7 th WORKSHOP ON LABOUR FORCE SURVEY METHODOLOGY

DATA PROCESSING AND DATA QUALITY

Madrid, Spain, 10 – 11 May 2012

C. Data processing: Methods for editing and imputation, weighting, non-response adjustment

C3– Calibration by age groups and its impact on key labour market indicators

Katharina Puch– Germany

Calibration by age group and its impact on key labour market indicators

With the introduction of the monthly Labour Force Survey, two weighting models were developed both using calibration to auxiliary variables: One of the models allows examining the quarterly and annual results for nearly all variables at highly regionalised levels and the other allows analysing monthly results at national level.

Due to the different weighting methods, (but also for other reasons) the monthly results differ from the annual results. To quantify the impact of the weighting model on those differences, a project was carried out testing a new weighting model for the quarterly and annual results. For this test the detailed regionalisation as well as a deeper breakdown of auxiliary variable "age" was introduced in the annual model.

First, the two existing and the tested weighting models are briefly described to give an overview. Then, the tested weighting model is applied to be able to compare the annual results. The effects on the main labour market indicators are presented below and in the associated presentation.

1 The weighting models for quarterly and annual results

The weighting model of the annual results is a two-stage procedure. In a first step, to eliminate sampling errors and non-sampling errors, known non-response is compensated for by means of information on households who did not respond. In a second step, selected auxiliary variables are adjusted to calibration marginals from Current Population Statistics and from the Central Register of Foreigners. The auxiliary variables used here are three age groups (under 15 years, 15 to 44 years, 45 years and over) and four citizenship groups (German, Turkish, EU-25 and non-EU-25), always in a breakdown by sex. Adjustment is made per quarter at different regional levels (Land [equals NUTS1], administrative region [similar to NUTS2] and regional adjustment stratum[1]). The compensation factors and weighting factors are obtained by means of a Generalized Regression Estimation (GREG), so that very small frequencies in the calibration classes can be avoided.[2]

2 The weighting model for the monthly results

The weighting method of the monthly results is one-stage. For reasons of timeliness, the monthly results are expanded without separate compensation for non-response. However, those results, too, are adjusted to calibration marginals from Current Population Statistics, which means that the data are compensated indirectly. The selected calibration variables are age (in 13 age groups with five age years each) and two citizenship groups (German, non-German), each broken down by sex.

---

[1] Regional adjustment stratums are usually a grouping of several administrative districts ("Kreise") or even urban municipalities with at least 500 000 inhabitants.

[2] For detailed information on the weighting procedure see Afentakis, A./Bihler, W. 2005: Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005. In Wirtschaft und Statistik 10/2005, p. 1039ff.

As data are available only for part of the sample when the monthly weighting is done, the calibration variable "age" is broken down to a more detailed level than is done for the annual calibration so that a highly differentiated adjustment can be achieved. However, the highly detailed regionalisation was not applied to the calibration of monthly results. The regional breakdown (Land, "Nielsen areas", former territory of the Federal Republic and New Länder) is less detailed than that of the annual estimation.

The final weighting factors are calculated by means of the same calibration method as described above (Generalized Regression Estimation). This allows adjustment to separate marginal distributions.[3]

Table 1 displays the differences between the two above mentioned weighting models and the tested one for quarterly and annual results using a more detailed age structure. Therefore, the only difference between the actual model for the quarterly and yearly results and the tested model is found in the stratification of the auxiliary variable "age".

Table 1: Overview of auxiliary variables for calibration of the different weighting models[4]

Regional units	Month	Quarter/Year	Variant tested
Total Germany	• 13 age groups × sex • calibration to weeks	no calibration	no calibration
2 units (west-east)	• 2 citizenship groups × sex	no calibration	no calibration
8 Nielsen areas	• 6 age groups × sex	no calibration	no calibration
16 Länder (NUTS1)	• sex	• 3 age groups × sex • 4 citizenship groups × sex • soldiers, persons doing compulsory military service • total population per month	• 19 age groups × sex • 4 citizenship groups × sex • soldiers, persons doing compulsory military service • total population per month
39 administrative regions (NUTS2)	no calibration	• 2 citizenship groups x sex	• 2 citizenship groups x sex
132 regional adjustment strata	no calibration	• total population	• total population

3 Tested weighting model - integration of the more detailed auxiliary variable "age"

Combining the two weighting models allows, for the analysis of annual results, maintaining the desired level of regional breakdown and, in addition, an adjustment to more detailed age groups. The tested model here involves the same approach as the annual weighting method - with the exception of the auxiliary vector of "sex and age per Land", where 19 age groups are used instead of three. Apart from that, the comparisons shown below are based on the same annual survey results. Due to the calibration method applied in the weighting model, empty cells are avoided also in the adjustment to 19 age groups. The random sampling error (relative standard error) is even somewhat smaller than in the actual annual weighting method.

---

[3] cf. Statistisches Bundesamt 2008: Quality Report "Monthly unemployment statistics according to the ILO concept".
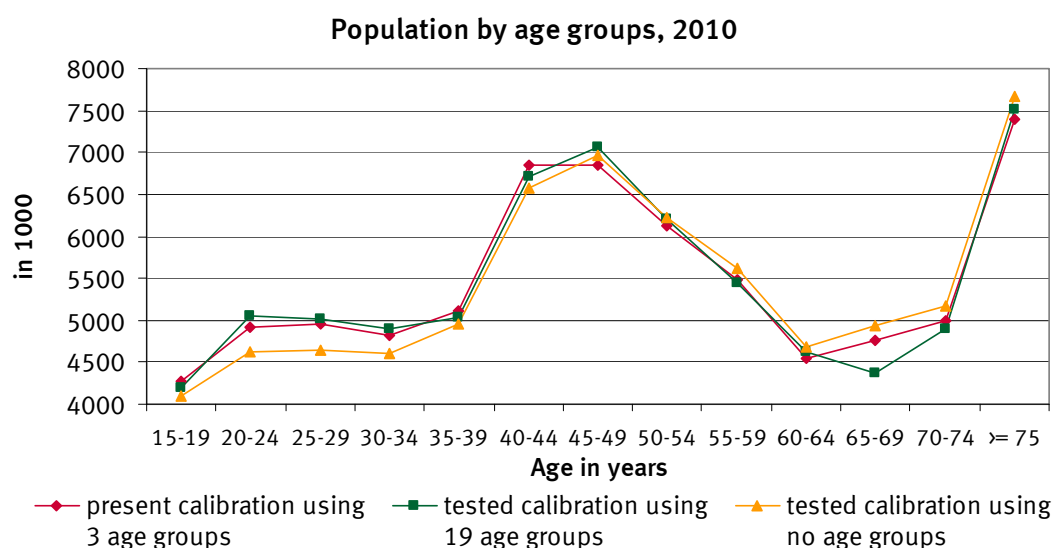
[4] Simplified illustration.

4 Comparing the results using different weighting models

Performing an adjustment to the more detailed breakdown of age groups in the weighting procedure has two consequences. First, cohort-related jumps in (time series) results are reduced. Second - and this is more important for the interpretation of labour market data - it provides detailed results regarding analyses for the highly correlated survey variable of age.
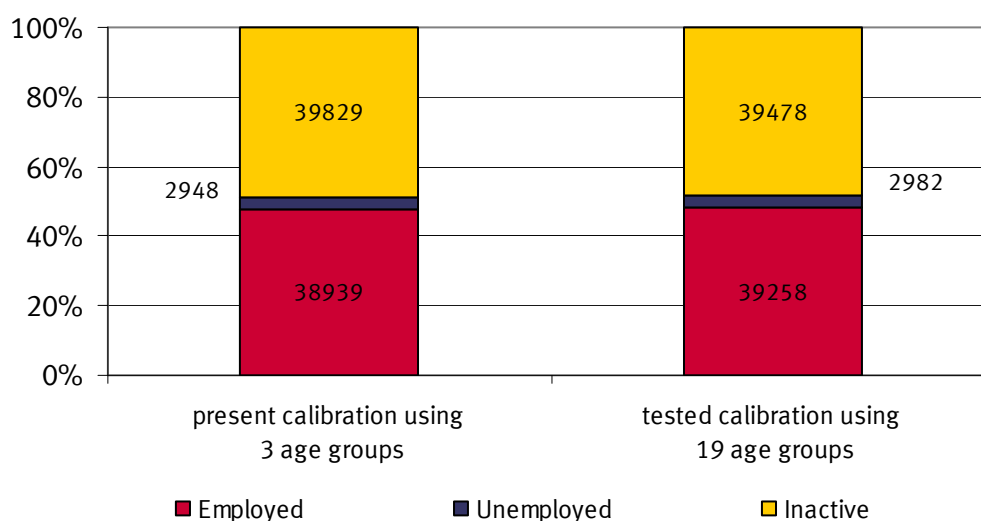
Applying the modified weighting factors for the calculation of results for 2010 provides new results when analysing the population by age. The population in the age groups of the 45 to 49 year olds is by more than three percent higher and those belonging to the age group 50 to 54 years by another 1.5 percent (see figure 1). It should be taken into account here that the group of the 45 to 54 year olds (1) is part of the labour force and (2) belongs to the highest of the three age groups in the current weighting model of the Labour Force Survey. As the estimation models accounting for age both have an increasing and a decreasing effect to specific age groups, the quite substantial group of 45 to 54 year olds (16% of the total population and 27% of the labour force) are important. Consequently, there is an impact on the results of variables like employment or unemployment, where that age group is particularly strongly represented. Examining the results for the overall population shows that the weighting-related increase in some age groups is levelled out by weighting-related decreases (e.g. in the age group of the 65 to 69 year olds). For results where only specific age groups are examined, however, this counterbalancing effect does not occur.

Figure 1: Comparison of the age structure of the population using different weighting models

The adjustment in the weighting procedure has an effect in particular on the results by employment status (see figure 2). The number of persons in employment and of unemployed persons rises by 0.8%, respectively 1.1%. The number of inactives is by 1.2% lower in the weighting model tested. In some age groups, the more detailed breakdown of age groups is particularly obvious, leading to a difference of more than 3% between the results.

Figure 2: Comparing the results for the population by employment status (2010)

5 Conclusion

Comparing the two weighting models for the yearly data shows that the standard errors caused by the tested weighting procedure do not change much - they even decrease slightly. The results for individual groups of people, however, change in part considerably, especially if analysed by age.

It can generally be assumed that a calibration to age groups with a more detailed breakdown leads to more coherent results. As such detailed information from the Current Population Statistics is currently not available at a quarterly basis it is not possible at present to do an adjustment to 5-year age groups. However, that weighting model - or the element of age adjustment to 5-year age groups - should be included in a modification of the weighting procedure in the context of the redesign of the system of household surveys planned in 2017. This applies in particular with a view to the Regulation on the organisation of the LFS[5], which requires a breakdown of the calibration variable of age by 5-year age groups also for the weighting of the quarterly and annual results.

Katharina Puch, tel: +49 (0) 611 / 75 41 06, e-mail: katharina.puch@destatis.de

---

[5] See Council Regulation (EC) No 577/98, Article 3 (5): "The weighting factors are calculated taking into account in particular the probability of selection and external data relating to the distribution of the population being surveyed, by sex, age (five-year age groups) and region (NUTS II level), where such external data are held to be sufficiently reliable by the Member States concerned." Official Journal of the European Communities, L 77 of 9 March 1998.