-(0)611 / 75 26 89, e-mail: walter.joachimiak@destatis.de

Events

Federal Statistical Office grants the 2015 Gerhard Fürst Award at the 24th Scientific Colloquium "Understanding Statistics – Orientation in the Information Society"

⇒ "The sexy job in the next ten years will be statisticians"
⇒ "Do physicians understand health statistics?"
⇒ "The 'Unstatistik des Monats' (bad statistics of the month)"
⇒ "Orientation in the information society"
⇒ "Statistics-theatre"
⇒ "Telling figures"

Those are just some of the buzzwords used on the occasion of the 24th Scientific Colloquium on the topic "UNDERSTANDING STATISTICS – Orientation in the information society", held in Wiesbaden on 19 and 20 November 2015.

More than 200 participants attended the conference which was organised jointly by the Federal Statistical Office and the German Statistical Society (DStatG).

Sibylle von Oppeln-Bronikowski, Head of Department at the Federal Statistical Office, welcomed the participants and emphasised the importance of the conference topic and the indispensability of data literacy in most of the rather demanding professions today, as well as the ensuing high significance of improving and promoting each individual's data literacy. She highlighted the importance of understanding words and figures, taking the timely topic of "refugees" as one example.

Prof. Dr. Walter Krämer (Dortmund University, Faculty of Statistics) moderated the colloquium and also introduced the topic. In his introduction, Prof. Krämer touched upon the most frequent misconceptions and barriers to comprehension (such as the incorrect use of percentages and probabilities or the confusion between correlation and causation) and illustrated them with the help of interesting examples.

With Prof. Dr. Gerd Gigerenzer, Director at the Max Planck Institute for Human Development and the affiliated Harding Center for Risk Literacy, Berlin, and Prof. Dr. Walter Krämer of TU Dortmund University, two of the most renowned authors of scientific books on statistics were attracted for the colloquium. Those two university professors wrote a book together with Prof. Dr. Thomas Bauer from the RWI in Essen (Rhine-Westphalia Institute for Economic Research) and Ruhr University Bochum entitled "Warum dick nicht doof macht und Genmais nicht tötet - über Risiken und Nebenwirkungen der Unstatistik" (Why being fat does not make you stupid, and why genetically modified corn does not kill - risks and side effects of bad statistics), which they presented at the Dialogue Forum of the Berlin Information Point in October 2014. It was a great honour to be able to welcome all three authors as speakers at the colloquium.

Prof. Dr. Gerd Gigerenzer showed in his contribution that studies carried out in Germany and other countries indicate that 70 - 80% of all physicians do not understand health statistics, and that physicians may be strongly influenced in their assessment of tests and treatments by the way in which the statistics are presented. He presented pertinent studies and explained

principles of an efficient risk communication. "Better medical education can mean better health care for less money", he emphasised.

Prof. Dr. Thomas Bauer presented the "Unstatistik des Monats" (bad statistics of the month) in his contribution. Under this label, Prof. Dr. Gigerenzer, Prof. Dr. Krämer and Prof. Dr. Bauer have been dissecting recently published statistics as well as their interpretations since 2012. In the presentation, two "turbo generators" that are responsible for the emergence of "Unstatistiken" (bad statistics) were explained in detail: the use of rankings and claiming causal relationships where there is only correlation.

Further renowned representatives from the media (Dr. Philip Plickert, FAZ Frankfurter Allgemeine Zeitung GmbH, Frankfurt am Main), didactics (Prof. Dr. Rolf Biehler, University of Paderborn), social sciences (Anouk Zabal, GESIS - Leibniz-Institute for the Social Science, Mannheim), statistical consulting (Katharina Schüller, STAT-UP Statistische Beratung & Dienstleistungen, München), and culture (Sebastian Brünger, Rimini Protokoll, Berlin) gave speeches at the Wiesbaden Museum on the question as to how to make statistical information easier to understand for the public at large.

The Federal Statistical Office was represented by Sibylle von Oppeln-Bronikowski and Kerstin Hänsel as lecturers. Sibylle von Oppeln-Bronikowski, together with Christoph Bergmann, presented various forms of visualisation using a number of practical examples and showed new ways of storytelling in her speech entitled "Telling figures. Understanding statistics."
Kerstin Hänsel introduced a new online service of the Federal Statistical Office in her speech entitled "Welcome to the StatistikCampus!", whose purpose is to provide easier access to official statistical data for students in higher education.

Prof. Dr. Krämer not only moderated the event and held the opening speech but also gave the laudatory speeches in his function as chairman of the expert jury during the formal ceremony in which the Gerhard Fürst Awards were presented, i.e. the research awards granted by the Federal Statistical Office for outstanding scientific contributions closely related to official statistics.

In the category "doctoral theses", Dr. Christian Troost was granted the 2015 Gerhard Fürst Award for his contribution entitled "Agent-based modelling of climate change adaptation in agriculture: A case study with MPMAS for the Central Swabian Jura". The award-winning doctoral thesis had been supervised by Prof. Dr. Thomas Berger of Hohenheim University. The award money in the "doctoral thesis" category amounts to 5,000 euros. In the "Master's/Bachelor's thesis" category, the experts chose the Master's thesis of Philip Rosenthal on "Optimising algorithms to estimate robust spatial small area models", which had been supervised by Prof. Dr. Ralf Münnich by Trier University. The Gerhard Fürst Award money in this category amounts to 2,500 euros.

The laudatory speeches delivered by Prof. Dr. Walter Krämer on the award-winning papers were published in the June 2015 issue of the journal "Wirtschaft und Statistik" (Ausgabe 6/2015 der Zeitschrift „WISTA – Wirtschaft und Statistik"). In early 2016, the award winners will publish detailed articles on their papers in that journal.

 Short versions of the award-winning papers and more detailed information on the presentation of the Gerhard Fürst Awards can be found on the Federal Statistical Office's website. The documentation comprising all abstracts of the contributions made to the colloquium as well as the accompanying presentations is available from the theme page Kolloquien.

Birgit Baptistella, tel: +49-(0)611 / 75 26 03, e-mail: birgit.baptistella@destatis.de