

METHODS – APPROACHES – DEVELOPMENTS

Information of the German Federal Statistical Office

Edition 2/2006

The catchword

Measuring the anonymity of longitudinally linked economic statistics microdata..... 3

Methods of federal statistics – Further development

Jackknife confidentiality protection versus the utilisation of scientific use files
examined on the basis of a Cost Structure Survey 6

Treatment of dwelling services in national accounts 8

Earlier publication of employment data for the economy as a whole in the
national accounts context 10

Expenditure on and investments in information and communication technologies,
ICT investments - survey pursuant to Art. 7 (2) of the Federal Statistics Law to meet new
EU requirements in the context of structural business statistics 12

Effects of the recent Collective Agreement for the Public Service (TVöD)
on the public service personnel statistics 15

Events

The 1996 - 1999 Microcensus Panel as a Scientific Use File: An Introduction
ZUMA Workshop on 14 September 2006 17

1st Users' Conference on the Sample Survey of Income and Expenditure
(*Einkommens- und Verbrauchsstichprobe - EVS*) 18

Published by: Statistisches Bundesamt (Federal Statistical Office), Wiesbaden

Subject-related information

on this publication:

Division IB,
tel: (+49-611) 75 20 77
fax: (+49-611) 75 39 50
christian.koenig@destatis.de

General information

on the data supply:

Information service,
tel: (+49-611) 75 24 05
fax: (+49-611) 75 33 30
<http://www.destatis.de/allg/e/kontakt/contact2.htm>

**Release calendar
of the Press Office:**

www.destatis.de/presse/deutsch/cal.htm

Periodicity: (generally) twice a year

Published in December 2006

© Statistisches Bundesamt, Wiesbaden 2006

Reproduction and free distribution, also of parts, for non-commercial purposes are permitted provided that the source is mentioned. Any distribution, also of parts, via electronic systems/data media requires prior permission. All other rights reserved.

The catchword

Measuring the anonymity of longitudinally linked economic statistics microdata

Scientists increasingly express the desire to use official statistics microdata for their own empirical economic and social research. The road prescribed by the legislator is that microdata should be converted into a factually anonymised form, before they are made available to scientists. Accordingly, data items are regarded as sufficiently anonymised, if the expenditure needed for a possible reallocation is “unreasonably high”.

While the work done by statistical offices in cooperation with scientists during 2002-2005, which was sponsored by the Federal Ministry of Education and Research (*BMBF*)¹, on the project “Factual Anonymisation of Economic Statistics Microdata” has, fortunately, shown that factual anonymisation of cross-sectional economic statistics microdata can, as a rule, be achieved by using special information-reducing and data-altering methods, evidence of that kind has still to be provided in respect of data items that are linked longitudinally (so-called panel data). That is why, since the beginning of the year, the Research Data Centre of the Federal Statistical Office, in cooperation with the Research Data Centre of the Statistical Offices of Germany’s Federal States, the Institute for Applied Economic Research (IAW), Tübingen, and the Institute for Labour Market and Vocational Research (IAB), Nuremberg, has been conducting a BMBF-sponsored project “Economic Statistics Panel Data and Factual Anonymisation”.

Apart from expensive work on linking the various waves of surveys examined in the context of the project and complex investigations into the potential analytical capabilities of (anonymised) panel data, one essential module of the project consists in the development of a technique for simulating data-intrusion scenarios a potentially attacking data intruder might apply with the goal to re-identify units in a stock of data that need to be protected. One important constellation is the so-called database cross match scenario. In a database cross match scenario, an attacking data intruder tries to assign as many external database units as possible (extra knowledge) unambiguously to units of an anonymised target database in order to extend the external database by target database information.

In a first phase, the database cross match scenario is mathematically modelled as a multicriterial assignment problem, which is then converted, by way of suitable parameterisation, into an assignment problem with one target function to be minimised. Then, the main concern is to choose the best-fitting coefficients of this target function. Whereas in the past a distance-measure, generated across all matching variables of the two data sources (key variables and overlaps), proved to be well suited for the examination of cross-sectional data, it now seems that the examination of panel data requires the use of additional measures. As the information on variables, which is now available to a potential data intruder, has been collected in several waves, it is obvious that this more complex structure should be reflected in the coefficients as well. With that goal in mind we have considered various approaches, of which the following three are most promising:

- **Distance-based approach:** The square deviation is calculated for every metrical overlapping variable and every pair of matching units in the two data sources. Then, these deviations are standardised to intervals of [0.1] and summed up. It may be advisable in some cases to assign additional weights to the various deviations on unit level. A weakness of that measure becomes apparent in cases where the definition of some overlapping variable slightly differs between the two data sources, for example, if a

¹ This was already described in MVE 2/2003.

variable such as “number of employees” relates to the number of all employees in absolute terms in one data set, whereas that number is converted into full-time workers in the other data set. Another example would be a case in which one of the data sets includes total turnover figures, whereas the other data set shows total turnover less a certain sub-proportion of that turnover. In these cases we are confronted with the fact that there is a bias in the overlapping variables, even if we presume that the circumstances of the surveys are comparable.

- **Distribution-based approach:** In a panel data situation we can take it for granted that an attacking data intruder will have information over several years for every overlapping variable, for example, total turnover of an enterprise from 1999 to 2002. In general, we can assume the existence of a bias between the two sources of data. Using the chi square measure and presuming independence of distributions from data collection, we can measure the similarity of total turnover distribution between two different units over four years. Apart from the presumed independence, another weakness of this approach becomes apparent, when overlapping variables have high degrees of variance. In that case, one should forbear from that approach or use a simplified measure which considers but the monotony behaviour of the variables over some years.
- **Correlation-based approach:** A data attacker may have information for two overlapping variables over several years in both sources of data, for example, “total turnover” (u_1, \dots, u_n) and “number of employees” (b_1, \dots, b_n) of an enterprise for the period of n years. If one interprets the pairs of values (u_i, b_i) as manifestations of two arbitrary variables, those units that belong together in the different data sources can be expected to reveal empirical correlation coefficients that are “similar”. It should, however, be considered that what is measured by correlation is just the linear interrelation of two variables. In special cases the two estimated correlation coefficients can diverge from each other very clearly, even if there the variables are linked by a direct functional relationship.

Once the coefficients are calculated, one can solve the linear assignment problem using classical methods such as the Kuhn-Munkres algorithm or the simplex method. For larger data blocks (typically generated when dealing with VAT statistics) it is recommendable for reasons of efficiency that approximation heuristics should be used.

Because of the mentioned weaknesses of the various measures described above it is planned to combine them in a suitable way. Here we distinguish between two types of combination:

The first type consists in combining the various measures so that an overall measure is obtained and then, by way of solving the assignment problem, in forming pairs of assignments (*hybrid assignment*), where, for example, $w_dist(a,b)$, $w_chi(a,b)$ and $w_corr(a,b)$ are those weights for pair (a,b) that pertain to the three measures. The easiest way to combine these weights is convex combination (i.e. a kind of weighting such as, e.g., that used for arithmetic means). This may also include special cases such as

$$w_1(a,b) := \min(w_dist(a,b), w_chi(a,b), w_corr(a,b)) \quad \text{or}$$

$$w_2(a,b) := \min(1_{[c,\infty)}(w_dist(a,b)), w_chi(a,b), w_corr(a,b))$$

In respect of measure w_2 there will be one pair (a,b) containing the biggest possible value 1, which will accordingly be rejected as a possible assignment, if it exceeds a preset threshold c with regard to w_dist .

The second type consists in a special function to be applied to the suggested pairs of units, which the various measures have identified with this method and which already represent a preselection of possible assignments (*compound assignment*). Z_dist , Z_chi and Z_corr be sets

of assignments obtained by the various measures. First examples of the function mentioned above are generated by a composition of operations applied to the sets of assignments:

$$f_1(Z_dist, Z_chi, Z_corr) := Z_dist \cap Z_chi \cap Z_corr,$$

$$f_2(Z_dist, Z_chi, Z_corr) := (Z_dist \cap Z_chi) \cup Z_corr \text{ or}$$

$$f_3(Z_dist, Z_chi, Z_corr) := Z_dist \cap (Z_chi)'$$

For example, the only pairs that function f_1 selects are those suggested by all three measures together. That would mean that a data intruder has made just a small set of assignments, which, however, are very probably correct. Function f_2 allows multiple assignments as well and in these cases the data intruder has to take a final individual decision in favour of one assignment, which is unambiguous.

After implementation, the approaches described above will experimentally be applied to different trial anonymisation cases of the Cost Structure Survey in Manufacturing (1995-2003), the Monthly Reports in Manufacturing (1995-2004), VAT statistics (2000-2004) and the IAB Panel of Local Units. Risk tables discussed in advance can serve as a basis for the decision whether an experimentally anonymised file can be rated as factually anonymous. For example, table 1 below can be used to enter risks of reidentification by size class of employees of enterprises to be identified and approach to the coefficients of the target function. A_1, \dots, A_k mean special weights of the three measures and Z_{A_1}, \dots, Z_{A_k} denominate the resulting assignments (hybrid assignment).

Table 1: Risks of reidentification by size class of employees and assignment approach.

	Z_dist	Z_chi	Z_corr	Z _{A₁}	...	Z _{A_k}
Class 1	0.42	0.33	0.12	0.23	...	0.06
Class 2	0.03	0.31	0.62	0.50	...	0.24
...
Class n	0.88	0.37	1.00	0.02	...	0.38

Table 2 below can serve to enter global risks of reidentification when using the symmetric binary operation on sets $f(A,B)=A \cap B$ (compound assignment).

Table 2: Global risks of reidentification

\cap	Z_dist	Z_chi	Z_corr
Z_dist	0.40	0.18	0.25
Z_chi	0.18	0.33	0.00
Z_corr	0.25	0.00	0.66

In conclusion, it has to be mentioned that one should deal very sensitively with the risks estimated in the above manner. It is obvious that a growing number of combinations will also lead to an increase in the maximum values of risks emerging in the tables. That is why, in order to remain realistic, one should assess a file's factual anonymity on the basis of just some selected

combinations which appear to correspond as closely as possible to the approaches a data intruder is likely to use.

Dr. Rainer Lenz, tel: (+49-611) 75-2636, e-mail: rainer.lenz@destatis.de

Methods of federal statistics – Further development

Jackknife confidentiality protection versus the utilisation of scientific use files examined on the basis of a Cost Structure Survey

Introduction

The 1/2005 edition of *Methods – Approaches - Developments* presented a new confidentiality protection method, the so-called Jackknife Technique (JT), which was introduced by Jobst Heitzig.² This technique is a compromise of output- and data-related anonymisation, which is supposed to ensure absolute confidentiality, rather than factual anonymity. The underlying idea is that single microdata items in a cell of the original record one at a time are systematically replaced by values from an appropriate substitute distribution and that the implications that procedure has for the results of the analysis are calculated for each individual operation, so that in the final analysis one can be sure that the published result actually includes the true value, but does not in any case allow individual values to be deduced. The article mentioned above describes first positive evaluations concerning the accuracy of the results achieved.

Below we will show how the jackknife confidentiality method performs in comparison with the practiced scientific use files regarding the conservation of analytical potentials, an examination based on the 1999 Cost Structure Survey (CSS), which is used as an example. Scientific Use Files (SUF) relate to more than 13,000 minor and medium-sized enterprises covered by the CSS. Although large enterprises are not considered, what we have is, in fact, a skew distribution of the various variables.³

Analyses

Apart from the consideration of distribution measures, the estimations we made by using regression analysis were the same as those in the analyses of the SUF data mentioned above, which were used in the project “Factual anonymisation of economic statistics microdata” (FAESM), and logistic regression was evaluated additionally. In terms of substitute distributions, we made use of the normal distribution (with a variation of standard deviations), apart from the rectangular distribution of categorical variables, and of a specifically developed skew distribution for the metric variables. What we did was to compare the error of the result obtained by using the SUF method with the deviation between the middle of the interval of the published JT result and the true result. The criteria for the assessment of analytical potentials for the various

² See “*Methods-Approaches-Developments*”, 1/2005 edition, pp. 7-8. For more details, please, refer to: The “Jackknife” method: confidentiality protection for complex statistical analyses. Monographs of official statistics (2006 edition) - Work session on statistical data confidentiality, Geneva, 9-11 November 2005, pp.325-332.

³ For more information on the CSS and SUF, please, refer to the manual mentioned later (see footnote 3).

analyses were laid down in conformity with the Manual for the Anonymisation of Economic Statistics Microdata developed under the FAESM project.⁴

Results

Descriptive evaluations

The JT was applied repeatedly in respect of distribution measures and correlations and the yardstick used for all cycles was the mean value. As to the distribution measures we found out that the results obtained were somewhat better when using a normal distribution rather than the specific skew distribution. The JT performs more poorly than the SUF as far as the definition of mean values and standard deviations is concerned, however, it should be considered that the share of the mean value's deviations is in line with the requirements of the critical threshold as defined by the FAESM project. Concerning deviations in respect of the median, the SUF is outperformed by the JT.

First evaluations of distribution measures by sub-population (economic sector, size class of employees) show deviations which are clearly higher: In all statistics considered we fail to meet the requirements of FAESM criteria, although we almost succeed in meeting them in the case of the median.

As to correlations, the use of the various dispersion plots between two variables has confirmed that skew distributions make it problematic to measure a linear correlation with a Pearson correlation coefficient. That is the reason why logarithmic variables are used in such case. The JT and the SUF methods yield similarly good results concerning the Spearman rank correlation coefficient: absolute deviations of more than 0.1 do not occur in the correlations. The same is also true of the Pearson correlation with logarithmic variables. Here, however, the average deviation of the JT is at least three times as high as that of the SUF.

Econometric analyses

Regression analyses with a Cobb-Douglas production function (the first model excluding and the second model including economic sectors as dummy variables).⁵ Similarly to the evaluation of descriptive statistics, the JT was applied repeatedly to achieve robustness regarding the results. The evaluations show that the first model ensures a very good analytical potential both with the SUF and the JT: it meets all the criteria laid down in the FAESM project. Similar results can be observed regarding the average deviation in respect of coefficient values and robust values, with the SUF performing slightly better. The second model shows some clear deviations between the various cycles regarding the JT with dummy variables. Furthermore it occasionally happens that the interval, which the JT has defined for some economic sectors, is too broad to allow the identification of the level of significance. Here, the SUF should be preferred to the JT.

Logistic Regression

As an extension to the analyses made in the FAESM project, we constructed a model, which apart from the input factors of the Cobb-Douglas production function used the East-West categorisation and the changes in stocks of self-produced products, raw materials and supplies as well as goods for resale as explanatory variables and *Research and Development* (R&D) as a binary dependent variable (yes or no).

⁴ See "Manual for the Anonymisation of Economic Statistics Microdata" (*Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten*) in: Statistics and Science, volume 4, Federal Statistical Office, ed. (2005).

⁵ For more details about the model, please, read chapter 21.1 of the Manual mentioned earlier.

The explanatory variables were divided into two or three size classes each. Then, in contrast to the linear regression model we estimated the extent to which the probability that an enterprise is doing R&D work is influenced by the given size class in relation to the corresponding reference class. For example, it has turned out that East-German enterprises are nearly twice as likely as West-German enterprises to have R&D. Considering the discrepancies between these results and the original data we can state that the analytical potentials of both the SUF and the JT are very good. In the final analysis, if we include all estimates and quality criteria in the consideration, the JT performs a little better. For example, the average deviation of the regression coefficients amounts to as little as 0.38% in the case of the JT, whereas it is 1.1% in the case of the SUF, which is nearly three times as high.⁶

Current state and outlook

The present study shows that the jackknife method of confidentiality protection of CSS data is particularly suitable for econometric estimations and as good as the SUF.⁷ As in the latter case the change in the data was relatively weak⁸, it may be assumed that the JT possibly provides not only a higher degree of confidentiality (in absolute rather than factual terms), but also a better analytical potential in such analyses. This is to be more closely examined in the near future.⁹

As far as distribution measures are concerned, the JT is not at all inferior to the SUF, if robust statistics (median, Spearman's rank correlation) are available. However, higher deviations were found with regard to estimates of standard deviation and evaluations by subpopulation.

It can also be stated that the two substitute distributions considered for the metric variables (normal distribution and specific skew distribution) lead to similar results.

On the whole, the evaluations made confirm that the jackknife method of confidentiality protection is an extremely promising technique that may be used to provide scientists with high-quality results of complex analyses without extra expenditure in terms of manual checks.

Michael Scheffler, tel: (49-611) 75-2574, e-mail: michael.scheffler@destatis.de

Treatment of dwelling services in national accounts

Calculating dwelling services in national accounts is based on a complex calculation model briefly presented below.

In the practice of German national accounting, the industry of dwelling services comprises the entire letting of housing space and the services from owner-occupied dwellings as part of division 70 of the WZ (national classification of economic activities). Unlike the allocation of enterprises to industries according to their main economic activity, which is usually applied in the production approach, the area of dwelling services is defined in functional terms. Any

⁶ Similar to the estimations of regression analysis, several cycles were run to confirm the robustness of the results.

⁷ An exception is estimation in terms of regression analysis which uses economic sectors as dummies and, like in the descriptive case, indicates problems of evaluation by subpopulation.

⁸ The approach to anonymisation, which was used to change the data, was separate micro aggregation. For more details, please, refer to chapter 6.2.4 of the manual mentioned above.

⁹ For example, an option is to consider the SUF for all enterprises covered by retail trade statistics. The data-changing approach used here is one of random overlay.

transactions connected with dwelling services are shown in that industry, irrespective of the industry to which the owner's main economic activity is allocated.

The output of dwelling services comprises the rents actually paid and the imputed (fictitious) rents for owner-occupied dwellings. For owner-occupied dwellings, the output is calculated by means of a model analysis using the stratification method based on comparable actual rents, with the strata being formed according to the criteria of rents (size and equipment of the dwelling, age of the building, type of financing, federal Land).

The German calculation thus is fully in line with the computation method required by the EU. What must be complied with, apart from the European System of Accounts 1995 (ESA 95), is especially the Commission Decision of 18 July 1995 (95/309/EC, Euratom) specifying the principles for estimating dwelling services and EU Regulation No 1722/2005 of 20 October 2005 on the principles for estimating dwelling services for the purpose of Council Regulation (EC, Euratom) No 1287/2003 on the harmonisation of gross national income at market prices.

In the old Länder, the calculation of the output of dwelling services is based on results of the census of buildings and housing conducted in 1968 and 1987; more recent data are provided by the sample survey of buildings and housing of 1993 and the supplementary microcensus components of 1998 and 2002. The calculations for the new Länder – as far as the stock of dwellings is concerned – are based on the results of the census of buildings and housing of 1995, which was conducted as a complete enumeration in the new Länder, while information on the level of rents is obtained from the sample survey of buildings and housing of 1993 and the supplementary microcensus components of 1998 and 2002. For years where quantity and/or price data are not available in a detailed breakdown from dwelling surveys, the output of dwelling services is updated through combined quantity and price updating.

Updating of the overall stock (number and floor space of dwellings) is mainly based on the completions and on the units leaving the statistics on building activity. Other units added or leaving that are not covered by the statistics on building activity are also included. This refers, for example, to the transformation of a dwelling used for residential purposes into a dwelling used for commercial purposes (e.g. medical practice, lawyer's office, etc.) or the conversion of two or more dwellings into one. In the former territory of the Federal Republic, the units added or leaving that are not covered by the statistics on building activity are determined by comparing the results of the census of buildings and housing (GWZ) 1987 with the results of the GWZ 1968 that were updated up to 1987, and by making plausible assumptions. As a result of those corrections, the total number of dwellings is slightly below, and the total living floor space above the figures published otherwise. As a final step, the dwellings and living floor spaces of unoccupied dwellings as well as the living floor spaces of commercially used rooms are calculated and excluded from that overall stock.

The average rents (gross rents exclusive of heating expenses) are updated by means of rent indices from the statistics of consumer prices. Imputation factors take account of the quality changes not included in the price index (e. g. quality improvements through new construction or renovation). The average rents in the new Länder are based on the figures of the sample survey of buildings and housing 1993 and are updated using the data of the Federation of German Housing Companies (GdW). In a further step, the updated average rents per m² are then adjusted to the results of the supplementary microcensus components of 1998 and 2002. As regards the charges for heating and warm water (also referred to as incidental rental expenses), the calculation performed for the old and new Länder is the same as for gross rents. Net rents are obtained by deducting the charges for heating and warm water from gross rents.

The rents for garages and parking spaces, which belong to the area of dwelling services are calculated on the basis of the average rents and the number of garages/parking spaces from results of the sample survey of buildings and housing 1993. The number of garages is updated

on the basis of the trend for the number of dwellings, while the average rents are updated using the consumer price index for privately let garages, the level of detail being rented dwellings and owner-occupied dwellings, in a breakdown by Länder. The output of dwelling services is obtained by adding up net rents and rents of garages.

The intermediate consumption in the area of dwelling services is the consumption of goods and services occurring as costs of maintenance and repair in the context of dwelling services (letting or owner occupancy of dwellings). Calculating the intermediate consumption or the ratio of intermediate consumption to output of dwelling services is mainly based on the continuous household budget surveys, the sample surveys of income and expenditure, and data from the Federation of German Housing Companies (GdW). Intermediate consumption is calculated on the basis of average ratios of intermediate consumption to output, broken down by rented dwellings and owner-occupied dwellings and available separately for the former territory of the Federal Republic and the new Länder incl. Berlin-East, and on the basis of the relevant output per Land. The FISIM (financial intermediation services indirectly measured), which since the 2005 revision of national accounts have been part of intermediate consumption of any industry, are allocated to the industries by means of indicators – as no other information is available – and are added to other intermediate consumption. Gross value added for dwelling services is then obtained by subtracting intermediate consumption from output.

Those results, which are also available at Land level, are used by the working group on national accounts of the Länder to calculate regionalised results.

The results are, first, integrated into the gross domestic product (GDP) through the generation of income and, second, they are shown as final consumption expenditure of households in the use of income. What is shown as a subgroup of the item “Housing, water, electricity, gas and other fuels” is, among other things, the “rents actually paid”, the “imputed rents” and the charges for heating and warm water, here referred to as “water supply and other services in connection with the dwelling”.

Also, results for the dwelling services are included in the distribution of income approach. As regards the income of households, the primary income from dwelling services is part of the entrepreneurial income.

Jürgen Bopp, tel: (+49-611) 75-2774, e-mail: juergen.bopp@destatis.de

Earlier publication of employment data for the economy as a whole in the national accounts context

Against the background of the contemporary process of globalising economic relations, the perspectives in analysing labour supply and demand in the framework of short-term economic and labour market monitoring have become increasingly comparable at the international level. Hence, the demand for timely short-term and internationally comparable information on total employment and the related changes over time has increased, too.

In this context, an EU-US benchmarking project was launched at the European level, whose current aim is to publish quarterly results in the form of a first release for 19 Principal European Economic Indicators such as gross domestic product (GDP) and employment data as early as 45 days after the end of a reference quarter. Also, for the purposes of labour market monitoring, monthly unemployment rates are to be released in a breakdown by demographic variables such as sex and age and the level of qualification by all Member States of the European Union (EU) 30 days after the end of a reference month.

As for Germany, the EU requirements regarding quarterly GDP results and employment data for the economy as a whole have already been fulfilled.

Since the introduction of the monthly and internationally comparable labour market statistics based on the ILO concept from reference month January 2005, the Federal Statistical Office has published monthly and quarterly employment data for the German economy as a whole each about 30 days after the end of the given reference period. First quarterly employment figures are released together with the monthly result for the last reference month of the previous quarter.

Hence Germany has succeeded in shortening the publication time lag for monthly and quarterly employment figures from 70 and 43 day, respectively, to 30 days each.

The question to be discussed below relates to the conditions and methodological advancement of employment accounts which have permitted the above-mentioned improvement of timeliness.

1. Methodological and conceptual aspects

- *Monthly calculations:*

The employment accounts as part of the German national accounts are based on a calculation model which is used to determine the average number of persons in employment in Germany during a given reference period. The model has been designed to allow, in a complex and integrated calculation procedure based on monthly computations, direct derivations of all other employment-related data with reference to periodicity, breakdown by economic activity, status in employment and labour accounting concepts. Monthly estimations as part of continuous labour market monitoring allow assessing and comparing the employment data with other labour market and short-term economic indicators that are available on a monthly basis, too (unemployed persons, registered unemployed, notified vacancies, orders received, business climate index, etc.). Monthly estimates ensure coherence in the overall employment and labour market statistical system which is basically geared to monthly presentations. Another characteristic feature of monthly calculations is that seasonal, irregular and other short-term effects can be covered earlier and estimations based on time-series analyses be made on a broader basis and with higher precision. As the quarterly employment figures are directly derived from the monthly data, plausibility of the quarterly results is ensured, too. Besides, continuous checks are made of the plausibility and consistency of the quarterly results with other national accounts aggregates such as gross value added, compensation of employees, and hours worked.

- *Combination of methods:*

The monthly employment accounts in Germany are not based on a single (rigid) procedure. They rather combine different methodological approaches like mathematical and statistical forecasting methods and also calculations and expert estimations based on all existing statistical employment data. The results of calculation, which are obtained independently of each other and in different ways, are then matched with the help of the primary data collected in the Federal Statistical Office's monthly telephone survey of the ILO employment status and are eventually combined to obtain an independent final result. This combination of different methodological approaches and procedures, on the one hand, ensures high flexibility and, on the other, permits the realisation of synergy effects. Thus, mathematical and statistical procedures, in particular the time-series analyses, are an important instrument to assess current trends and seasonal movements in employment. However, specific circumstances and irregular effects, such as the impact of new legal regulations in terms of employment policy, can only be considered in expert estimations.

2. Organisational and technical aspects

- *Multi-stage calculation procedure:*

To observe the specified publication time lag of 30 days after the end of a given reference period, the organisational and technical conditions of performing the calculation procedure must allow a rather easy and quick calculation of employment while, at the same time, supplying sufficiently precise results. To this end, the following procedure, which includes several successive stages of calculation, has been designed for the purposes of the German employment accounts:

1st stage: The results of the first stage of calculation are released after 30 days. They are largely based, on the one hand, on the results of single-step forecasts resting on time-series analyses which can be made rapidly and do not depend on the availability of data of other sources of employment statistics and, on the other, on source-based expert calculations and estimations on an aggregated, i.e. overall economic level. The above results obtained independently of each other are subsequently assessed and finally combined, considering the relevant data of the ILO telephone survey, which become available about seven days after the end of the reference month, and other macroeconomic labour market indicators and general economic trends.

2nd stage: The second stage encompasses employment calculations in an economic breakdown and in a breakdown by employment status. Those calculations include and process successively all employment information which becomes available at a later time. The computation system is complex and its conceptual approach is geared to the requirements of national accounts. The calculations allow the presentation of results in great detail. The Federal Statistical Office is in a position to publish quarterly data on persons in employment and employees in an economic breakdown about 45 days after the end of a given quarter.

3rd stage: Annual results of basic statistics with longer periodicities, which become available with time, allow the presentation and publication of annual employment accounts data in a highly detailed breakdown of up to 60 branches of economic activity.

More national accounts-related information on employment accounts and the results of labour market monitoring based on the ILO concept can be accessed at:

http://www.destatis.de/themen/e/thm_erwerbs.htm

Stephan Lüken, tel: (+49-611) 75-2016, e-mail: stephan.lueken@destatis.de

Sigrid Fritsch, tel: (+49-611) 75-2054, e-mail: sigrid.fritsch@destatis.de

Expenditure on and investments in information and communication technologies, ICT investments – survey pursuant to Art. 7 (2) of the Federal Statistics Law to meet new EU requirements in the context of structural business statistics

Users and data needs

Detailed information on the use of information and communication technologies and the resulting economic effects on productivity is required in many areas of economic policy. The European Commission has put the related data requirements into words, for instance, in the

context of the so-called Lisbon process¹⁰ and the i2010 initiative¹¹. Information on ICT usage and its effects is of decisive importance across the whole of Europe both to define current positions and fields of political activity and to control and review the results achieved.

In Germany, too, reliable information on ICT usage in enterprises and public institutions has gained in political importance. Part of the information requirements in this area have been met by the results of the survey of ICT usage in enterprises and the data issued in cross-section publications¹². However, information has so far been available neither on the extent and structure of ICT expenditure in enterprises nor on its economic impact. In particular, reliable information has not been obtained with regard to the effects of ICT expenditure on the level of productivity. For this reason, a methodology has been developed together with Eurostat with the aim to gather the information needed.

The goal of the survey of ICT investments is to apply the above methodology in practical circumstances and to study the conditions under which results of high informational value can be obtained. Data collection has been designed with the final aim to provide a statistical basis for deriving indicators which will show the relationship between ICT investments and productivity. To this end, the availability of quantitative information is of particular relevance.

Concept and implementation

To begin with, the current survey of ICT investments has above all been designed to test the methods of collecting and evaluating the relevant information. The frame of the data collection concept was worked out in close co-operation with Eurostat and co-ordinated with the WPIIS¹³ expert group of the OECD. In line with the formulated information requirements, data collection has two parallel aims: on the one hand, to obtain monetary information on the investment activities of enterprises and public institutions in the ICT sector and, on the other, to gather expert information on the effects of ICT usage on productivity.

The national approach to implementing the framework concept encompasses the following three content-related survey modules:

- ICT investments in enterprises
- ICT investments in public institutions
- expert information on the productivity effects of ICT usage

The EU project on Investments in information and communication technologies has been carried out in Germany as a survey pursuant to Art. 7 (2) of the Federal Statistics Law. In co-operation with the statistical offices of the Länder of Brandenburg, Lower Saxony, North Rhine-Westphalia, Saxony and Thuringia, voluntary-response surveys have been conducted among approximately 20,000 enterprises and institutions of almost all branches of the economy since early 2006.

¹⁰ The goal of the Lisbon process is to increase productivity and economic dynamism in the European Union. In this context, the promotion and evaluation of ICT usage plays a decisive role (see also: Common Actions for Growth and Employment: The Community Lisbon Programme, http://ec.europa.eu/growthandjobs/pdf/COM2005_330_en.pdf).

¹¹ i2010 is a comprehensive strategy for modernising and deploying all EU policy instruments to encourage the development of the digital economy (cf.: Commission launches five-year strategy to boost the digital economy, http://europa.eu.int/information_society/eeurope/i2010/docs/press_release_en.pdf).

¹² See also: IKT in Deutschland, Informations- und Kommunikationstechnologien 1995 – 2003, <http://www.destatis.de/download/d/veroe/iktvoe.pdf>.

¹³ Working Party on Indicators for the Information Society

The enterprises covered were selected from the business register based on random sampling in a breakdown by economic branches (11 strata), employee size classes (3 strata) and regions (4 strata).

Before the survey actually started, a draft questionnaire had been sent to selected enterprises and institutions, asking them to make proposals for improvements which would help respondents complete the questionnaire (pretest). As a result of that test, the list of questions was – as expected - adjusted and made more precise above all with respect to the questions regarding ICT usage and its effects on productivity.

List of variables

The above-mentioned three survey modules are components of the same questionnaire. The structure of the questionnaire is characterised by two main sections. In the first section, both enterprises and public institutions are asked to give information about quantitative investment items. In this context, they are asked to enter data regarding their expenditure on IT (hardware) and telecommunication goods and also the software acquired and the enterprises' own-account production of software. Other questions concern the expenditure on ICT services and the leasing of ICT goods.

Almost all answers to the questions of the above section can be derived from the accounting records of enterprises. However, an exception is the expenditure on enterprises' own-account production of software since, pursuant to the provisions of the German Commercial Code, the latter must not be entered as an investment on the assets side and is consequently not included in the cost accounting system of enterprises. For this reason, the relevant expenses will be determined based on information requested on staff assignment (person-days) and related estimates of costs per day.

The second section of the questionnaire includes questions on ICT usage and its effects on the productivity of enterprises. In this context, information is requested on both the fields of ICT usage and the relevant success factors. Further questions concern the use of open-source software. Those questions are primarily directed to the IT experts of enterprises and public institutions.

Including the above three modules in one survey provides the opportunity to combine information on the types and scope of ICT investments with the IT experts' assessment of the development potential and the effects of new technologies. In this way, for instance, it can be determined whether enterprises which operate all their ICTs via an external service provider judge the potential effects of ICT usage differently from organisations which develop and operate their information and communication technologies themselves. Also, a cross-branch comparison could be made regarding this issue. Many studies on productivity and ICT usage support the thesis that the effects of using information and communication technologies depend largely on the organisational framework and the ICT usage strategies followed in an enterprise¹⁴.

Current experience

The survey is currently in the field phase. As early as now, however, some information can be provided regarding the trends observed in implementing the methodological concept. The completed questionnaires so far received show that the response rate can be expected to be similar for all strata of the sample. However, it is still difficult to state anything about possible nonresponse biases. Some few enterprises and institutions indicated that they did not fill in the questionnaire because the questions posed were not relevant to them. Particularly small units in traditional branches of economy activity (crafts, conventional services, construction) refused to

14 Cf. also: Information Technology and Productivity: It ain't what you do it's the way that you do I.T., http://www.eds.com/services/whitepapers/downloads/lse_productivity.pdf

answer the questions, indicating that the latter were not of any relevance to them. This trend has been confirmed by the number of questionnaires returned from the individual strata.

The responses received do not give any indication that the questions caused any specific difficulties for the respondents – neither in terms of content, nor because of their structure or wording. Also, the necessary thematic structure of the questionnaire comprising a quantitative and a qualitative section (the latter to evaluate the ICT effects on productivity) did in general not adversely affect the completeness and plausibility of the answers. However, this statement applies only to a limited extent to affiliated enterprises whose IT has been centralised or outsourced. Those enterprises often face problems in providing information on strategic orientation. Besides, they frequently have difficulty allocating expenditure and investments to (parts of the) enterprise.

Response burden

To begin with, the sample of the ICT investment survey has been designed in a way to avoid, to the extent possible, the inclusion of enterprises which are among the respondents reporting for the purposes of the survey of ICT usage. Though the questions of the two ICT-related surveys do not overlap in terms of their content, sending two questionnaires to the same respondents would not only increase the statistical burden on them, but also raise concerns regarding a potential deterioration of the response behaviour in terms of quantity and quality.

To determine the actual burden imposed on the respondents of the ICT investment survey, the questionnaire includes questions about the time a respondent has required for completing it and the problems he/she has encountered in answering the questions. For the time being, the following conclusions can be drawn from the feedback received: The majority of enterprises require less than 25 minutes to fill in the questionnaire. Hardly any problems have been reported with respect to answering the quantitative questions (expenditure and investments) as almost all of the data required could be derived from the accounting records of the enterprises. As regards the evaluation of the effects of ICT usage on productivity, it can be stressed that the problems faced in answering the relevant questions become smaller the larger the enterprises are.

Prospects

The field phase will be completed in early July 2006. First results will be available by the end of October. A comprehensive project report including both the methodological evaluation regarding the implementation of the framework concept and detailed results will be published in the first quarter of 2007.

Jörg Feuerhake, tel: (49-611) 75-4116, e-mail: joerg.feuerhake@destatis.de

Effects of the recent Collective Agreement for the Public Service (TVöD) on the public service personnel statistics

Data collection method so far used for compiling statistics on public service personnel

The public service personnel statistics provide data on the persons employed by public employers with the reference date of 30th June each year. The relevant data are largely supplied by payroll accounting centres.

The statistics on public service personnel serve as a basis for detailed studies of the personnel and organisational structure of the public service sector. The results of the studies are used especially by policy makers, on the one hand, for further developing the Public Service Law and the Laws Governing the Remuneration of Public Officials, Collective Agreements and Public

Officials' Pensions and Allowances and, on the other, for medium-term financial budgeting. In addition, the information provided by the statistics on public service personnel is used as a basis for developing gender equality concepts and carrying out benchmarking comparisons particularly at the municipal level. For this purpose, data are collected, for instance, on pay grade assignment and the individual remuneration components.

In addition to the provisions of the Law on Public Officials, the statistics on public service personnel in the past covered the grading and remuneration variables of the Federal Collective Agreement for Salary Earners in the Public Service (BAT), the Framework Collective Agreement for Wage Earners in the Public Service (MTArb) and the Federal Framework Agreement for Wage Earners in Local Authority Administrations and Operations (BMT-G). In very few areas of the public service like, for instance, the health insurance sector, separate collective agreements were applied to salary and wage earners. To the extent possible, however, the pay arrangements etc. of those agreements were similar to the regulations of the BAT and MTArb.

Changes due to the recent Collective Agreement for the Public Service (TVöD)

On 13 September 2005, the Federation and the Local Authorities Employers' Association signed with representatives of the trade unions the Collective Agreement for the Public Service (TVöD).

The new collective agreement, which applies to public employees in the employment of the Federation or of an employer which is a member of the Local Authorities Employers' Association, entered into force in October 2005. The new agreement has caused a number of changes in the statistics on public service personnel, too.

Changing over from the BAT and BMT-G to the TVöD has been based on the provisions of what are called the Collective Agreement on Changing Over to the TVöD and the Transition Rights Agreement (TVÜ). The latter provides for the protection of the employees' vested rights.

The TVöD is another 'principal' collective agreement for the public service, which applies to the salary and wage earners in the employment of the Federation and to the by far largest part of the employees of local authorities. In addition, separate collective agreements were signed for certain branches of the public service (e.g. local transport, public utilities, water supply and distribution). However, the BAT and MTArb arrangements still applied to the vast majority of public employees at the Länder level on 30 June 2006.

The variables which form the basis for pay rate calculations differ between the individual collective agreements. Whereas the payment of salary earners in accordance with the BAT was based on pay grades and steps/steps depending on age, the TVöD provides for pay scales which do not correspond to the BAT pay grades. Furthermore, payment in the context of the TVöD does no longer depend on age. It is rather based on work experience.

To permit a comparison of structures despite the wide range of agreements, new items will be added to the grading (e.g. pay grade or pay scale) and step (e.g. step depending on age or depending on experience) variables and the variable **Type of collective agreement** be newly included in the statistics on public service personnel. Introduced for differentiation purposes, the new variable will only consider the major collective agreements in the public service sector.

With the aim to adequately depict the public service structure, the publication tables of the statistics on public service personnel have included data for both employees and public officials in a breakdown by service classes. Apart from pay grades, case groups - as they are called - have in some cases been used as a differentiating criterion for service classes.

Some of the pay scales of the TVöD (particularly E9) now include employees who were previously grouped under different classes of service. The TVöD grading has been structured in line with the pay grades of remuneration scheme A for public officials. The allocation to service classes does no longer distinguish between entrance and top offices. Hence all employees grouped, for

instance, under pay scale E9 are assigned to the class of higher intermediate service. As a result, the personnel structure will change in service class-related statistical presentations. The data will not be comparable to those of previous years. Pay scales 1 to 4 are assigned to the ordinary service, while scales 5 to 8 are assigned to the intermediate, 9 to 12 to the higher intermediate and 13 to 15 to the higher service.

The TVöD does not distinguish between salary and wage earners. Consequently, this approach will be abandoned in public service personnel statistics, too. Both wage and salary earners will in future be covered by the concept of **employees**. However, the difference between nursing care staff and other employees will continue to exist as the TVöD provides for different pay rates in this area.

Prospects and further changes in the public service personnel statistics

For a couple of years, the standard weekly working time has differed at the Länder and municipal levels so that the magnitude of full-time equivalents cannot serve as a basis for comparing the actual volumes of labour. Full-time equivalents are calculated with the help of so-called working time factors. The latter are however based on payment rather than working time items, meaning that a working time factor equals 100% if 100% of the amounts listed in the remuneration tables are paid out. Consequently, the different working time arrangements are only partly reflected by full-time equivalents. As the underlying working time may differ, full-time equivalents can indicate the actual volumes of labour to a limited extent only. From reference year 2006, information will be additionally collected on the **standard weekly working time**.

More changes can be expected to be necessary in the context of the planned reform of federal structures in Germany because the responsibility for all matters concerning the public officials of the Länder under the Laws Governing the Remuneration of Public Officials and Public Officials' Pensions and Allowances will be shifted to the Länder authorities. However, the extent to which it will be possible to reflect the specific regulations of the Länder in the statistics cannot yet be determined at the moment.

Susan Kriete-Dodds, tel: (+49-611) 75-4105, e-mail: susan.kriete-dodds@destatis.de

Events

The 1996 – 1999 Microcensus Panel as a Scientific Use File: An Introduction ZUMA Workshop on 14 September 2006

The microcensus is a rotating panel sample survey that interviews households in a sampling district over four years: Every year one quarter of the sampling districts is replaced. For the first time, a Microcensus Panel has recently been made available as a Scientific Use File. It relates to a period from 1996 to 1999, covering some 120,000 individuals and 55,000 households for each survey reference period.

The focus of the workshop will be to impart basic application-oriented knowledge of this new type of microcensus data. The workshop is primarily addressed to scientists, who have already worked with cross-sectional data of the microcensus, but it is open also to interested scientists having no previous knowledge of the microcensus.

The workshop will give an overview of survey design, consolidation of cross-sectional data and the construction of new identifiers and extrapolation factors for longitudinal analysis. Examples will be given to describe evaluation cases occurring in practice.

The workshop will be held in close cooperation with the Federal Statistical Office (division VIII C – Microcensus, Labour Force Sample Survey, Household and Family). Talks will be contributed by members of the Shared Methods Project “Processing and Provision of Microcensus Data as a Panel Sample Survey”.

- [Programme](#)
- [Overview](#)

Notice of intention to attend: by 31 July 2006

Attendance fee: € 60, students € 40

Number of participants: max. 18

Interested individuals are requested to inform the ZUMA-Meetings Secretariat of their intention to attend (workshop@zuma-mannheim.de, tel.: 0621-1246-221).

1st Users' Conference on the Sample Survey of Income and Expenditure (*Einkommens- und Verbrauchsstichprobe - EVS*)

The Federal Statistical Office, in cooperation with ZUMA, will hold the 1st EVS Users' Conference at Wartburg Hotel in Mannheim on 19 and 20 October 2006.

The conference addresses scientists, who have already worked with EVS microdata or are interested in using EVS data.

For more information about this conference, please, refer to the destatis Internet website under: <http://www.destatis.de/allg/e/veranst/proservke.htm>.