

Comparison of two different BTS weighting systems in the services sector of Latvia

Ieva Vanaga
Central Statistical Bureau of Latvia

10.10.2016

Historical introduction on BTS in Latvia

The history of Business Tendency Surveys in Latvia dates back to 1993. The first ones were Industry and Construction Surveys, followed by the Retail Trade Survey (1996) and the Investment Survey (2001). The last survey of such kind was the BTS of the services sector in 2002. These surveys were initially carried out by the Statistical Institute of Latvia, but in 2007, when the Institute was reorganised, BTS were overtaken by the Central Statistical Bureau of Latvia, where the Business Tendency and Territorial Statistics Section was established.

Role of the CSB in the field of BTS

The main objective of the CSB is to provide users of statistical information with independent official statistics that are of high quality. We plan and organise surveys, collect data, calculate and publish results in the field of Business Tendencies. Business tendency statistics has developed in step with the general statistical system of Latvia – we have introduced new technologies both in data collection and in the development of data processing. However, the CSB does not perform in-depth economic analysis of Business Tendency data, as this work is carried out by our professional users.

Historical weighting system of BTS data in Latvia

One of the conditions for providing high quality data is the revision and improvement of the methods used. The object of interest in the field of BTS is the primary data weighting system.

The weighting system used for all BTS in Latvia is very simplified.

Enterprises are stratified by NACE and by number of employees (3 size groups). Primary data of each enterprise are weighted with coefficient 1, 2 or 3 depending on the size group. The thresholds of size groups differ depending on the survey.

These weights are used in order to calculate the results in the necessary NACE breakdowns. In order to calculate the overall results of the whole sector (manufacturing, construction, retail trade or services sector), the share of every NACE group in the respective sector is used.

I presume that this weighting method was initially used as a very simplified and unified model of significance of small, medium and large enterprises.

The advantage of this method is its simplicity – individual weights of enterprises stay unchanged throughout the whole year, while the current list of respondents is effective.

Correlations with hard data/reference series are quite good (80-90%), and they are slightly lower for industry survey. The BTS data looks flat in comparison with hard data trends.

Steps of the activity

The real question is how adequate can the description of real tendencies be if such a simplified method is used?

In order to answer this question, several steps of methodological work must be carried out:

- 1) Calculation of BTS results by using multiplication of sample weights and the number of employees as individual weights for primary data.
- 2) Comparing results of both weighting systems.
- 3) Comparing BTS data (both systems) with hard data.

We have currently completed only the first two steps of the list. We must also carry out seasonal adjustment in order to realise the last step, however, the accumulated time series of the alternative method are too short for seasonal adjustment. Therefore, all data series of this paper are not seasonally adjusted.

Subject of the activity

We chose a BTS in the services sector as the subject of the experiment, because:

- 1) The services sector includes many subsectors (30).
- 2) The respective subsectors are very different (by nature of economic activity, by number of enterprises in the sample frame, by distribution of enterprises in size classes, and by their contribution within the total services sector).

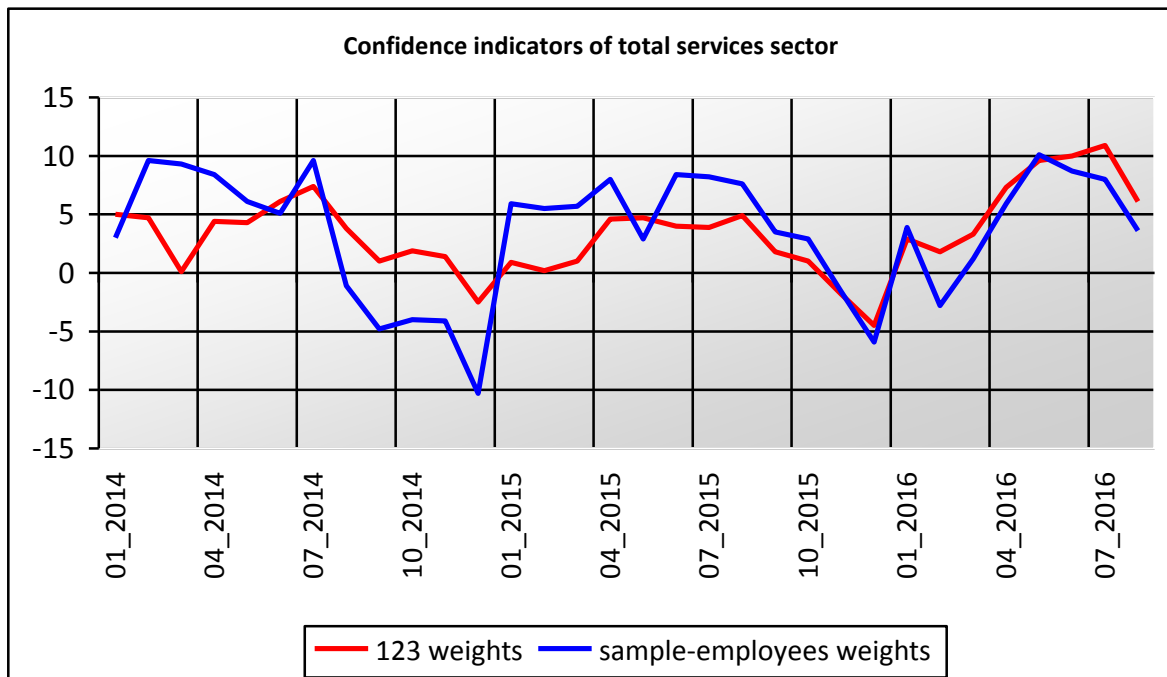
The frame of BTS in the services sector in Latvia includes approximately 26 000 (in 2016) enterprises stratified by NACE (30 groups at the 2-digit level) and by the number of employees (3 groups: 1-9; 10-49 and 50+). The fact that characterises Latvia as a small economy is that there are only 600 frame enterprises that belong to the “50+” group.

General results

We have done parallel calculation of results beginning with data on January 2014.

The first general results have been summarised in Chart 1.

Chart 1



Overall the red line that indicates 123 weights is more flat. The maximums and minimums of the blue line (sample-employees weights) are more expressive. The correlation between

the lines over the first six month of 2014 looks poor. May 2015 also stands out from the rest of the data set. The correlation between the series is 0.65.

During the work we moved from general to more detailed data, therefore the next step was analysis of subsectors.

We compared confidence indicators that were calculated by using both methods in all 30 NACE subsectors included in our survey. As a result we came to the conclusion that the confidence indicators of half of the subsectors that were calculated by using both of the methods correlated very well. Medium correlation was recorded in 7 subsectors, whereas the estimates of correlation were quite poor in 8 subsectors (Table 1). The most surprising results were observed in the accommodation sector – the lines almost overlapped.

Table 1

Coefficient of correlation (r)	Subsectors (NACE Rev 2)
High ($r \geq 0,7$)	50, 51, 55, 56, 59, 60, 61, 64, 66, 72, 77, 79, 80, 81, 82
Medium ($0,5 \leq r < 0,7$)	49, 58, 65, 68, 70, 75, 78
Low ($r < 0,5$)	52, 53, 62, 63, 69, 71, 73, 74

We try to carry out a detailed analysis of every NACE group that shows low correlation, as well as detect reasons for the significant differences between results.

Factor 1 – dominating enterprises

The initial hypothesis was that the factor posing negative influence on the correlation was the existence of dominating enterprises within the NACE group.

This factor was topical in 8 NACE groups: 50, 51, 52, 53, 60, 61, 63 and 65, however, after a more detailed analysis we established that the situation in groups is quite different (Chart 2).

In groups **50** and **51** the total number of enterprises in the sample frame is very low (38 and 18), and answers of large enterprises possess high influence on the results using both methods. Values of confidence indicators differ significantly, whereas intervals of increase and decrease are the same in general. This means that by using sample-employees weights in the respective groups, we increase the absolute value of confidence in comparison with the 123 method.

In group **60** the lines are very close to each other and the reason for this is that the dominating enterprise of this subsector provides only neutral answers year after year.

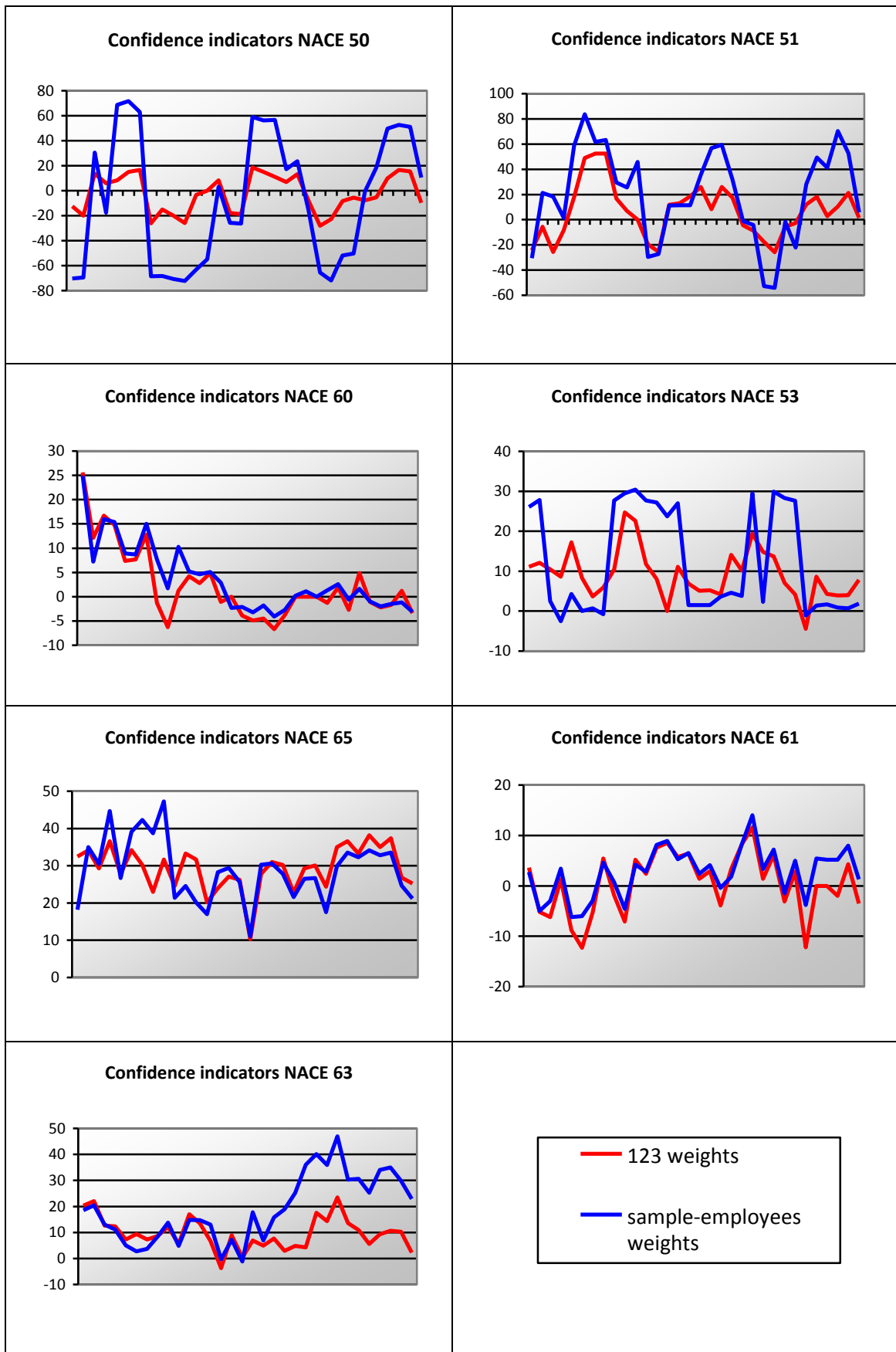
The dominating enterprise of group **53** also submits neutral answers, however, when sample-employees weights are used, the influence of the respective enterprise is so high that in separate cases, in which the answers differ from neutral, the correlation decreases significantly.

The situation is more complicated in group **65**. In 2015, both dominating enterprises submitted answers that were almost neutral, therefore the lines are very close to each other. However, in 2014, the answers of the respective enterprises varied, therefore medium correlation has been recorded throughout the whole research period.

The reason for the correlation being very high in group **61** is only the neutrality of the answers provided by the dominating enterprise.

The low correlation in group **63** was caused by significant changes in the list of large enterprises the middle of the research period. As a result the lines are very close to each other until May 2015, but the appearance of a significant dominant enterprise leaves a noticeable impact on the correlation.

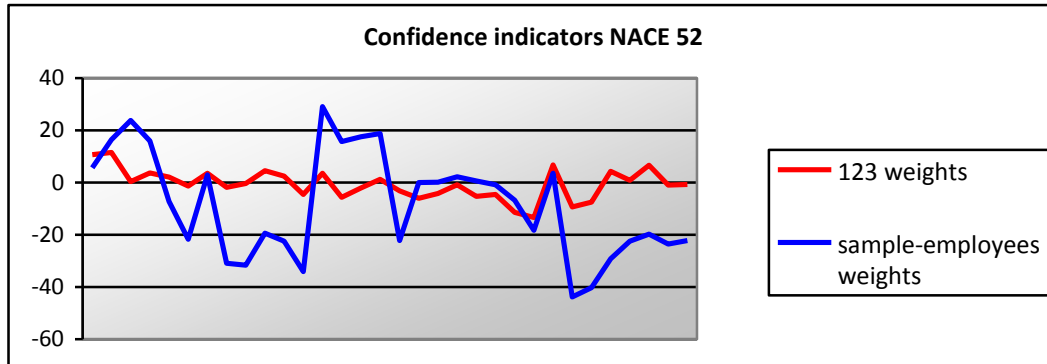
Chart 2



Special attention – NACE 52

Factor of dominating enterprise is also topical in group 52 and the correlation between confidence indicators in NACE 52 was low (Chart 3).

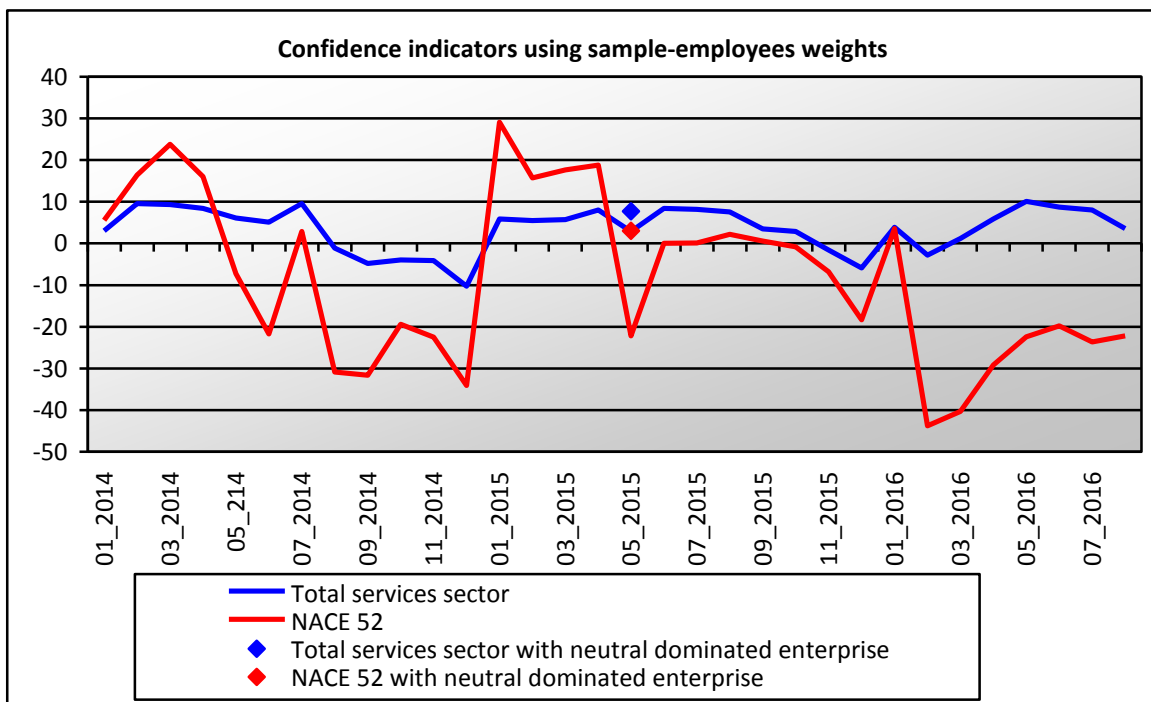
Chart 3



The difference between group 52 and the groups analysed above is that in most of NACE groups all enterprises of the size group “50+ employees” are included in the sample, because there is a rather small number of them. The number of large enterprises in group 52 allows carrying out real sampling in this strata. As a result the answers of the dominating enterprise are multiplied by a sample weight (~2.5) and the number of employees, and the results of NACE 52 are largely dependent on the answers of the respective enterprise. Since NACE 52 is one of the three subsectors providing significant contribution in the total services sector, the use of sample/employees weights without any corrections in stratification may also impact the results of the total services sector (Chart 4).

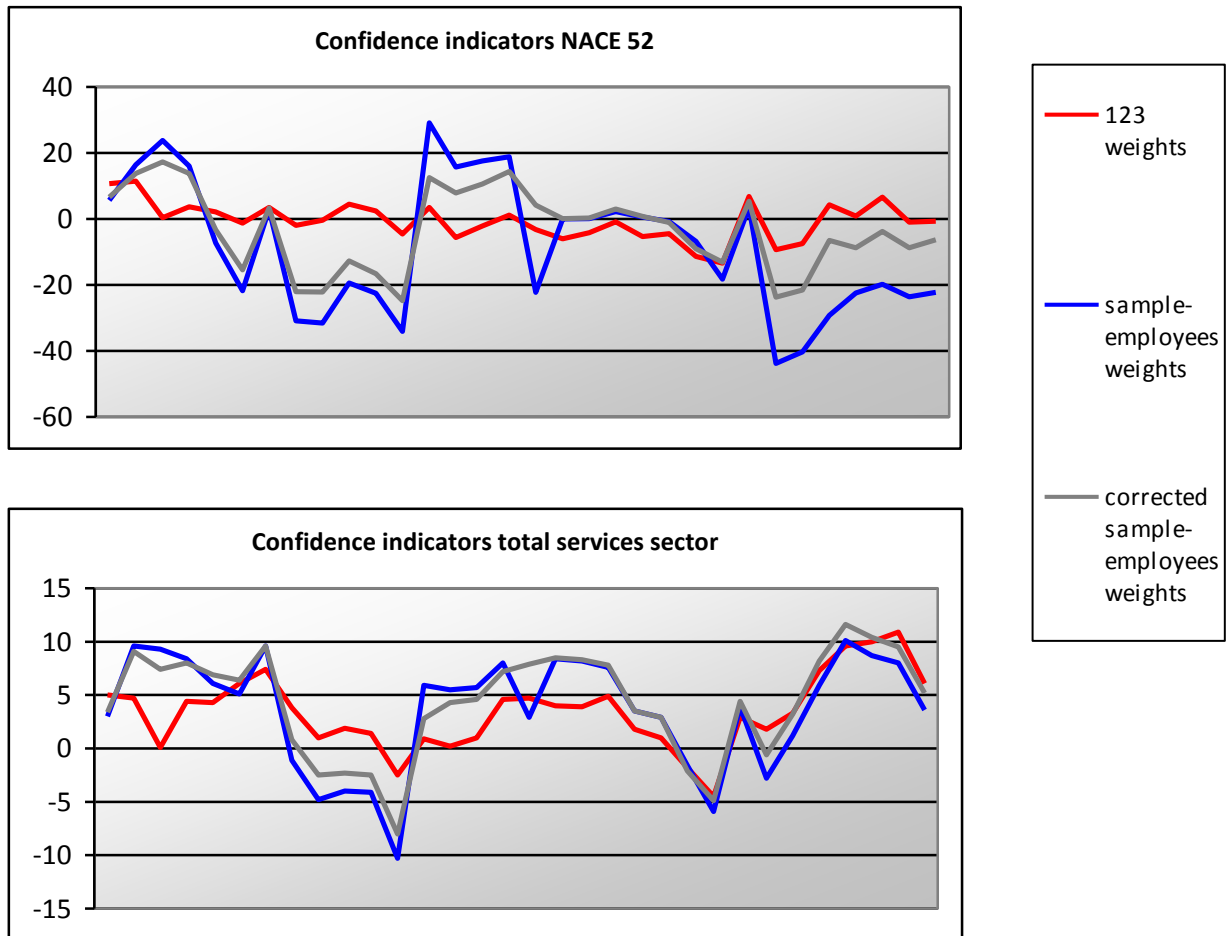
In Chart 4 it is possible to observe how the confidence indicators in NACE 52 and total services sector differ in May 2015 if real answers of dominating enterprises are changed with neutrals.

Chart 4



In order to prevent the inadequate influence that the dominating enterprise has on the total results, we marked the respective enterprises (and also all dominating enterprises of other NACE groups) as outliers with sample weight of 1 and did recalculations by using corrected sample-employees weights. There were significant changes in the results of group 52 and total services sector (Chart 5). The correlation between the confidence indicators of the total services sector calculated by using the 123 weights method and the sample-employees weights increased from 0.68 to 0.80.

Chart 5



Factor 2 – underrating small enterprises

Another factor that has influence on the correlation is the fact that the 123 method currently underrates the significance of small enterprises (1-9 employees) in the groups, where the role of small enterprises is the most important (decisive).

In small the small economy of Latvia the respective groups are 62, 68, 69, 71, 73 and 74. The distribution of frame and sample enterprises by size classes is presented in Table 2.

Table 2

NACE group	Number of enterprises within the sample frame by employees; 2016		
	1-9	10-49	50+
62	1 686	104	23
68	3 496	323	42
69	2 639	59	10
71	1 256	147	21
73	1 302	106	8
74	986	37	2

Statisticians constantly feel a greater pressure to decrease the burden on the respondents, especially small enterprises. Nevertheless, the data quality in the respective NACE groups can be improved by increasing the number of small enterprises within the sample.

Next steps

In order to carry out the further assessment of both weighting methods, it is necessary to carry out the seasonal adjustment of the acquired results and compare them to hard data. However, the time series acquired with the sample-employees weights method that are available at the moment are too short for seasonal adjustment.

Conclusions and questions for further activities

BTS indicators calculated by using 123-weights are relatively flat. When using this weighting method, the results may now show significant changes in large enterprises.

When carrying out yearly sampling, it is important to be very careful when considering the distribution of enterprises by size classes in each NACE group in order to prevent the underrating of small enterprises.

Certain explanations about BTS surveys and the significance of the answers of each respondent should be given to the enterprises that provide only neutral answers year after year, especially large enterprises.

Having carried out the analysis of the weighting methods, I am sure that a switch-over from the currently used 123-weighting method to the widely used sample-employees weighting would allow us to prepare BTS data that are of a higher quality and are more suitable for users.

Questions concerning the possible change of the weighting method in the future:

- 1) How long should the back calculation be in the case of a switch in methods?
- 2) Would it be necessary to change the weighting system of all 4 surveys all at once, or would it be possible to do this over a period of several years?