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

Number 2/2002

The Catchword

Cooperation in research and development

Methods of federal statistics – Further development

The Catchword

Census test: a new method for conducting a population count

Population counts provide data on the population of a country, on economic activity and on the housing situation as a basis for political and economic planning and decision-making. However, only part of the results of a population count can be regularly updated, and deviations from the actual stock are inevitable in the course of time. The major functions of a census are to determine the official number of inhabitants, to provide housing data and small area data, and to supply sampling frames and expansion frames for important sample-based statistics (like the microcensus).

Like in other European countries, potential ways of using the existing registers for the above purposes have been studied in Germany as well. As a result, a model has been developed together by working groups of the German statistical offices and the German Association of Cities and Towns, which relies on the data of existing administrative registers, a postal survey among the owners of buildings, and a detailed combination of those data. In this context, an analysis of the data of residents' registers is intended to replace the survey of all individuals. The data collected are checked for errors (double entries), and the necessary statistical corrections are made, if necessary, on the basis of related queries. Besides, the data of the residents' registers are internally combined with various data files, for instance on employment and unemployment as produced by the Federal Institute for Employment. The aim of an additional postal census of buildings and housing is to collect data on dwellings and the names of the owners or main tenants of the dwellings. By matching the data of the census of buildings and housing with the data of the residents' registers, information is generated about internal household relations. The implausibilities in data stocks which remain after the combination of various data sources are then checked and, upon related queries, internally corrected.

Based on the Act on preparing a register-based census dated 27 July 2001, a test procedure has been carried out whose survey date was 5 December 2001. A comparison of the results with the register data, and a supporting postal census of buildings and housing are intended to provide information about the quality of the method and to contribute to an enhanced method in the future.

The test procedure encompasses several components with various aims.

Checking for multiple registration in residents' registers

With the aim to check for multiple registration (more than one main place of residence) in residents' registers, all citizens are covered who were born either on 1 January, 15 May or 1 September and, in addition, those citizens whose date of birth is not completely recorded in the given residents' register. This corresponds to a sample size of 1.5% of the population. The Federal Statistical Office centrally checks the data records for double registration. The Land statistical offices settle the unclear cases by contacting the citizens concerned.

Testing the accuracy of residents' register data by comparing them with the data obtained through conventional interviews

A total of approximately 38,000 buildings with about 250,000 dwellings and 550,000 persons in about 570 communities of different size have been selected in two-stage sampling.

The people covered in the sample are interviewed in the conventional manner by interviewers. At the same time, the data recorded on them in the residents' registers are collected from the communities. The information obtained through the conventional interviews and the data stored in the registers are compared with the aim to reveal the extent of over-recording or under-recording in the residents' registers. Taking into consideration the expected areas of difficulty of a register-based census, the

sample survey focuses on larger towns and there on big buildings so that it cannot be regarded as representative.

Testing the new procedures in a subsample

Actually, 16,000 buildings (in 230 communities) of the total of 38,000 buildings included in the sample have been selected for the subsample. The aim of the procedure is to check whether the owners of buildings can provide adequate information on names and the buildings and dwellings concerned and to compare those results with the data provided in interviews by the owners or main tenants of the dwellings. The procedure includes the components that are outlined below.

In the context of testing automatic household generation, the data records obtained from the residents' register for the buildings selected are processed to generate households for each building and are allocated to concrete dwellings based on the names of the owner or main tenant of a dwelling.

A combination check is aimed at discovering inconsistent results for buildings and also hidden register errors. Another goal is to ascertain the effort that would be required for improvements in the quality of statistical data by purposive interviewing in the context of a register-based survey.

To obtain data on economic activity, personal data are requested from three administrative registers of the Federal Institute for Employment for the addresses of the buildings covered in the subsample. By means of variables such as addresses, names and birthdays, those personal data are linked to the personal data records (which have been grouped to form households) of the residents' registers.

Protection of data privacy

The concept of the census test is strictly in compliance with the provisions of the Federal Constitutional Court's population census judgement of 15 December 1983. The data, which are subject to statistical confidentiality, remain in protected areas of the statistical offices. Auxiliary variables such as names and addresses are deleted as soon as possible.

Hopefully, a register-based census will considerably reduce the overall costs of a population count as it makes use of existing administrative registers and the required results do not have to be obtained in an expansive manner by surveying all individuals. The new census model includes primary surveys only in the form of postal surveys among the owners of buildings, of (mostly postal) enquiries to settle cases of multiple registration, and of purposive enquiries in the case of implausible results for individual buildings. And finally, programming costs are included in the overall costs of the test.

In accordance with the current schedule, the analyses of the test results should be completed about 18 months after the survey date of 5 December 2001.

Hans-Gerd Siedt, tel: (+49-611) 75 28 45, e-mail: hans-gerd.siedt@destatis.de

Cooperation in research and development

Research Data Centres in official statistics

The "Memorandum zur Lage der empirischen Wirtschaftsforschung in Deutschland" (Memorandum on the situation of empirical economic research in Germany) published by Prof. Richard Hauser, Prof. Gert G. Wagner and Prof. Klaus F. Zimmermann in summer 1998 initiated a very intensive discussion on the scientific community's access to official statistical microdata. In this context, the fact that the scientific community in many respects is either not at all provided with the opportunity or, as seen by the researchers, given not more than very inadequate chances to use the microdata of official statistics as

a data basis for their analyses was considered to be one of the reasons why empirical economic research in Germany is in danger to lose its importance at the international level.

The discussion was continued at a symposium promoted by the Federal Ministry of Education and Research on the "Cooperation between the academic community and official statistical bodies – Practice and prospects", which was organised by the Federal Statistical Office together with the Deutsches Institut für Wirtschaftsforschung (German Institute for Economic Research - DIW) and the Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (German Social Science Infrastructure Services - GESIS) in summer 1999. The discussion finally led to setting up a Commission to improve the informational infrastructure by cooperation of the scientific community and official statistics (KVI) chaired by Prof. Hans-Jürgen Krupp, which handed over its report on "Ways towards an improved informational infrastructure" to the Federal Ministry of Education and Research in spring 2001. That report of experts discusses in detail various ways of microdata access.

One of the results of the KVI's work is the establishment of Research Data Centres in official statistics. The Federal Statistical Office set up its Research Data Centre as of 1 October 2001, while the statistical offices of the Länder launched their common Research Data Centre on 1 April 2002. Currently, both Centres are at the developmental stage. They co-ordinate their work very closely with the aim to provide the scientific community with a joint proposal for improved access to official statistical data. During the Statistical Week which took place between 7 and 11 October 2002, the Research Data Centres introduced themselves to the public and for the first time presented the main stresses of their work in the next few years.

The common goal of the two Centres is to facilitate and further expand access to the microdata of official statistics by

- providing scientific use files and public use files whose transmission will not be subject to the science privilege,

- creating a network of workplaces for guest scientists from where they will have access to the statistical offices' microdata, and

- offering the opportunity of controlled data teleprocessing.

Based on the resources which are made available in the context of research funding, methodological research projects will be carried out, the required infrastructure be created and a system designed to provide detailed information on the data of official statistics and their use be developed to achieve this goal in the next few years.

Contact:

- *Forschungsdatenzentrum des Statistischen Bundesamtes (Research Data Centre of the Federal Statistical Office):*

  Markus Zwick, tel: (+49-611) - 75 42 20, e-mail: forschungsdatenzentrum@destatis.de

- *Forschungsdatenzentrum der Statistischen Ämter der Länder (Research Data Centres of the statistical offices of the Länder):*

  Dr. Sylvia Zühlke, tel: (+49-211) - 94 49 55 14, e-mail: sylvia.zuehlke@lds.nrw.de

De facto anonymisation of economic microdata
– anonymisation procedures: trends and tests in practical circumstances

Sophisticated techniques are required to anonymise economic microdata. Their application in practical circumstances which, even internationally, is still in its infancy, poses new challenges. Since anonymisation is laid down by law as a requirement for providing the scientific community with access to microdata of federal statistics for research purposes, making progress in this areas is a major matter of concern of the statistical offices.[1]

In this context, the Federal Statistical Office is participating in an international project on statistical confidentiality (CASC)[2], which is supported by the EU Commission. The project was touched upon earlier in Methods … Approaches … Developments (1/2001). The CASC project is aimed at further developing various methods for safeguarding statistical confidentiality and integrating them into the ARGUS program package. ARGUS consists of a program for table confidentiality (τ-ARGUS) and a program for microdata secrecy (μ-ARGUS). Currently, version 3.1 of the μ-ARGUS program is available for testing purposes. It includes data grouping, top-coding and bottom-coding procedures, and the post-randomisation procedure[3] (PRAM) which was developed by Statistics Netherlands. The PRAM procedure permits a post-randomisation of confidential variable values in a similar way as the so-called randomised response techniques in survey research.

A major project contribution by the Federal Statistical Office consists in the further development of a technique for random noise addition. Like the algorithms for microaggregation developed by another partner in the project, the technique is planned to be integrated into the μ-ARGUS program. Subsequently, both components can be tested and applied in the statistical offices together with procedures that were included at an earlier time. In a second step, a comparison is planned of all data-modifying procedures examined in the course of the project so that their characteristics will become transparent to the user. To this end, the Federal Statistical Office intends to carry out empirical checks with economic microdata.

Initially, a procedure of random noise addition developed by Sullivan/Fuller[4] was closely studied at the Federal Statistical Office. The procedure was implemented in a development environment and expanded in a way permitting the masking of individual variables while at the same time generally preserving the correlations structure between the variables subject to noise addition and the other variables.

Tests with synthetical data have proved that the procedure can be applied in a technically reliable way. Univariate distributions and correlation coefficients can be largely maintained, while the procedure shows bigger inaccuracies for very wide ranges of values. In cases where the variable distribution is extremely skew and masking considerable, however, the correlations cannot always be maintained. Apart from that procedure, another method of random noise addition developed by Kim[5] was tested as well. Unlike the procedure developed by Sullivan/Fuller, that one cannot maintain univariate distributions. It however allows an analysis of subsets.

---

1) Cf. Sturm, R.: Wirtschaftsstatistische Einzeldaten für die Wissenschaft, in: Wirtschaft und Statistik (Economy and Statistics), 2, 2002, pp. 101 – 109.
2) Computational Aspects of Statistical Confidentiality.
3) Kooiman, P.; Willenborg, L.; Gouweleeuw, J.: PRAM: A Method for Disclosure Limitation in Microdata, Research Report, Centraal Bureau voor de Statistiek, Voorburg 1998.
4) For a summarising description cf. Fuller, W. A: Masking Procedures for Microdata Disclosure Limitation, in: Journal of Official Statistics 9, 1993, pp. 383 – 406.
5) Kim, J. J.: A Method für Limiting Disclosure in Microdata Based on Random Noise and Transformation, Proceedings of the Section on Survey Research Methods 1986, American Statistical Association, Alexandria, VA, pp. 370 – 374.

Measured by a simple distance criterion, the tests have shown that the procedure developed by Sullivan/Fuller displays a higher degree of protection than other techniques of that group (simple additive random noise addition, Kim's technique).[6]

First tests based on real data have revealed that none of the procedures is fully convincing. It is in particular the requirement of analysing subsets while preserving univariate distributions that can hardly be fulfilled. For this reason, it has to be known before the actual application what kind of analyses will be carried out with the anonymised material and what estimation procedures will be applied. As regards the transmission of anonymised material to scientists, the first test results have led to the conclusion that the above procedures should be only applied to data which cannot be adequately anonymised by means of other procedures.

Roland Sturm, tel: (+49-611) 75 25 80, e-mail: roland.sturm@destatis.de

Scientists accessing microdata from their workstations
– controlled teleprocessing to use microdata from tax statistics

Microdata from the wage and income tax statistics have been accessible for scientific work for quite some time. After 1992 and 1995, the 1998 income tax statistics, including about 30 million microdata records, for the third time covers a complete year of assessment. Based on more than 500 variables, the data material sheds light on the generation of income for nearly 40 million citizens of this country. Like in the previous years, various samples will be drawn for analysis purposes because of the large size of the material.

The information potential of the above data is accessible to the scientific community in several ways. In the past, a well-established method for giving scientists access to microdata included in particular specialised analyses of the statistical material. Apart from that method, other opportunities of data use have been developed. Thus, for instance, a scientific use file based on the 1998 income tax statistics will be made available from mid-2003. And there are intentions to provide similar files for the assessment years 1995 and 1992.

Scientists have time and again expressed the wish to gain access to microdata from their own workstations. By means of controlled teleprocessing, that wish can be fulfilled to a greater extent in the future. So far, the method has been applied in a limited way only. Thanks to process standardisation measures, however, it can be used more effectively now.

A scientist who is interested in getting access to data by controlled teleprocessing transmits his questions in a program syntax form to the statistical office. Within the office, the syntax is checked for potential deanonymisation strategies and, if quite safe, applied to the original stock of microdata. Upon subjecting the data output to a confidentiality check, the results are returned to the scientist. In this way, program modifications, which will be necessary on the basis of the first results, can be implemented by the scientist relatively soon, and a final result be gradually obtained. Preconditions for that procedure are that, on the one hand, the program syntax can be run faultlessly at the statistical office and, on the other, the statistician concerned has sufficient knowledge of the program and the method applied.

However, the number of programs which can be used for controlled teleprocessing has to be limited. Currently, SPSS and SAS can be applied. Besides, the question whether and to what extent STATA (which has been additionally included by other EU statistical offices) could enlarge the program portfolio here is examined at the moment.

---

6) For an introductory survey cf. Brand, R.: Masking microdata through noise addition, in: Domingo-Ferrer, J. (Ed.): Inference Control in Statistical Data Bases – From Theory to Practice, Berlin and oth. 2002.

An elementary step in any form of electronic data processing is the preparation of faultless program syntax. Unlike logical errors or errors in data, syntax errors can be realised in a relatively simple way as they lead to abortions of program execution. Sometimes, however, the actual troubleshooting is rather time-consuming. Usually, it should be the programmer who performs troubleshooting on the basis of the data provided. For this purpose, tax statistics for the first time offers a data structure file. The structure of the respective data material fully corresponds to that of the original data material, i.e. the data record structure and the names of the variables correspond to those of the original material, though the concrete variable values do not include any content-related information, i.e. the microdata material is absolutely anonymised. Based on that material, the data user can now develop his program, check it for syntax errors and then transmit to the statistical office a program without any syntax errors.

For the time being, such data structure files are available for the 1992 and 1995 wage and income tax statistics. When the 1998 wage and income tax statistics will be published, a data structure file will be provided for that year of assessment, too.

The data structure file of the wage and income tax statistics can be obtained free of charge from the Federal Statistical Office. Analyses by means of external program syntax are carried out on a chargeable basis, that is the data user has to pay all accruing costs.

Markus Zwick, tel: (+49-6 11) 75 42 20, e-mail: markus.zwick@destatis.de

Methods of federal statistics – Further development

Estimation of inland waterways transport

Statistics on shipping and goods transport on inland waterways have been collected and analysed in Germany since the 50s. Among other things, the statistics are required to monitor and assess the structures of inland waterways transport, to calculate the proportions and determine the development of the different modes of transport (road goods vehicles, railways, ships), and also as a supporting instrument for decision-making with regard to potential infrastructural measures. Usually, annual data are sufficient for the aforementioned purposes. In addition, however, monitoring the short-term economic development of inland waterways transport is very important for businesses and public authorities who are active in the transport sector. For this purpose, monthly data are required.

The main pillar of the statistics are the reports on the goods transhipments of all German ports with waterway connection, which are submitted to the statistical offices of the Länder. The data are largely derived in an electronic manner from the internal accounting systems of the ports. Usually, however, the detailed federal results can be issued not earlier than three months after the expiry of a reporting period. And, due to data errors or capacity constraints in the agencies involved, additional delays have not rarely to be accepted. In particular the Federal Ministry of Transport has called upon the Federal Statistical Office to sooner provide its reports on the development of inland waterways transport in Germany. The procedure for estimating inland waterways transport has been designed to fulfil that requirement.

The estimation relies on a regression model. The basic assumption is that the official trend of goods transport on inland waterways can be estimated based on the early business data which are issued, on the one hand, by big ports on their goods transhipments and, on the other, by the Federal authority in charge of waterways on transport through major lockages and on canals. The presently used method is based on preparatory work which originally led to the introduction of a first estimation procedure for

inland waterways transport in 1995.[1] The new procedure has however not changed the basic concept of that time. Since the focus is on monthly data and inland waterways transport is definitely subject to strong seasonal variations, an original aim of the new development was to estimate the results on the basis of a time series model (ARIMA[2]). Test results showed, however, that the ARIMA model produced hardly other, and partly even worse results than simple regression. The reason is probably the largely irregular seasonal component of inland waterways transport which depends to a considerable extent on weather effects such as high water and low water or ice-caused blockages. For the purposes of the new estimation, the data basis has been substantially enlarged, and new methodological details have been developed. Now data are included from six ports and six lockage points and from the statistics on goods transport on northern German canals, which is compiled within the sphere of competence of the Federal Ministry of Transport. Every month, the bodies concerned each submit about 15 values on a voluntary basis. Unlike the data of the official statistics on inland waterways transport, the data transmitted by the reporting units are aggregated to a great extent and are not harmonised in methodological terms. However, this is not important as regards the application of a regression model. The data supplied are used to estimate the results of the regular statistics on inland waterways transport as shown in the table below. The new procedure has been developed on the basis of PC applications, i.e. statistical software and a spreadsheet program.

The estimation is based on the assumption of a linear multiple regression model and the application of the least squares method[3]. It was not possible to identify a non-linear relation, and the usually applied least squares optimisation method fulfils a number of major quality criteria, particularly with regard to estimate variance.[4] To exclude to the extent possible illusory correlations, the independent variables were determined on the basis of expert evaluations and, where necessary, upon additional empirical studies. Those 'potential' explanatory variables are gradually included in the regression equation in accordance with certain criteria and, when necessary, are again excluded.[5] To check whether the model selected was suited for regression, a number of graphical and mathematical analyses were carried out (i.a. Durbin-Watson test for autocorrelation of residuals). The results were highly satisfactory. There was no indication of any violation of the conditions desirable for regression.

The underlying time series starts in 1994. Since part of the relations between the variables have presumably changed to a considerable extent in the course of time, the weights applied to recent years are by far larger than those considered for previous years. In this respect, detailed experiments were performed with changing weights. In estimating the results for months affected by high water, for instance, higher weights were applied to earlier periods characterised by similar weather conditions, and the effects of that approach on the quality of the estimates were studied. As those changes in weights did not yield adequate results, they have not been integrated into the estimation procedure.

The regression equation is redetermined every month. An estimation procedure is "able to learn" in this sense. Therefore, availability of the data of all reporting units is not an absolute necessity for estimation though, of course, estimations based on incomplete data material, as they were performed in the past to keep the specified deadlines, bear the risk of quality losses.

---

1) Lindenstruth, T.: DV-gestützte, monatliche Schnellmeldung Binnenschifffahrtsstatistik (unpublished manuscript), Statistisches Bundesamt, Wiesbaden 1994.
2) See Hartung, J.: Statistik, 10th edition, München, Wien 1995, p. 678 ff.
3) See Hartung, J.: Statistik, 10th edition, München, Wien 1995, p. 569 ff., particularly p. 595 ff.
4) See Gauss Markoff theorem, e.g., in: Fox, J.: Applied regression analysis, linear models and related methods, Thousand Oaks, London, New Delhi 1997, p. 217 f.
5) In each case the variable is included in the regression equation which displays the strongest relation to the dispersion of the dependent variables which has not yet been explained by the existing regression equation. The criterion for selection is the probability of a linear relationship between the explanatory variable and the residual which has not yet been explained. That probability can change, in particular rise, with the inclusion of further variables in the regression equation. A variable is again excluded from the regression equation when the probability of no influence exceeds a certain value. In our case, the probability values were specified based on ex-post analyses as a compromise between (illusory) accuracy and robustness. Subjective expert evaluations regarding the regression equation which seemed to be most adequate were eventually decisive, too. The approach described also allowed to discover and largely exclude statistically significant dependencies between explanatory variables (to avoid multicollinearity).

On the whole, the quality of the estimated results was really satisfactory. Half of the series reached a coefficient of determination (R^2) of 0.9 or more. For none of the series R^2 was below 0.7. Due to the partly enormous range of the monthly values, right from the beginning the accuracy was not expected to be too high. The following graph shows that the trend and extent of change from month to month were estimated rather adequately. Naturally, the estimation of special or "outlier" months remains a problem.

The below table with the results obtained is made available free of charge to all interested parties on a monthly basis. Most recent estimates are successively added in that table, while estimates for earlier months are replaced by the official results. The table is usually dispatched two months after the end of a given reporting period at the latest, i.e. at least one month before the official results are released. Further acceleration of the whole process would be desirable. Due to voluntariness of the provision of information and to limited capacities in the ports, however, this goal has not yet been achieved.

Further improvements to the procedure have been planned for the future. On the one hand, there is the intention to cover more ports so that the data basis will be enlarged. On the other, the methodology applied will be continuously examined and enhanced.
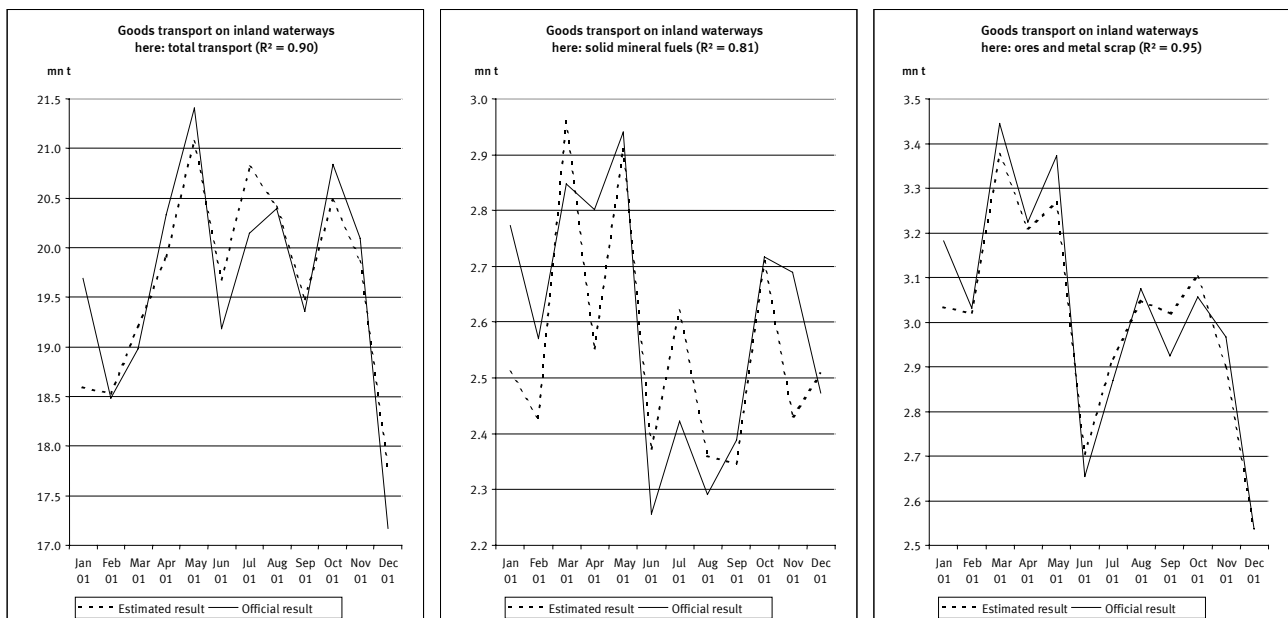
Inland waterways transport of goods
Preliminary results (estimates)

Main transport links/ Types of goods	February			January to February		
	2001	2002	Change in %	2001	2002	Change in %
Transport	mn tonnes					
Total	**18.5**	**18.1**	**-2**	**38.2**	**34.5**	**-10**
By main transport links						
National transport	4.3	4.0	-6	8.7	8.1	-7
Dispatches to other countries	3.8	3.6	-5	7.7	6.7	-12
Arrivals from other countries	8.4	8.4	1	17.7	16.1	-9
Transit traffic	2.0	2.0	-1	4.1	3.5	-13
By types of goods						
Agricultural products and the like	0.8	0.9	21	1.6	1.6	-4
Other food and animal feed	1.2	1.2	-1	2.5	2.3	-7
Solid mineral fuels	2.6	2.9	13	5.3	5.2	-3
Mineral oil products and the like	3.0	2.9	-6	6.7	6.1	-10
Ores, metal scrap	3.0	2.5	-18	6.2	4.9	-21
Iron, non-ferrous metals	1.1	1.0	-6	2.4	2.0	-15
Stones and earths	3.7	3.6	-3	6.9	6.1	-11
Fertilizers	0.7	0.6	-9	1.5	1.2	-21
Chemical products	1.4	1.4	-1	3.1	2.9	-6
Other semi-finished and finished products	1.0	1.1	12	1.9	2.2	16
German flag	**6.8**	**6.7**	**-2**	**13.7**	**12.7**	**-8**
	mn tkm					
Transport performance	**5086**	**5015**	**-1**	**10446**	**9353**	**-10**
	1 000 TEU					
Containers	**110**	**106**	**-4**	**222**	**209**	**-6**

Horst Winter, tel: (+49-611) 75 22 30, e-mail: horst.winter@destatis.de

New developments in the microcensus

The results of the microcensus in Germany, which is a representative household sample survey of the population and of working life[1], have so far been based on an annual reference week in spring. Against the background of social change as a result of which some of the previously stable social structures have acquired some dynamism, experts started to discuss, among other things, the adequacy of the reference week concept. The discussion focused on the question whether a firmly specified reference week in spring would still be adequate to produce relevant data for continuously monitoring the labour market and socio-economic changes at the level of individuals, families and households.

---

1) Literature recommended as an introduction regarding the microcensus: Emmerling, D.; Riede, Th.: 40 Jahre Mikrozensus, in: Wirtschaft und Statistik (Economy and Statistics), 3, 1997, p. 160 – 174.

The Scientific Advisory Council concerned with microcensus issues recommended in its final report on *Mikrozensus im Wandel – Untersuchungen und Empfehlungen zur inhaltlichen und methodischen Gestaltung* (Microcensus in change – studies and recommendations for its organisation in terms of content and methods), which was presented in June 1989, to change over to an at least quarterly survey frequency. The Council based its recommendation on the thesis of individualisation[2] and the related consequences. It stated that in particular the individualisation and mobilisation of social conditions required a more substantial consideration of subjective feelings which increasingly tended to deviate from the objective social patterns. And, due to the increasing pace of the processes under change, data were required which would be collected and analysed at shorter intervals and suited to explain the processes of social change also at the microlevel.[3] Apart from those socio-structural reasons for an at least quarterly frequency of the survey, Heinz Grohmann in his 1991 paper entitled *Zum Übergang vom Berichtswochenkonzept zum Konzept der unterjährigen Erhebung beim Mikrozensus* (Change-over from the microcensus reference-week concept to a concept of continuous surveys) pointed out the statistical and mathematical limits of the results obtained. He showed that the concept based on a firmly specified reference week corresponds to a reference-date survey as is typically conducted in population statistics. That type of survey provides structural information but can hardly be used to obtain average annual stock figures or to produce volume estimates, e.g., of the hours worked.

Apart from national requirements, the current microcensus also fulfils EU requirements. In 1998, however, the EU expanded its requirements based on Council Regulation (EC) No. 577/98 (and others) by the following issues:

1. provision of average quarterly and annual results,

2. smooth distribution of interviews throughout all calendar weeks of a year.

In the context of preparing an new continuous microcensus, the Conference of Heads of the Statistical Offices in spring 1999 commissioned the working group on the *Organisational and technical advancement of the microcensus*, which includes microcensus experts of the Länder and the Federation, to conduct pretests for changing over to a continuous microcensus survey.

As the change-over to a continuously moving reference week requires substantial changes in the survey design, five statistical offices of the Länder and the Federal Statistical Office carried out a three-month organisational study on the basis of the current 2000 microcensus, which was promoted by EU funds. The study was designed to achieve the following goals in view of a continuous survey:

- improved timeliness by:

  – a shorter time-lag between interviews and the publication of results

  – quarterly results

- improved quality to the extent possible by:

  – applying modern data collection tools

  – reducing the staff and increasing the level of professionalism of the group of interviewers

- compliance with the requirements of the EU Labour Force Survey (Council Regulation (EC) No. 577/98):

---

2) The thesis of individualisation was originally developed by Georg Simmel. In 1986, Ulrich Beck succeeded in his book entitled "Die Risikogesellschaft" in establishing that thesis both as a theoretical paradigm and an interpretation of everyday life.

3) H. Esser u.a.: Mikrozensus im Wandel – Untersuchungen und Empfehlungen zur inhaltlichen und methodischen Gestaltung, Stuttgart 1989, p. 29.

- upper limit of the relative standard error regarding annual averages at the level of administrative regions

- upper limit of the relative standard error regarding changes between two successive quarters at the national level

- maintenance of the level of regional detail

- organisational and content-related co-ordination with other official population sample surveys

The results of the organisational study based on a subsample of 11,000 households showed that, basically, a continuous survey is feasible in the tested organisational circumstances. However, the results also pointed to the fact that, under the conditions of an infra-annual continuous survey, such areas as work/training of interviewers, logistics and process organisation and also the survey methods as such would require a considerable advancement.

The following can be stated with regard to data quality and regionalisation:

- The average annual results and the results of the current reference date survey display similar standard errors. Consequently, the detail of regionalisation will be maintained for the annual result when changing over to a continuous survey. The different character of average results in terms of their content should however be taken into consideration.

- The standard errors of quarterly results are about twice as big as those of annual results. At the level of adjustment strata (male/female; German/non-German), sections of the population whose proportion is about 5% can still be covered if the relative standard error is 15%.

- Due to missing overlaps, the sampling errors are relatively big for changes between two successive quarters. The error variance is about twice as large as that of a quarter. To comply with the accuracy requirements of Council Regulation (EC) No 577/98 of 9 March 1998 regarding the change in a given quarter compared to the previous quarter, the EU Labour Force Survey has to be based on the overall sampling rate of 1%. At the level of the regional adjustment stratum, sections of the population whose proportion accounts for a minimum of 8% can be covered – provided the relative standard error related to the total value of the subgroup is not bigger than 15%. This, however, is only possible if a significant change has occurred which by far exceeds the standard error of 15%. Hence the feasibility to include a subgroup does not only depend on the size of that subgroup, but also on the extent of change.

- The degree of precision achieved in presenting changes improves with an increasing extent of overlap and variable correlation. The degree of correlation differs from variable to variable and depends additionally on changes in relation to the households and members of the households covered in the sample. With a relative standard error of 15% related to the total value of the subgroup and a correlation coefficient of the variable of 0.85, sections of the population whose proportion amounts to a minimum of 3% can be covered at the level of the regional adjustment stratum, provided the change on the same quarter of the previous year is significantly bigger than the standard error.

With the aim to enhance the currentness and quality of the data and to maintain the level of regional detail, the Conference of Heads of the Statistical Offices which took place on 13 and 14 November 2001 decided to introduce a new microcensus survey design. From 1 January 2005, the data collection so far carried out in spring of each year will be replaced by a continuous survey based on a moving reference week. The legal basis required for this purpose at the national level is currently being prepared.

Sabine Lotze, tel: (+49-18 88 / 644 87 21, e-mail: sabine.lotze@destatis.de

The methodological treatment of e-commerce and software in foreign trade statistics

Both the new type of distribution "e-commerce" and the product "software" are characteristics of our modern information society. However, the importance of e-commerce and software has so far hardly been reflected by official statistics in Germany, exceptions being first pilot surveys on e-commerce as part of wholesale and retail trade statistics. Data on *cross-border* trade with (standard) software are included in the results of German foreign trade statistics. Where e-commerce is performed across national borders or software is sold to clients abroad, it is an obvious choice to show such transactions separately in foreign trade statistics or possibly to include new values of characteristics in that area of data collection because of the general interest shown in such information.

The following common definitions of e-commerce and software are applied in this article.

E-commerce: Transactions performed via Internet or other electronic networks. Goods or services are ordered through such networks, while the goods are delivered, or the services rendered, online or offline.

Software: Machine-readable information stored on data media (e.g. diskettes, CD-ROMs, magnetic tapes) for the reproduction of programmes, sound or images.

The coverage of e-commerce transactions and software in foreign trade statistics is based on a fundamental principle of such data collection, that is, only the *physical trade in goods* is represented. This means in particular:

1. Generally, information transmitted online is not covered by foreign trade statistics. This is true of both types of data collection in German foreign trade statistics (direct declaration of intra-Community trade to the Federal Statistical Office and customs-based data collection for third-country trade).

2. Goods ordered online which then physically cross the border are covered as "normal" foreign trade but are not presented separately. Introducing an additional statistical characteristic such as the "way of initiating the transaction" is currently not a subject of discussion because there is a general trend towards reducing and simplifying statistical surveys in order to disburden respondents.

3. Parallel to customs regulations, the legal provisions of foreign trade statistics applicable until the end of 2000 required that only the value of the data medium be declared, regarding *any* software deliveries. In practice, even those data were presumably incomplete because of exemption thresholds in terms of value. This means that until the relevant statistical EU Regulations were modified at the end of 2000, the value of the software itself – which in many cases is considerable – was not included in the results of foreign trade statistics.

As explained unter item 3 above, the fact that in the past only the hardware value was covered by foreign trade statistics was based on the view that software is really a service rather than a commodity. However, "standard software" has the disctinct character of a commodity; just think of common word processing programmes that are available as final products, similar to books or music CDs, which have to be declared for foreign trade statistics when crossing the border. This is why the two statistical EU Regulations No 1901/2000 and 1917/2000 ("Intrastat and Extrastat Implementing Regulations") were modified in September 2000, rendering compulsory the declaration of standard software (on physical data media) for foreign trade statistics. Since that time, a diskette containing standard software has to be declared with the value of the diskette itself plus the market value of the software, using the relevant eight-digit commodity code for "recorded diskettes" of the Commodity Classification for Foreign Trade Statistics. What is exempted from declaration for foreign trade statistics still today, however, is software produced for an individual customer ("individual software") because it is a generally accepted view that in those cases the service character is prevailing.

From methodological aspects, the modifications to the above-mentioned statistical Regulations are reasonable, and this shows that even surveys that in part are characterised by tradition, such as foreign trade statistics, can be adjusted to new information requirements. It should not be disregarded, however, that in declaration routine, difficult questions do arise. For example, the borderline between individual and standard software is fluid. Also, software may be linked to hardware and in this combination be moved abroad, which means that different provisions for declaration would become applicable.

Andreas Kuhn, tel: (+49-611) 75 29 93, e-mail: andreas.kuhn@destatis.de

First survey in the services sector

As in other leading industrial nations, the services sector in Germany has experienced a dynamic development over the past years and has evolved into one of the factors decisive for economic growth.

This trend, however, was not sufficiently reflected by official statistics until the late 90s. Following years of preparatory work, the Law introducing Services Statistics was finally adopted in December 2000. That law provides for yearly enterprise surveys, which are to be conducted starting with reference year 2000 and are to describe economic trends in the services sector. Those surveys at the same time permit to fulfill the requirements laid down in EU Council regulation No 58/97 concerning structural business statistics, which prescribes the annual delivery of comprehensive data on the services sector to Eurostat. On top of that, the results of the services statistics will contribute to further enhancing national accounts, since with their provision, data on the economic activity of entire economic sectors will become available for the first time.

The services statistics cover the following headings of the Statistical Classification of Economic Activities in the European Community (NACE Rev. 1):

- Section I   Transport, storage and communication

    Division 60   Land transport, transport via pipelines

    Division 61   Water transport

    Division 62   Air transport

    Division 63   Supporting and auxiliary transport activities; activities of travel agencies

    Division 64   Post and telecommunications

- Section K   Real estate, renting and business activities

    Division 70   Real estate activities

    Division 71   Renting of machinery and equipment without operator and of personal and household goods

    Division 72   Computer and related activities

    Division 73   Research and development

    Division 74   Other business activities.

The statistical units are enterprises and units established for the purpose of carrying out freelance activities, together with their local units in Germany. The survey is conducted in a decentralised manner using a sample including not more than 15 percent of all statistical units. The sample is drawn on the basis of the business register, which also serves as a frame for expanding the results.

Variables collected for the statistics are:

- enterprise identification data,

- number of persons employed, wages and salaries, social security expenses,

- turnover or income, intermediate consumption, and taxes and subsidies,

- investment.

To avoid an excessive burden on the units obliged to respond - most of which are subject to an official statistical survey for the first time - the programme of the services statistics includes 31 different survey characteristics. Small enterprises, i.e. enterprises with an annual turnover of less than 250 000 euros, have to report not more than twelve variables. That approach will greatly contribute to the public acceptance of the new statistical reporting system. For the first time, official statisticians are discriminating between enterprises in such a way. Large enterprises with local units in several German federal Länder report their data concerning turnover, persons employed, wages and salaries and investment in a breakdown by federal Länder. This way, results may be calculated for those enterprises also according to the regions in which their local units are seated.

Since the enterprises or institutions obliged to provide information are a rather heterogeneous group, three different questionnaires are used to implement the services statistics. The variable definitions and explanations used in those questionnaires are customized to the respective group of respondents. In fact, a so-named basic questionnaire is used for roughly 80 % of the respondents, while freelancers and holding companies are surveyed using specific questionnaires. In the processing stage, we ensure that the results originating from those three different questionnaires are properly linked.

The results at federal level for reference year 2000 are available since September 2002.

Dr. Bernd Petrauschke, tel: (+49-18 88) 644 85 62, e-mail: bernd.petrauschke@destatis.de

New law passed on trade statistics

A balanced and harmonised system of statistical surveys in trade and the hotel and restaurant industry, which constitute an important sector of the national economy, is indispensable for monitoring short-term economic trends and structural changes. In the Federal Republic of Germany, the following legal provisions have been applicable to that sector:

- Council Regulation (EC, EURATOM) No 58/97 of 20 December 1996 concerning structural business statistics,

- Council Regulation (EC) No 1165/98 of 19 May 1998 concerning short-term statistics,

- the Law on Statistics in Trade and in the Hotel and Restaurant Industry of 10 November 1978 and

- the Law on Cost Structure Statistics of 12 May 1959.

As far as their regulatory content was concerned, the surveys laid down by the above legal provisions were not coordinated with each other.

The main objective of the new Trade Statistics Law is to harmonise national and EU requirements. To this end, the former national working programme was examined in cooperation with all institutions represented in the specialist committee on trade statistics, and adjusted accordingly. This did not only result in suggestions to reduce the programme, but also brought about a number of proposals for necessary extensions. As a result, it was possible to take into account the information requirements of the EU as well as of the Federation, the Länder and other major national users. Furthermore, it was

possible to avoid conducting multiple surveys at the enterprises in trade and in the hotel and restaurant industry.

The "Law on the Reorganisation of Statistics in Trade and in the Hotel and Restaurant Industry - Trade Statistics Law (HdlStatG)" was passed in December 2001 and went into force on the day following its promulgation on 10 December 2001 (Federal Law Gazette, Volume 2001, Part I p. 3438).

The Law lays down that statistical results on trade (NACE Rev. 1, Section G) and on the hotel and restaurant industry (NACE Rev. 1, Section H) shall be compiled through compulsory representative sample surveys with monthly, annual and multi-annual periodicity. Hence, the system of trade statistics now also covers repair activities and activities of commission agents, which are classified under Section G, by means of continuous surveys.

As before, the monthly surveys collect data on turnover, full- and part-time employment. At the national level, monthly figures on the development of those variables are considered indispensable for economic analysis and short-term statistics. The EU Regulation concerning short-term statistics, however, provides for monthly statistics only of retail trade, while quarterly statistics are stipulated for the remaining branches "sale of motor vehicles", "wholesale trade" and "hotels and restaurants". What the monthly surveys do no longer include are repairs of personal or household goods (due to insignificance). The obligation to collect data in great detail for the Länder has, however, been retained. As of reference month January 2002, the monthly surveys have been conducted in accordance with the new legal provisions.

The annual surveys cover data on earnings and costs as well as on the structure of employment and investments. The results are meant to throw light on, among others, the enterprises' profitability and productivity. Besides, the annual surveys are used as instruments to revise the economic coding of the enterprises and to bring it up-to-date - within the context of monthly statistics as well. The annual surveys' scope of coverage also encompasses repairs of personal or household goods. As compared to the former Trade Statistics Law, the list of variables has been considerably extended in order to meet the requirements of the EU Structural Business Statistics Regulation (the annual surveys in 1999 and 2000 already covered the extended list of variables). While the former annual surveys collected information only for enterprises (this also holds for the 1999 and 2000 surveys), the subsequent annual surveys (including the 2001 survey) will cover the variables "number of persons employed", "gross wages and salaries" and "total gross investment" with particular detail for the Länder.

At several years' intervals, i.e. every five years, data are collected on the floor space of retail shops.

The aim of imposing as little burden on enterprises as possible and of using every opportunity to further reduce that burden has been incorporated into the Law by laying down branch-specific cut-off thresholds for the monthly statistics. In addition, the quarterly cost structure statistics in trade and the hotel and restaurant industry have been suspended, and the Law lays down that the census of trade and the hotel and restaurant industry, which was last held in 1993, shall no longer be taken with a regular, ten-year periodicity. That census will in future be executed only if an appropriate legal act has been passed with the consent of the Bundesrat.

Gustav Grillmaier, tel: (+49-611) 75-2135, e-mail: gustav.grillmaier@destatis.de