

METHODS – APPROACHES – DEVELOPMENTS

Information of the German Federal Statistical Office

Edition 2/2013

The catchword

Consumer price index rebased to new base year 2010 3

Methods of federal statistics – Further development

User conference on tax statistics: use and prospects of tax statistics for policy advice
and science 5

How to report statistical data – the (more) easy way 7

Methodological enhancement of the indices of agreed earnings and working hours
extends the range of data offered 11

New SAS tool to identify the main causes of asymmetries in foreign trade statistics 13

Defining potentially environmental goods on the basis of the GP 2009 classification
(combined in the 'environmental goods list') as another measure to identify the group
of respondents of the supply-side environmental sector 15

Published by: Federal Statistical Office, Wiesbaden

Subject-related information
on this publication:

Division B 2,
tel: +49 (0) 611 / 75 20 77
fax: +49 (0) 611 / 75 39 50
institut@destatis.de

General information
on the data supply:

Information service,
tel: +49 (0) 611 / 75 24 05
fax: +49 (0) 611 / 75 33 30
<https://www.destatis.de/.../kontakt>

**Release calendar
of the Press Office:**
<http://www.destatis.de/.../Terminvorschau>

Periodicity: (generally) twice a year
Published in December 2013

© © Federal Statistical Office, Wiesbaden 2013

Reproduction and distribution, also of parts, are permitted provided that the source is mentioned.

The catchword

Consumer price index rebased to new base year 2010

The consumer price index (CPI) is subject to regular revisions at five-year intervals. This includes updating the weighting patterns and implementing methodological changes. In February 2013, the Federal Statistical Office published the results referring to the new base year 2010. In Germany, this involves recalculating the consumer price index as from the beginning of the base year (from January 2010). Since the results have been published with reference to the new base year 2010, users will have comparable time series again for at least three years, without structural shifts or methodological changes influencing them. In addition, the recalculation allows comparing the previous results with the new ones. In *Wirtschaft und Statistik*¹, issue 05/2013, the differences are shown and the main backgrounds explained. On the whole, the recalculation leads to a slight reduction of inflation rates on an annual average for 2011 and, consequently, to a correction by -0.22 percentage points on an annual average. For 2012, the inflation rate was confirmed (the correction amounts to -0.02 percentage points). The differences for the individual months of 2011 and 2012 vary considerably.

Introduction of the 2010 weighting basis

The regular revision focuses on updating the weighting pattern for goods and services. The weighting pattern reflects the spending pattern of 2010 for all products in the total consumption expenditure of households in Germany. For such updating, the consumption habits of households for the calendar year of 2010 were ascertained. A major basis for this calculation was the results of the household surveys, which comprise the sample survey of income and expenditure of 2008 and the continuous household budget surveys until 2010. Possible undercoverage was offset by using tax statistics (e.g. for tobacco and alcohol). Further differentiation was achieved by using a variety of other statistics (e.g. health monitoring system, vehicle licensing statistics of the Federal Motor Transport Authority). In addition, non-official statistics were used such as those compiled by business associations or results of market research institutes and queries with associations and big companies. National accounting results were applied to adjust the results of various sources to the concepts of consumer price statistics, for example, to the domestic concept.

Introducing the 2010 weighting basis involves recalculating the weights of individual shop types and the weights of the Länder. The shop types reflect the importance of different retail shops such as specialised dealers, discount stores or mail order business. Major bases are the results of retail statistics and market research data on turnover distributions in retail trade.

The weights of the Länder show their share in private consumption expenditure in Germany and are based on the results obtained by the national accounts working group of the Länder regarding final consumption of households, which were produced in 2009. This is why those results were adjusted to prices of 2010, using the Land-specific development of consumer price statistics.

The effects of updating the weighting patterns are the reason for many of the differences between recalculated and previous results for 2011. Here a downward correction by 0.18 percentage points occurred on an annual average. For 2012, the downward correction amounts to 0.05 percentage points. Most differences can be explained by the readjustment of the weights for all goods and services. The effects of the updated weights of shop types are rather small. The main reason is that only about a third of the consumption expenditure is affected by shop type updating. The effect reweighting has on the Länder is negligible.

¹ Egner, U.: *Verbraucherpreisindex auf neuer Basis 2010* in *Wirtschaft und Statistik* 05/2013, pp. 329ff.

Methodological improvements

The regular revision involved in rebasing the consumer price index to 2010 was also used to implement methodological improvements in the index. A major goal in this context is to apply internationally accepted methods that are comparable on a European level when producing the results. Such methodological steps taken in the revision included the following:

- extending the coverage of seasonal products, and
- modifying the way of covering package holidays and holiday dwellings.

The extended coverage of seasonal products is based on an EU regulation on the harmonised index of consumer prices (HICP)¹. This regulation applies to the HICP and was implemented for the CPI with the introduction of base year 2010. According to those rules, products are included which are not available, or whose purchased volumes are small or negligible, for specific periods in a typical annual cycle. This requirement made the definition of seasonal products more narrow, but stricter criteria were applied to include them in the index calculation. In Germany, this had led to an extended coverage for the food groups of fish, fruit and vegetables. In addition, the implementation rules for the regulation include specific updating methods applying to seasonal products not only in the food sector but also in the clothing and footwear sector. In the HICP, the regulation on seasonal products started having an effect on the index as from January 2011. Information on the implementation is given in articles contained in an earlier edition of this publication and in the periodical *Wirtschaft und Statistik*². These articles also refer to the implementation in the CPI and contain not only a description but also first estimates of the impact the methodological changes have on the results. The effects which the CPI adjustment to the HICP has on the CPI results are very small, although they can be identified. The corrections differ between the months. For 2011, there is an annual average difference of less than +0.01 percentage points and for 2012 a difference of -0.01 percentage points.

The modified way of collecting data on package holidays and holiday dwellings should take better account of the concept of consumption segments³. In this context, for package holidays, selecting a survey item has been linked to the place/region of destination or to a concrete offer. In the past, the decisive criterion for selection or replacement was the link with the travel agent; this is no longer the case. Price collection for holiday dwellings has markedly been extended. Small providers are more widely covered now and all holiday regions in Germany are covered.

The improved sample has led to weaker seasonals, in particular for holiday dwellings. The overall effect of the modification regarding holiday dwellings in the CPI, however, is small because of the small weight of expenditure in this area. However, the improved sample of package holidays and holiday dwellings has a larger effect on individual months, which is due to the seasonal price development and which explains the differences in overall monthly inflation rates. On an annual average in 2011, the CPI is corrected by -0.03 percentage points and for 2012 by -0.04 percentage points.

The regular revision also included national updating and improvements in price collection. Based on the market development, a more detailed breakdown of the price development in education, for example, was established. Price observation was extended especially for private education providers (such as private schools, private institutions of higher education, private lessons), so that it is now possible to show some of these education price developments separately. Due to the new sample structure, the strong price decrease of the last few years is slightly slowed down.

¹ Commission Regulation (EC) No 330/2009 of 22 April 2009 laying down detailed rules for the implementation of Council Regulation (EC) No 2494/95 as regards minimum standards for the treatment of seasonal products in the Harmonised Indices of Consumer Prices (HICP).

² Elbel, G.: "Including seasonal products in the Harmonised Index of Consumer Prices from January 2011" in *Methods-Approaches-Developments* 2/2010 pp.14f. and Elbel, G.: *Behandlung saisonaler Erzeugnisse in der deutschen Verbraucherpreisstatistik* in *Wirtschaft und Statistik* 11/2010 pp.1022ff.

³ Commission Regulation (EC) No 1334/2007 of 14 November 2007 amending Regulation (EC) No 1749/96 on initial implementing measures for Council Regulation (EC) No 2494/95 concerning harmonised indices of consumer prices.

This leads to average corrections of +0.01 percentage points for 2011 and of +0.04 percentage points for 2012.

Other changes refer, among other things, to an improved representation of the price development in health care and further updating of the sample, for example, by a marked extension of the selection of reporting units in the mail order business (including e-commerce). The various measures are complex, their effects are small to negligible. On the whole, however, the effects of the other “small” changes explain the remaining difference (“residual” of the overall correction).

As consumer price statistics in Germany comprise the calculation of various indices which are based on data of a common price collection system, the measures taken in the regular CPI revision do have an impact on other indices. For example, the early implementation of the EU regulation on seasonal products in the HICP affects the HICP inflation rates in 2011 (but not in 2012). The national methodological changes will not have an impact on the HICP until 2013, which is due to chain-linking (without back-calculation). Apart from the time shift, there are differences to different extents. For example, the retail price index is affected only by changes implemented for goods.

Ute Egner; tel: +49 (0) 611 / 75 41 55; e-mail: ute.egner@destatis.de

Methods of federal statistics – Further development

User conference on tax statistics: use and prospects of tax statistics for policy advice and science

1. Introduction

Official tax statistics are one of the most important data bases that policy-makers can use to estimate the fiscal effects of tax law amendments. Based on forecasts by the Working Party on Tax Revenue Estimates, financial estimates of those effects are incorporated into budgetary planning at national, Land und municipal level. Likewise, scientists are happy to use the wide range of tax statistics data for research purposes.

To meet the requirements of official statistics users more effectively, the Federal Statistical Office regularly organises expert committee meetings. The aim is, on the one hand, to inform the users about current developments and, on the other, to enable them to influence the content and design of statistics.

Last year's expert committee meeting was held in Berlin as a two-day user conference and was organised in cooperation with Humboldt-Universität zu Berlin und the Federal Research Data Centre. On 11 October 2012, Dr. Sabine Bechtold (Federal Statistical Office) und Prof. Dr. Ralf Maiterth (HU Berlin) welcomed about 100 participants in the Senate Hall of Humboldt-Universität. These representatives from the political and scientific communities, associations and official statistics agencies came together to discuss the status und prospects of tax statistics with a view to policy advice und scientific analyses.

2. Policy advice

The first conference day focused on the use of tax statistics for purposes of policy advice. In their presentation, Dr. Ulrich van Essen and Gregor Schlick (both of the Federal Ministry of Finance - BMF) drew attention to the importance of estimating the effects of tax policy decisions on public budgets. Before the Tax Statistics Act was amended on 11 October 1995, BMF had no access to microdata of official tax statistics. Analyses could only be made using aggregated tables from the Federal Statistical Office's subject-matter publications and by applying simulation models that were based on those tables. Apart from enabling the Federal Statistical Office to carry out comprehensive special evaluations, the availability of new microdata permits to develop microsimulation models. Dr. Sven Stöwhase (Fraunhofer-Institut) presented the microsimulation model on income tax (MikMod), which is run by the Fraunhofer-Institut on behalf of BMF.

In his talk, Prof. Dr. Joachim Merz (Leuphana Universität Lüneburg) stressed the importance of tax statistics as a microdata basis. Using the Federal Government's Reports on Poverty and Wealth, he underlined the informational value of the data for very high incomes and self-employed incomes. Prof. Dr. Henriette Houben (HU Berlin) described how inheritance tax statistics can be used to obtain a nearly complete picture of the probable effects of inheritance tax reform projects. Against the background of the ongoing discussion on a municipal finance reform, Prof. Dr. Michael Broer (Ostfalia-Hochschule für angewandte Wissenschaften) set out the enormous potential of the data from tax statistics.

Finally, Klaus Wolter (BMF) drew attention to the future demands on tax statistics and science. He cited the "Common Consolidated Corporate Tax Base" project, whose aim is to harmonise the taxation of corporate profits at European level, as evidence of the continuously increasing importance of tax statistics.

3. Science

Several contributions presented research projects based on data from tax statistics. Dr. Alexander Vogel (Statistikamt Nord) used the turnover tax panel to examine the dynamics of the import and export involvement of industrial companies in Hamburg und Schleswig-Holstein. With the help of the taxpayer panel, which is based on income tax statistics, Maja Adena (FU Berlin) set out to identify changes in the patterns of charitable giving by those liable for income tax after the tax reforms of 2004 and 2005. Georg Struch (Universität Potsdam) described how he linked the de-facto anonymised sample of income tax statistics (FAST) with the socio-economic panel (SOEP) in the EITDsim microsimulation model. Jost Henrich Heckemeyer (Universität Mannheim) explained that ZEW (Centre for European Economic Research) was using turnover and corporation tax statistics to transfer the results from a microsimulation model used to analyse corporate taxation in Germany in approximate form to the German business population.

4. Production by the statistical offices

As far as the production of tax statistics is concerned, a variety of innovations have been made in the last few years. As outlined by Stefan Dittrich (Federal Statistical Office), the multi-annual federal statistics were complemented by annual administrative statistics. He explained that work was underway to set up an Integrated System of Tax Statistics (StSys). The aim is to have annual federal statistics available according to the central production and data management concept. This could make it possible to rationalise processing and to achieve synergy effects through the application of cross-sectional statistical methods. The supply of data which are mostly contained in tables was expanded by various forms of access to microdata, such as, in particular, de-facto anonymised data for researchers, microdata accessible through remote data execution, panel data and linked-up data of turnover, trade and corporation tax statistics. The two latter topics – panel data and the linking of statistics – were further elaborated in separate contributions by Ulrike Gerber and Natalie Zifonun-Kopp (both of the Federal Statistical Office). Apart from that, Natalie Zifonun-Kopp described considerations concerning additional statistics on the e-balance sheet. She explained that model calculations on tax law amendments for policy-makers and the

complementation of official data by cartographic results down to municipal level were new focal topics for official statisticians.

The conference then focused on the availability of tax statistics. While access to microdata is specially regulated by law for the federal and Land ministries of finance, scientists can only access official microdata through the research data centres, as was pointed out by Prof. Dr. Markus Zwick (Eurostat) and Rafael Beier (Federal Research Data Centre). Statistical confidentiality plays a central role in this context because especially sensitive data on individuals and businesses are processed and published in tax statistics. The contribution by Juliane Gude and Sarah Giessing (both of the Federal Statistical Office) discussed confidentiality procedures for tables. Eric Schulte Nordholt (Statistics Netherlands) rounded off the topic with his paper on data access for researchers in the Netherlands.

5. Results and outlook

At the user conference, the main players involved in official tax statistics, i.e. the fiscal authorities, data producers and the scientific community, presented their current work and ideas for future activities. The juxtaposition of different interests promoted mutual understanding. The users stressed the major improvements in the provision of tax statistics for policy advice and research in the last ten years. Development potential was identified, in particular, in the further linking of longitudinal and cross-sectional tax statistics, the statistical editing of the e-balance sheet and the accessibility of microdata for researchers. Also, there is some interest in even more timely data and in person-related information of wage and income tax statistics instead of information at the level of individuals liable to pay tax. The policy-makers were asked to incorporate and extend microdata accessibility for researchers in the Federal Statistics Law, while the policy advisers hope that researchers will find new and robust forecast models.

The conference programme and the individual contributions can be found on the Federal Statistical Office's homepage at <https://www.destatis.de/DE/UeberUns/Veranstaltungen/VeranstaltungenArchiv/VeranstaltungenArchiv.html#doc52436bodyText2>

Ulrike Gerber, tel: +49 (0) 611 /75 23 85, e-mail: ulrike.gerber@destatis.de

Stefan Dittrich, tel: +49 (0) 611 /75 41 24, e-mail: stefan.dittrich@destatis.de

How to report statistical data – the (more) easy way

Electronic alternatives to the paper questionnaire

For many reporting units, getting a letter from a statistical office means unwelcome extra work and bureaucracy.

Therefore, for a long time already, the statistical offices have made efforts to further reduce the burden involved for businesses and public authorities when complying with their statistical reporting obligations. The main tool they use in this context is safe online reporting tools, which will soon replace the traditional paper questionnaires. The current practice of using a multitude of different reporting channels and data formats will soon be entirely replaced by the two standardised online reporting tools of official statistics: IDEV and eSTATISTIK.core.

The IDEV tool provides browser-based online forms for the secure transmission of statistical data. As an online reporting tool, the eSTATISTIK.core package of technological solutions offers several

ways to achieve a largely automated process of obtaining, compiling and transmitting statistical data.

Both online reporting tools transmit the same standard XML data format (DatML/RAW) and thus ensure that the statistical data are directly further processed without switches between media using standard tools in the system of official statistics. The DatML/RAW data format used here is part of XStatistik, the XÖV certified and standardised data exchange format of official statistics. XÖV stands for *XML in der öffentlichen Verwaltung* (XML in public service).

The planned amendment of the Federal Statistics Law will oblige public authorities and businesses to report their data online by using the online reporting tool offered and, consequently, to use the standard format. The only exceptions are public authorities which use a different XÖV format for data exchange.

This legal obligation of the reporting units is in line with the strategy of official statistics to extend the utilisation of online reporting tools. The statistical offices consider it a commitment to offer such tools for as many surveys as possible. For every survey, at least one secure online reporting tool should be introduced. The decision as to which tool is offered should be based on how a maximum reduction of the burden on reporting units can be achieved. The reduction is achieved mainly by automation. A second means is to increase the utilisation of online tools. This can be achieved – apart from the legal obligation – by continuously optimising the reporting tools. Generally, paper questionnaires will no longer be sent together with the recruitment notification.

eSTATISTIK.core – using data that are available electronically

The idea behind eSTATISTIK.core is quite simple. In the majority of reporting units, the data requested for statistics are already available in an electronic form in their software systems. Thus the main requirement for using eSTATISTIK.core is met because this is exactly where the starting point of the reporting channels of eSTATISTIK.core is. The electronic data available in the reporting unit are generated and transmitted directly and in an automated way where possible.

The highest level of automation in eSTATISTIK.core is achieved by implementing into the applied software a module for the generation of the data package to be transmitted. Another component generally integrated in addition is the CORE.connect software library, which has been developed by the Federal Statistical Office. It is provided free of charge and serves as a communication interface. Using the statistics module with these features, the reporting unit can directly generate the statistical data in its software system, compile them into a data package, perform first checks and transmit them in an encrypted form through a secure internet connection to the central data portal of the statistical offices.

There the data are checked again and immediately passed on to the responsible statistical office for further handling without switches between media. At the same time, the reporting unit gets an electronic check log.

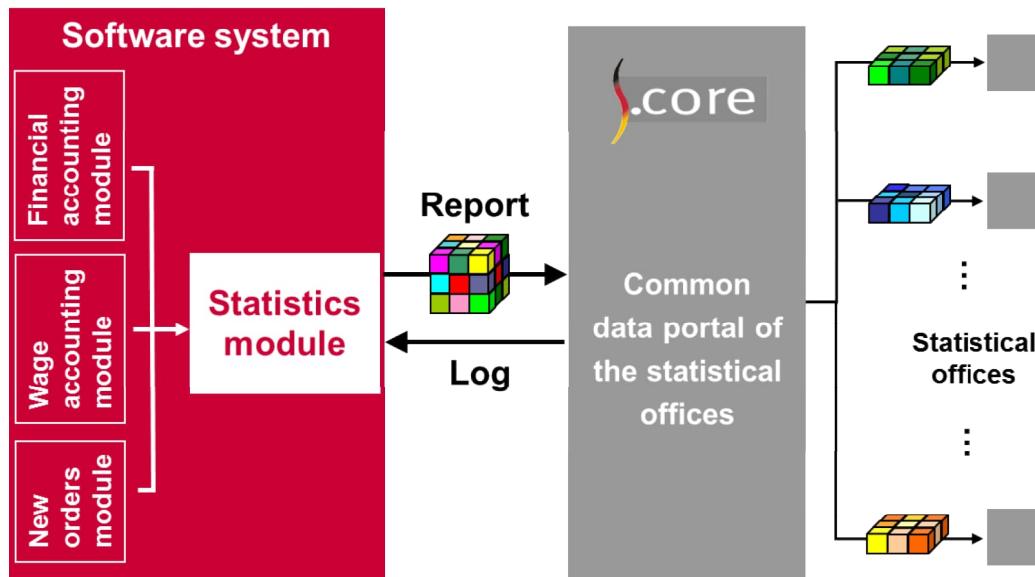


Figure: The reporting channel when using eSTATISTIK.core in an enterprise

eSTATISTIK.core provides synergy effects for reporting units which have to provide data for several surveys and/or for several establishments which may be located in different Länder. This is because the online reporting tool has been designed also for data packages containing several statistical reports from different reporting units for several recipients (statistical offices).

Many software producers have integrated statistics modules into their products or are willing to do so upon request by their customers. Reporting units whose software has just a statistics module without a communication interface can upload their statistical reports – created in the DatML/RAW format – through the web application of eSTATISTIK.core. If the reporting units has no statistics module in its software, the PC application CORE.reporter, which is provided free of charge by the statistical offices and has a graphical user interface, is another solution for electronic online data transmission to reduce the response burden.

eSTATISTIK.core has been in use since 2005. In 2012 alone, just under 69,000 reports with a total of about 12,700,000 data records were transmitted for 33 sets of statistics. For another 29 sets of statistics, the required subject-related specifications have been provided by the statistical offices, while a statistics module has to be supplied by a software provider. For many other statistics, subject-related preparations are now being made to offer eSTATISTIK.core as a reporting channel there, too.

eSTATISTIK.core – the (more) easy way to report statistical data

For wider use, eSTATISTIK.core is continuously optimised. A CORE.connect version released in 2012 provides many improvements for reporting units, software providers and official statistics.

Modern reporting tools allow checking the entered data through the user interface of the relevant application. The checks applied generally refer to the requirements of official statistics. The purpose is to avoid queries having to be made by the statistical offices after receiving the data. Depending on how the software has been arranged by the software provider, the data are checked in the individual dialog masks to see whether they meet formal and/or subject-related criteria. With the current version of CORE.connect, it is no longer necessary for the software provider to create the format of the data to be transmitted. Through the DatML/RAW generator provided, a data package in the DatML/RAW format can now be generated with minimum effort. Basically, the programming work for software development is limited to compiling the data.

In addition, for formal checks, the software provider can use the CORE.inspector component, which is also provided. In this way it is possible even before transmission to subject a whole data package to the same checks which are performed after transmission by the statistical offices at

the central data portal. This allows the reporting unit to ensure before transmission that the data package is correct from formal aspects.

Apart from formal checks, further checks are performed at the central data portal. A total of four check stages are distinguished.

Upon data receipt: automated checks		
Stage 1:	Syntax	In the syntax check, it is checked whether the DatML-RAW file received meets formal technical criteria. If the criteria are not met, the XML file cannot be read by the system.
Stage 2:	Semantics	If there are no syntax errors, the semantic check is performed to check the logical structure of the file. This is necessary for the further automated processing of the file.
Stage 3:	Authorisation	If the file can further be processed, the user ID, password and other things are checked in the next step. Where applicable, the reporting unit ID is checked in addition.
Stage 4:	Formal requirements	At the last stage, the reported statistical data are checked for formal correctness.
Subsequently: check of contents by the responsible specialised department		

Checking the reporting unit identifier

The reporting unit identifier is an identifier assigned by the statistical offices in a unique way to any reporting unit for a specific survey period. One of the purposes the identifier is needed for is to record – as part of the checks upon data receipt – the fact that a reporting unit has reported its data. When the data have been received, the reporting unit ID can be matched with the reference database in an automated way to check whether the reporting unit ID contained in the data report corresponds to a reporting unit in the recipient Land indicated. If the check shows that the reporting unit ID does not correspond to any reporting unit, the report cannot be further processed and is rejected. If the recipient Land indicated in the report does not correspond with the recipient stored in the reference database, the recipient is corrected and the report is passed on to the correct office. This avoids sending a report to an office which is not responsible.

Formal checks versus subject-related checks

Formal checks are to ensure that the quality of the received data allows subject-related checking for plausibility and further processing. In such automated and purely technical checks, it is checked whether the received data meet the subject-related requirements that are laid down in the statistics-specific data supply agreement. In the formal check, it is verified whether the data are within the value range permitted for the relevant variable (e.g. the number of persons employed by an enterprise cannot be negative) or whether all compulsory variables are contained.

In the subject-related check, which cannot run in an entirely automated way, the data are checked for subject-related plausibility by the responsible specialised unit. Implausibility may be suggested by the relationship between individual variables (if an enterprise has employees, it has to pay wages and salaries) or by complex calculations.

Extending the check logs

After the data have been checked, check logs are provided at the central data portal. In the past, the logs could not be provided for download until about 20 minutes after data receipt, which was due to some long runtimes. Due to the extensions released in 2012, the reporting units can now view the check logs in the application directly after transmitting the data. In this way, they get direct feedback on the statistical report they have sent.

To show the results of the checks in a more transparent and comprehensible way, the check logs were completely revised and additional information was included. When implementing the formal checks, care was taken in particular to ensure that comprehensible notes – and possible solutions where applicable – are available for the error codes in the log. In addition, the structure was changed to show the information in a clearer way.

Quality improvement through the rejection of reports

Reports which cannot be processed due to their poor quality can now be rejected. By rejecting such reports directly at the data portal, it is ensured that the data which are admitted to content-related processing are really suitable for processing. This avoids a large number of queries having to be made by the statistical offices. Information on whether a report has been accepted or rejected is contained in the check log. The reporting unit thus can directly correct the report. If a query were made when a faulty report is already undergoing content-related processing, things would be much more complicated for all parties involved.

Outlook

First experience regarding the above new features in individual surveys show that the quality of incoming data reports has considerably improved. The goal of the next step is to extend the range of online reporting tools and to apply the new functions and features of eSTATISTIK.core in all areas.

More detailed information is available at: <http://www.statspez.de/core>

The CORE web application is available at: <https://core.estatistik.de>

André Orth, tel: +49 (0) 611 / 75 43 77, e-mail: andre.orth@destatis.de

Marion Rizzo, tel: +49 (0) 611 / 75 47 32, e-mail: marion.rizzo@destatis.de

Methodological enhancement of the indices of agreed earnings and working hours extends the range of data offered

The indices of agreed earnings and working hours measure the average change in agreed monthly and hourly earnings of employees and in weekly working hours. They show the general development of agreed remuneration and are part of the indicators of the dissemination standard of the International Monetary Fund (IMF). In their current form, the indices provide information about the part of agreed earnings that has to be paid permanently and regularly. As the calculation is based on a constant weighting pattern and as pay increases are usually agreed upon for a longer period, the indices are relatively non-sensitive to short-term economic fluctuations. This is why the indices of agreed earnings are used by many transport associations and energy suppliers for stable-value clauses in long-term contracts. Other users are trade unions, employers' associations and ministries. Further potential users, such as research institutes and the national accounting system at the federal level, often use alternative indices because any agreed earnings received irregularly have so far not been included in the indices described here.

To illustrate the influence of irregular payments on the development of the indices of agreed earnings, reference is made here to the collective agreement reached in the chemical industry in 2010. Due to the economic crisis, the social partners agreed on a one-off payment of at least 550 euros per employee instead of a percentage increase. In the past, this payment was not included

in the relevant index, so that the index did not show any change after the agreement had been reached. As more and more collective agreements included one-off payments, the coverage of the indices of agreed earnings decreased accordingly.

Therefore, indices of agreed earnings including extra payments were published for the first time in May 2013. They now take account of the changes in irregular payments such as one-off payments, vacation and Christmas bonuses, contributions to employee savings schemes and pension schemes. These additional indices considerably extend the range of information provided by the indices of agreed earnings and working hours.

The methodological extension was implemented in the context of reprogramming the indices. The latter had become necessary because a decision taken by the statistical offices of the Federation and the Länder in 2008 said that the main Adabas/Natural applications would have to be replaced within five years. The new IT application – the database of the indices of agreed earnings and working hours (agreed earnings indices database) – was implemented by using only open-source products and applying a 3-tier architecture. The front end employed is a Java Rich Client using the OVIS framework in the system of official statistics. The application server is JBOSS, the database system is MySQL.

Parallel to that, the indices of agreed earnings were rebased, as is done regularly. The basis for the new weighting pattern of the indices – that is, for the selection of collective agreements included and of the number of employees – is the results of the 2010 structure of earnings survey and of the survey of earnings in agriculture for September 2010. In these surveys, selected establishments and holdings are asked, among other things, whether they pay their employees according to a collective agreement and, if so, according to which pay grade each of them is paid. As each of the establishments and holdings covered is classified to a specific economic branch, detailed information on the application of collective agreements is available for every branch. To obtain weights for the indices, a number of collective agreements are selected for every branch in a way to cover at least 75% of all relevant employees. The remaining employees covered by collective agreements are distributed proportionally to the selected collective agreements. This ensures that the weight of every branch in the total is not influenced by the selection of collective agreements. To make sure that the agreed earnings indices take account of different agreed pay grades in accordance with their frequencies, five pay grades per collective agreement are selected.

The company and regional collective agreements needed to calculate the agreed earnings indices are obtained from the Federal Ministry of Labour and Social Affairs, the remuneration regulations for public officials from the relevant pay offices. Then all data that are relevant for the statistics of agreed earnings are taken from the agreed earnings database. The data required to calculate the indices are extracted in a largely automated manner from the agreed earnings database. However, it is not possible to transfer all data from the agreed earnings database to the agreed earnings indices database. For example, extra payments and pay increases coming into force with retroactive effect have to be entered manually into the IT application. As, in addition, the index values are not revised, pay increases published after the calculations have been finished will not be included until the following quarter. To keep the number of such cases to a minimum, the index calculation includes pay settlements for which collective agreements are not (yet) available but for which the percentage increases are known.

Contrary to the former indices of agreed earnings, full-time equivalents are used instead of full-time employees, that is, all employees covered by collective agreements are converted into full-time employees according to their number of working hours. The new indices of agreed earnings include the development of the remuneration of public officials. Agriculture is now covered, too. Data on the development of agreed earnings in the overall economy applying to a total of 11 million employees covered by collective agreements and public officials are now available for the first time. Another new element is the calculation of quarterly indices of agreed earnings on the basis of the monthly values. The final results are released four times a year, about 60 days after the end of the calendar quarter. The monthly values are available to users as provisional results.

The biggest challenge in further developing the methodology of the agreed earnings indices is taking account of extra payments. It must be ensured that the time of payment is represented as exactly as possible. As no relevant information is available in many collective agreements, the social partners are contacted for information on one-off payments. Vacation and Christmas bonuses are included in the calculation for June and November, respectively. Also, the information on extra payments, which differs considerably between collective agreements, has to be harmonised. If, for example, the vacation bonus is fixed per vacation day, a total amount is calculated by means of typing, which is then included in the index calculation. If, however, the data on extra payments are expressed as a percentage of the individual agreed earnings, a weighted average is calculated on the basis of the agreed earnings data stored in the agreed earnings index database.

Due to the rearrangement, the indices of agreed earnings meet European requirements. At the European Policy Conference “Collectively agreed wages in Europe (CAWIE)”, which took place in Brussels in December 2012 and dealt with the availability of statistics of agreed earnings in the euro area, the indices of agreed earnings and working hours of the Federal Statistical Office and of the Italian statistical office (Istat) were presented as best practice.

Comprehensive information on the new features of the indices of agreed earnings was published for users in the *Fachserien* and on the website of the Federal Statistical Office. To be able to offer time series even after the indices of agreed earnings have been rebased, selected WZ headings (long time series) were mathematically linked to the existing time series. The free publication *Index der Tarifverdienste und Arbeitszeiten – Lange Reihen* contains data for over 40 economic branches.

Also, comprehensive information from regional collective agreements is available to users free of charge in the online version of the agreed earnings database (www.destatis.de/tarifdatenbank).

Jörg Decker, tel: +49 (0) 611 / 75 24 42, e-mail: joerg.decker@destatis.de

New SAS tool to identify the main causes of asymmetries in foreign trade statistics

Comparisons of corresponding trade flows of partner countries in the context of foreign trade statistics often reveal rather large asymmetries¹ that may occur at an aggregated, and even more typically, at a detailed data level. The reasons for such asymmetries are manifold. They range from statistical and methodological differences in data collection in the two countries concerned to ‘errors’ occurring in data declaration or collection in one of the countries. For several years, the German foreign trade statistics unit has carried out mirror exercises to study and correct such asymmetries and to solve the relevant problems together with the foreign partner institutions. At the EU level, reducing asymmetries in the bilateral trade data of the Member States is highly important as this is a significant prerequisite for focusing more closely on export (dispatch) data in the future, while reducing the degree to which import (arrival) figures are collected. The ‘mirror’ export data of the partner countries to be provided in a data exchange system are instead to be used to a greater extent².

¹ Theoretically, the exports of a country to a partner country should correspond to the partner country’s imports from the exporting country and vice versa. In practice, however, comparisons of the data of these two trade flows often reveal discrepancies, that is, so-called asymmetries. They are the higher the more detailed the level of comparison of the commodity data.

² Mention should be made here of the SIMSTAT (Single Market Statistics) project within the European Statistical System, promoted by Eurostat, the Statistical Office of the European Union, which also has lead responsibility for it.

Despite the wide range of reasons for asymmetries, the mirror exercises conducted so far have shown that a substantial part of them are attributable to a few main causes, such as differences in the eight-digit commodity codes of the same goods or the names of the partner countries indicated by the two sides concerned. Analysing the problems more thoroughly led to the idea of applying a new approach to study the asymmetries in the bilateral trade data more systematically and to filter and quantify the underlying causes by means of an automated procedure. The focal point was to analyse precisely the asymmetries at the detailed and also aggregated data levels in order to find out whether they remain, become smaller or even disappear, thus giving an indication of their possible causes.

The goal of a related project¹ was then to design and program an algorithm using the statistical analysis software SAS in order to automate the required operations and carry them out in a reliable manner even with large quantities of data for the purpose of filtering asymmetries and indicating the most likely causes.

The focus was on the following four reasons for asymmetries that were to be integrated into the SAS tool as they were identified as the main causes during the long-term practice of mirror exercises:

- a) different commodity codes indicated by the two sides concerned (so-called misclassifications)²
- b) different names of the partner country indicated by the two sides (so-called erroneous partner country allocations)
- c) distorted relation between value and quantity data (the values indicated by the two sides differ, although the quantities are the same)
- d) different times (time lag) of reporting the relevant data (different reference months)³

All four phenomena are assumed to be a reason for asymmetries, if the given discrepancies disappear at one of the next higher levels of aggregation. For instance, a commodity code can be incorrect (item a), if an asymmetry occurs at the lowest level of classification (eight-digit level), while it does not appear at the higher six-digit or four-digit levels. An incorrect partner country (item b) can be filtered in a similar manner, if the relevant asymmetry in the bilateral trade with an EU partner country disappears at the aggregated EU level of the trade in goods regarding all 26 EU partners. If, as shown in item c), the quantity data coincide, while the value data differ, checks are made to identify incorrect values⁴. As regards time-related asymmetries (item d), the monthly data are compared with the cumulative data for the relevant three months (reference month plus previous and subsequent month) in order to identify or rule out this reason for the given asymmetries.

The SAS model algorithm is fed with (detailed) data from COMEXT, the Eurostat reference database for international trade. As a first step, all asymmetries are calculated and filtered by their magnitude. The following processes are designed to successively identify or rule out the various reasons for the asymmetries. Each operation produces tables that show the results of the checks made to identify the relevant causes. The information which indicates the probable background of the asymmetries has to be verified by contacting the partner country or, if possible, conducting a mirror study with that country. The whole SAS tool named ATIMA

¹ The project was conducted in the framework of the MEETS (Modernisation of European Enterprise and Trade Statistics) programme.

² Misclassifications turned out to be the most frequent cause of asymmetries. Many declarants seem to have difficulty in correctly allocating the eight-digit codes of the Commodity Classification for Foreign Trade Statistics, which contains more than 9,000 items. This even applies to goods like motor vehicles. Although, superficially, the classification of such goods seems to be easy, they are often listed in great detail.

³ The time when exports are declared in a country may differ from the time when they are reported as imports in the partner country. Also, a delayed declaration may cause differences in the reference months recorded for the same imports/exports.

⁴ This happens, for instance, often with respect to goods transactions in processing trade when the value of the goods concerned changes as a result of processing.

(Analyzing Tool to Identify Main Causes of Asymmetries) was developed by the foreign trade unit of the Federal Statistical Office.

The main benefit of the ATIMA model is that it can be used to discover reasons for asymmetries prior to a mirror exercise. The model is designed to process and analyse considerable amounts of data and especially suited to identify systematic and recurrent asymmetries and their causes. As it suggests possible reasons for data discrepancies in the same trade flows, these can be examined and resolved more quickly. Although mirror exercises are not made redundant by ATIMA, they will be simpler and take much less time in the future as possible reasons for asymmetries can be looked into in a more targeted manner.

A test run of the ATIMA tool with German-Czech trade in goods data, the subsequent verification of outcome using the results of several EU-wide mirror studies and the application of the model in a current mirror exercise with Poland have yielded relatively good 'matches'. In many cases, the reasons for asymmetries identified by ATIMA before the exercises took place turned out to be correct. ATIMA was also applied or is planned to be used in other reconciliation exercises, too, for instance with Hungary and Denmark.

Please contact the author of this article for more information about the SAS tool ATIMA or the detailed project report.

Joseph Steinfeldt, tel: +49 (0) 228 / 99 643 84 74, e-mail: joseph.steinfeldt@destatis.de

Defining potentially environmental goods on the basis of the GP 2009 classification (combined in the 'environmental goods list') as another measure to identify the group of respondents of the supply-side environmental sector

Preliminary remark

The environmental sector is a part of the overall economy where special goods, construction operations and services are traded which address environmental problems, are environmentally friendly and use resources efficiently. Official data obtained by the survey of goods, construction operations and services for environmental protection inform about the significance which environmental protection has for sales as an economic factor and about employment and export opportunities offered by the sector.

The challenge of the survey is to clearly identify the group of respondents, that is, local units offering goods, construction operations or services for environmental protection (environmental producers): the cross-cutting character of the environmental sector, the dynamics observed in the development of environmental technologies and the coverage of any level of specialisation attained by environmental producers make it difficult to identify the group of respondents. In contrast to other industries, it is not possible to define the group of respondents in the environmental sector only by means of the economic activity to which they belong.

As the definition of environmental protection builds on the purpose of the goods or services considered, the idea was to include the product range of a local unit as another criterion into the process of defining the group of respondents. In cooperation with institutes and associations, goods protecting the environment and the climate (environmental goods) were listed for the environmental sector in Germany in accordance with the Classification of Environmental Protection Activities (CEPA), which means that a functional definition of the environmental sector

was provided. Only if goods protecting the environment and the climate can be allocated to a statistical classification, conclusions may be drawn as to the producers of these goods.

Project goal

A list of goods which are potentially relevant for environmental protection (environmental goods list) was to be prepared for all sectors in a project as an additional measure to determine the group of respondents for the survey of the environmental sector, building upon the national Product Classification for Production Statistics (GP, 2009 edition). The selection criterion established was that the goods must have an environmental and climate protection purpose as their prime objective. Environmentally friendly goods, as they are called, were not considered in the environmental goods list. Their production, use and disposal cause less pollution or are more climate-friendly than those of comparable conventional products which furnish a similar utility, for instance more energy-efficient refrigerators or organic soap, while, however, their primary use is not one of environmental or climate protection.

The GP product types which are potentially relevant for environmental protection and have been included in the environmental goods list will be used in the process of identifying the group of respondents in the environmental sector of manufacturing, mining and quarrying.

Approach

The environmental goods list was initially prepared using a theoretical method supplemented by an empirical procedure.

The theoretical method served to assign the environmental goods listed by means of the CEPA to the product types of the GP classification. In many cases, however, the goods identified as being relevant for environmental and climate protection could not be found in the GP classification. Allocation was made difficult, on the one hand, by the fact that the goods belonging to one GP product type also serve purposes other than environmental and climate protection (problem of dual use). On the other hand, definitions of GP product types are sometimes too general so that the environmental protection purpose cannot be identified. This is the more so as the GP product classification meanwhile focuses rather on remedial than on the now more important integrated environmental goods which are low-polluting and resource-efficient.

The theoretical method also included an analysis of existing international environmental goods lists based on the Harmonised System. As the international environmental goods lists did not primarily serve statistical purposes, not all of the potentially environmental goods listed were of relevance for the national environmental goods list.

The empirical procedure linked and analysed official data from the survey of goods, construction operations and services for environmental protection with data from production statistics, using the identity number at local unit level. This made it possible for the first time to evaluate GP product types empirically with regard to their suitability for environmental and climate protection purposes. The following cases were considered in the analysis:

- eco specialists whose total turnover equals their turnover derived from sales of environmental goods,
- local units which report data for a GP product class and produce one (one-to-one) or several (one-to-n) environmental goods, and
- local units assigned to a division of economic activity that is important for the environmental sector.

Local units which report data to production statistics for more than one GP product type and data to environmental statistics for more than one environmental good (N-to-N) were not considered because the allocation was not unique. It was checked whether the GP product types considered

were of relevance for environmental protection and if so, they were included in the environmental goods list.

Result

By means of the procedure described, 250 environmentally relevant GP product types could be identified. Just a few contain environmental goods only; in most cases the GP product types are only potentially relevant for the environment due to their somewhat broader definition. The GP product types could also be assigned to the environmental domains in accordance with CEPA, so information is available from the environmental goods list also on the expected use of a good. In cooperation with the Lower Saxony Institute for Economic Research (NIW), the environmental goods list was adopted in a workshop in summer 2012.

The environmental goods list was used for the first time in reference year 2012 as a supportive measure in determining the group of respondents for the survey of goods, construction operations and services for environmental protection. The list was used in production statistics to check which local unit reports data for at least one of the potentially environmentally relevant GP product types. As there are reasonable grounds for assuming that such local units produce goods for environmental protection, these were initially covered once by the survey of the environmental sector. The results of the survey are used also for a final assessment of the environmental goods list and – if necessary – to delete GP product types from the list which received less favourable feedback.

Due to the harmonisation of classifications at European level, GP product types which may be relevant for the environment can be transferred to the “list of industrial products of the European Community” (PRODCOM list), which was developed by the Statistical Office of the European Communities (Eurostat) in cooperation with the member states. The environmental goods list has already been requested by several other member states. This means that the environmental goods list contributes to a further harmonisation of statistical methods in the environmental sector at European level.

Gesine Petzold, tel: +49 (0) 228 /99 643 82 24, e-mail: gesine.petzold@destatis.de