Final Report

V0 - 21. 10. 2013

Task Force on quality of business survey data

Task force 5: weighting approaches

1. Introduction

Business tendency surveys (BS) differ from most business surveys by their early release, the qualitative nature of their questions and the originality of the derived indicators (balances of opinions, indicators of confidence …). Their specificity is also attached to their non-intrinsic utility; the interest of business tendency surveys is based on their ability to forecast macroeconomic evolutions hence the question of how responses of the individual firms should be quantified and combined.

A debate on the best way to aggregate individual answers has occurred from the beginning of 50's. A consensus seemed to be found in the 70's by economic forecasters; balances of opinions were then considered as the most efficient way to quantify and combine responses of individual firms. However, the issue how to weight the results remains an open question.

The most common way to weight individual BS answers is to take into account both the relative size of the firm (*weighting at firm level*) and the economic importance of its sector (*weighting at stratum level*). The question is how to choose the most efficient variables at firm and stratum levels. The question is also what "efficient" means.

The question of the real impact of weighting schemes may also be asked. The following statement can be found in the OECD handbook[1] on business tendency surveys: *"Furthermore, practical experience has shown that the balances are not very sensitive to the choice of weighting variables. In practice it is sufficient to use a single variable reflecting the general economic importance of the enterprise in weighting all the survey answers"*. However, the small size of samples and potentially evolving heterogeneity among firms suggest that this statement needs to be revisited.

2. Why and how to weight individual answers?

Two kinds of weighting schemes can be used in processing the answers of business tendency surveys: sample weights and size weights. In some cases, results are processed without weights.

2.1. Sample weights

Most BS samples are stratified samples. Processing results by using sample weights means that balances of opinions (B), calculated as the difference between percentages of respondents giving favourable (P_+) and unfavourable answers (P_-) are estimated by the "classic estimator" of stratified sampling.

Let's consider that the BS answers are associated with an auxiliary variable Y:

Y_i= 1 if the answer is "+ increased" *denoted* Y_{i+}

Y_i= 0 if the answer is "= remained unchanged" *denoted* $Y_{i=}$

Y_i= -1 if the answer is "- decreased" *denoted* Y_{i-}

$$B = P_+ - P_- = \frac{1}{N}\sum_i Y_{i+} - \frac{1}{N}\sum_i Y_{i-} = \frac{1}{N}\left(\sum_i Y_{i+} + \sum_i Y_{i=} + \sum_i Y_{i-}\right) = \frac{1}{N}\sum_i Y_i = \bar{Y}$$

B is the average of the variable Y. By using the "classic estimator" of an average for a stratified sample, the estimator of B (denoted $\hat{B}$) is as follows.

$$\hat{B} = \sum_{h=1}^{H} \frac{N_h}{N}\bar{y}_h = \sum_{h=1}^{H} \frac{N_h}{N}(p_{+h} - p_{-h})$$

where *h* is the reference index of stratum, Nh the number of entreprises in stratum h of the sample frame and nh the number of reporting units in stratum h of the sample.

All enterprises surveyed in stratum h have therefore the same weight $\boxed{w_i = \frac{N_h}{N \cdot .n_h}}$

$\frac{N_h}{n_h}$ is the sample weight of stratum h.

[1] Business Tendency Surveys: A Handbook (2003), page 37, §115

However, this way of weighting individual answers is generally speaking not used for business tendency surveys. Indeed, it is difficult to think that attributing the same weight to each firm whatever its size is the best choice for linking BS to macroeconomic data. Size weights are therefore generally preferred to sample weights in processing BS answers because the importance of the answers is expected to depend on the size of the reporting units.

2.2. Size weights

The idea of using size weights ***at firm level*** is that answers from a large firm carry more information on economic activity than answers from a small one.

If S_i refers to a firm size criterion (turnover, employment,…), the estimator of balances of opinions obtained in a stratum h is then defined as follows:

$$\hat{B}_h = \sum_{i=1}^{nh} \left(\frac{s_i \cdot y_{i+} + s_i \cdot y_{i-}}{s_h} \right) = \frac{s_{h+} + s_{h-}}{s_h}$$

where $s_h = \sum_{i=1}^{nh} s_i$

s_{h+} (resp. s_{h-}) is the total size of enterprises that declare ‘’+ increased’’ in stratum h (resp. “- decreased”)

In a second time, results ***at stratum level*** need to be weighted in order to correct the inequality of sample rates in each stratum and then to reflect the structure of the economy.

The so-called “Balance of opinions” is then obtained as a weighted average of the above estimator by stratum.

$$\hat{B} = \sum_{h=1}^{H} \frac{V_h}{V} \times \hat{B}_h$$

where V_h and $V = \sum_{h=1}^{H} V_h$ are non-random values. In general, V refers to value added, turnover or employment of the considered economic sectors.

This estimator is quite different from the “classic estimator” used for stratified samples. It is not based on sample weights but on weights calculated from a variable linked to firms’ activity (turnover, employment…).

Enterprises surveyed in stratum h have therefore an individual weight w_i depending on their size:

$$\boxed{w_i = \frac{V_h \cdot si}{V \cdot .s_h}}$$

3. Metadata and quality indicators

3.1. Weighting procedures used in the Member States

The answers of BS partners to the metadata questionnaire sent by DG-ECFIN illustrate the diversity of methods implemented for weights at firm level as well as for weights at stratum level (see tables below).

They confirm the fact that sample weights are generally not used: few BS partners don’t weight the answers. Turnover or employment data are the main variables used to aggregate individual answers.. At stratum level, value added is preferred to turnover. The majority of BS partners use a single variable to weight all the questions.

Industry:

On firm level		On stratum level	
Weight	**Number of replies**	**Weight**	**Number of replies**
Turnover (only)	**4**	Turnover (only)	**3**
Employment (only)	**8**	Employment (only)	**5**
Value added (only)	**3**	Value added (only)	**8**
Turnover or employment	**4**	Turnover or employment	**1**
No weight (due to sampling method	**2**	Sold production	**1**
No weight	**1**	Output, employment or exports	**1**
No answer or no detail	**4**	Employment or exports	**1**
		No weight	**1**
		No answer	**5**

Services:

On firm level		On stratum level	
Weight	**Number of replies**	**Weight**	**Number of replies**
Turnover (only)	**6**	Turnover (only)	**3**
Employment (only)	**9**	Employment (only)	**4**
Value added (only)	**1**	Value added (only)	**8**
Turnover or employment	**5**	Turnover or employment	**2**
No weight (due to sampling method	**2**	Sold production	**1**
No weight	**1**	No weight	**1**
No answer or no detail	**2**	No answer or no detail	**7**

Construction:

On firm level		On stratum level	
Weight	**Number of replies**	**Weight**	**Number of replies**
Turnover (only)	**3**	Turnover (only)	**4**
Employment (only)	**13**	Employment (only)	**6**
Value added (only)	**1**	Value added (only)	**5**
Turnover or employment	**3**	Turnover or employment	**2**
No weight (due to sampling method	**2**	Sold production	**1**
No weight	**2**	No weight	**2**
No answer or no detail	**2**	No answer or no detail	**6**

Retail trade:

On firm level		On stratum level	
Weight	**Number of replies**	**Weight**	**Number of replies**
Turnover (only)	**8**	Turnover (only)	**5**
Employment (only)	**6**	Employment (only)	**4**
Value added (only)	**1**	Value added (only)	**4**
Turnover or employment	**5**	Turnover or employment	**2**
Gross profit margin	**1**	Sold production	**1**
No weight (due to sampling method	**2**	No weight	**2**
No weight	**2**	No answer or no detail	**8**
No answer or no detail	**1**		

The choice of variables used to weight individual answers is limited to the information available for each reporting unit. In contrast, a wider range of information is available at stratum level, coming from either the sampling frame or from an external source such as national accounts for example. Whatever the information source, the choice of the variable may have an impact on the results; the weight of economic sectors usually differs in terms of production, value added or employment. The lower the level of dissemination will be, the more important the impact of the weighting variable.

To illustrate this point, let's take an example coming from French industry survey. In French industry survey, stratum level corresponds to the three-digit level of NACE. When calculating series at 'manufacturing industry level', balances calculated at stratum level are weighted by the weight of each sub-sectors in manufacturing industry (see table 1). The weight of sub-sector 30.3 will vary to 3 to 7, depending on the weight used: value added or turnover.

Table 1: Share of value added and production of 'Manufacture of other transport equipment' sub-sectors in manufacturing industry, in France

NACE		Value added	Production
30.1	Building of ships and boats	0	1
30.2	Manufacture of railway locomotives and rolling stock	1	1
30.3	Manufacture of air and spacecraft and related machinery	3	7
30.9	Manufacture of transport equipment n.e.c.	0	0
	Manufacturing industry	***100***	***100***

Source: Insee, National Accounts 2010

When calculating series at '*Manufacture of other transport equipment* level', balances calculated at stratum level are weighted by the weight of each sub-sectors in *manufacture of other transport equipment* (see table 2). The weight of sub-sector 30.3 will then vary to 74 to 86, depending on the weight used: value added or turnover.

Table 2: Share of value added and production of 'Manufacture of other transport equipment' sub-sectors in manufacture of other transport equipment, in France

NACE		Value added	Production
30.1	Building of ships and boats	10	6
30.2	Manufacture of railway locomotives and rolling stock	13	7
30.3	Manufacture of air and spacecraft and related machinery	74	86
30.9	Manufacture of transport equipment n.e.c.	3	1
	Manufacture of other transport equipment	***100***	***100***

Source: Insee, National Accounts 2010

For the French industry survey, the choice has been made to weight at stratum level balances related to activity by production. Balances related to employment are weighted by employment.

3.2. Quality indicators

DG-ECFIN has calculated two quality indicators for each confidence indicator (COF): **correlations** between COF and reference series and **months for cyclical dominance** (MCD).

Reminder*: calculation of COF indicators*

Balances taken into account in the COF		COF formula
INDU	**Q2** - Assessment of order-book levels **Q4** -Assessment of stocks of finished products **Q5** - Production expectations for the months ahead	(Q2 - Q4 + Q5) / 3
SERV	**Q1** - Business situation development over the past 3 months **Q2** - Evolution of the demand over the past 3 months **Q3** - Expectation of the demand over the next 3 months	(Q1 + Q2 + Q3) / 3
BUIL	**Q3** - Evolution of your current overall order books **Q4** - Employment expectations over the next 3 months	(Q3 + Q4) / 2
RETA	**Q1** - Business activity (sales) development over the past 3 months **Q2** - Volume of stock currently hold **Q4** - Business activity expectations over the next 3 months	(Q1 - Q2 + Q4) / 3

❑ Correlations between COF and reference series

Because of their early release, BS results are commonly used in economic forecasts. However, forecasters are required to bridge the gap between the qualitative nature of BS indicators and the quantitative nature of "hard" economic indicators. The econometric techniques used, called bridge-models, are usually linear, hence the importance to find balances or composite indicators which are the most correlated to economic variations. To track performances of BS indicators, correlations between COFs and references series have been calculated by DG-ECFIN.

***Reminder**: list of references series*

COF	Reference series
INDU	Production in industry - Percentage change compared to corresponding period of the previous year
SERV	Gross Value added in wholesale and retail trade, transport, accommodation and food service activities, information and communication, financial and insurance activities
BUIL	Construction production index - Trend cycle
RETA	Household and NPISH final consumption expenditure

The table below shows that correlation coefficients are larger than 0.75 for a third of countries. On the other hand, correlations are smaller than 0.5 for a third of countries in retail trade survey. They are even negative for one country in building survey.

Correlation coefficient	INDU	SERV	BUIL	RETA
[-1 ; 0[			1	
[0 ; 0.5[	5	2	2	8
[0.5 ; 0.75[	12	7	7	9
[0.75; 1]	9	12	9	7

Note: - correlation coefficients of 12 "COF industry" with reference series are between 0.5 and 0.75
- data are not available for all BS partners

These results have nevertheless to be qualified because of both the specific way to calculate COF indicators and the choice of reference series. For instance, it is not obvious that the industrial production should be calculated on a year on year basis since firms are supposed to assess the evolution of their production over the last three months compared to the three previous ones and not compared to the same months of last year. Besides, the weaker correlation observed for retail trade COF may not be a surprise for forecasters who commonly notice that bridge models are better suited to forecast some supply variables such as industrial output than to forecast demand components such as households final consumption. To forecast final consumption, behavioural equations are usually preferred to bridge models.

❑ Months for cyclical dominance (MCD)

The MCD is a measure of short-term volatility in time series. It is defined as the shortest span of months for which the **I/C** ratio is less than unity. **I** is the average month-on-month change (without regard to its sign) of the irregular component of the series and **C** is the trend-cycle component of the series.

The Months for Cyclical Dominance (MCD) is used to determine the minimum number of months before an improvement/deterioration in the time series can be interpreted with reasonable confidence as an improvement/deterioration in economic sentiment. The higher the MCD, the higher is the volatility of time series.

The table below shows that on a month-to-month basis, the average change in the irregular factor is larger than that in the cyclical factor in the majority of countries. Over four-month intervals, it is larger in a minority of countries except for the retail trade survey, and for the building survey to a certain extent.

MCD	INDU	SERV	BUIL	RETA
1	1	2	2	
2	11	7	5	1
3	10	9	11	11
4	3	7	5	9
5	1		3	3

Note: MCD=2 in 11 "COF industry" series

Once again these results have to be qualified because of both the specific way to calculate COF indicators and the real aim of reducing MCD. Indeed, if for some series the measurement errors are already relatively small it may not be possible to reduce MCD.

MCD indicates the number of months at which the average amplitude of the cycle-trend component will overtake the irregular one. Behind irregular component, two kind of residual erratic fluctuations can be distinguished: the part of the irregular fluctuations that is due to measurement errors (response errors, sampling errors, processing errors, ***such as non efficient weight of responses***,…) and the part that is due to real fluctuations (unusual weather, strikes,…). The first part of the irregular should be considered as "noise" by forecasters whereas the second part should be considered as real information. Balancing the costs and gains of reducing MCD will be possible only if the measurement errors are known.

4. Further empirical results

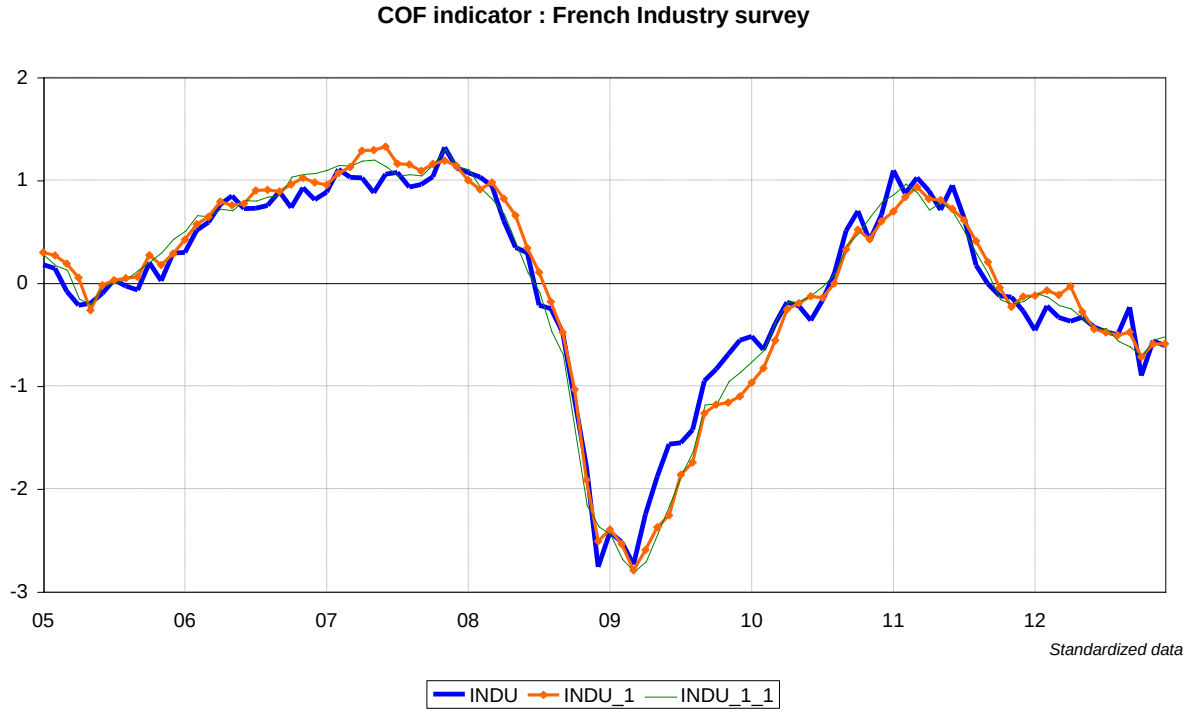
Even if size weights are commonly used, there is no theoretical evidence on the most efficient variable to use at firm and stratum level. This paper does not intend to give the definite solution to the question of weighting BS results.

More modestly we provide some experiments on the impact of a change in weighting schemes on time series in terms of correlations with reference series and MCD. However, a simple visual inspection of time series may already lead to some important conclusions.

More precisely, the rest of this paper presents the results of some experiments made on French industry survey data. First, series have been recalculated with no weight at firm level (denoted Y_1). Secondly, series have been recalculated with no weight neither at firm level nor stratum level (denoted Y_1_1). As recalculated series may not have the same long term average, graphs presenting these three series are based on standardized data (i.e. [Y-average]/standard error).

4.1. Impact of the weighting scheme on quality indicators

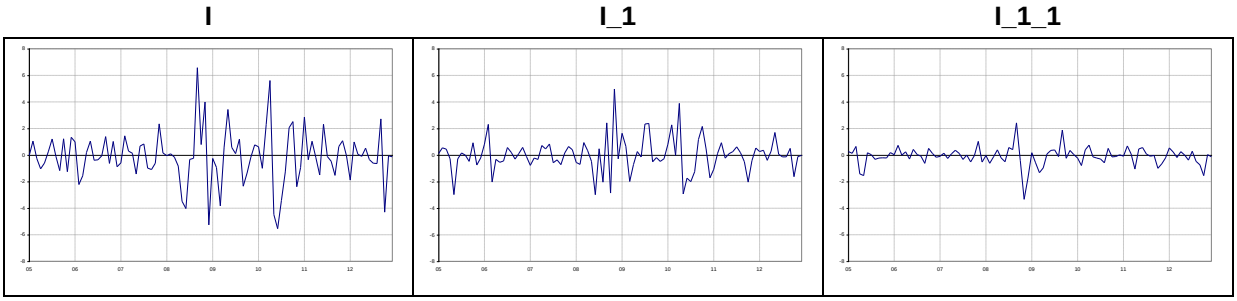
The graph below shows the three series of COF industry for France limited to the period 2005-2012. A quick visual inspection shows that in average all these curves vary in the same way. This is confirmed by correlation coefficents of each COF with the reference series: they are almost unchanged around 0.7.



Series INDU_1 and INDU_1_1 seem to smooth the original series INDU. This is confirmed by the visual inspection of irregular components ***I*** (see graphs below). If each curve highlights the end of

2008 period as an irregular one, I_1_1 seems to have softened all the irregular fluctuations even if they are due to real fluctuations.

Irregular components of the three COF industry series



The results of calculation of MCD for each series confirm that MCD of INDU_1_1 is lower than MCD of INDU (see table 3). MCD of INDU and INDU_1 is equal to 2. However, the ratio of the average amplitude of the irregular component to the cycle-trend one is lower for INDU_1.

Table 3: Average amplitudes of irregular and cyclical components for 1 to 10 months span

N° of months span	Weight at firm and stratum levels (INDU) MCD=2			No weight at firm level (INDU_1) MCD=2			No weight neither at firm level nor stratum level (INDU_1_1) MCD=1		
	$\overline{C}$	$\bar{I}$	$\bar{I}/\overline{C}$	$\overline{C}$	$\bar{I}$	$\bar{I}/\overline{C}$	$\overline{C}$	$\bar{I}$	$\bar{I}/\overline{C}$
1	1.22	1.87	1.54	1.09	1.28	1.17	1.11	0.76	**0.68**
2	2.37	2.09	**0.88**	2.12	1.53	**0.72**	2.19	0.87	0.40
3	3.50	0.96	0.56	3.12	1.53	0.49	3.25	0.76	0.23
4	4.59	2.07	0.45	4.11	1.45	0.35	4.25	0.77	0.18
5	5.65	2.00	0.35	5.07	1.40	0.28	5.22	0.76	0.15
6	6.67	1.93	0.29	5.99	1.42	0.24	6.13	0.76	0.12
7	7.63	1.84	0.24	6.89	1.47	0.21	6.97	0.81	0.12
8	8.50	1.97	0.23	7.72	1.44	0.19	7.78	0.80	0.10
9	9.31	1.89	0.20	8.45	1.41	0.17	8.51	0.82	0.10
10	10.07	1.83	0.18	9.13	1.40	0.15	9.18	0.72	0.08

$\bar{I}$: average percentage change (without regard to sign) in irregular component

$\overline{C}$: average percentage change (without regard to sign) in cycle-trend component

4.2. Visual inspection of time series

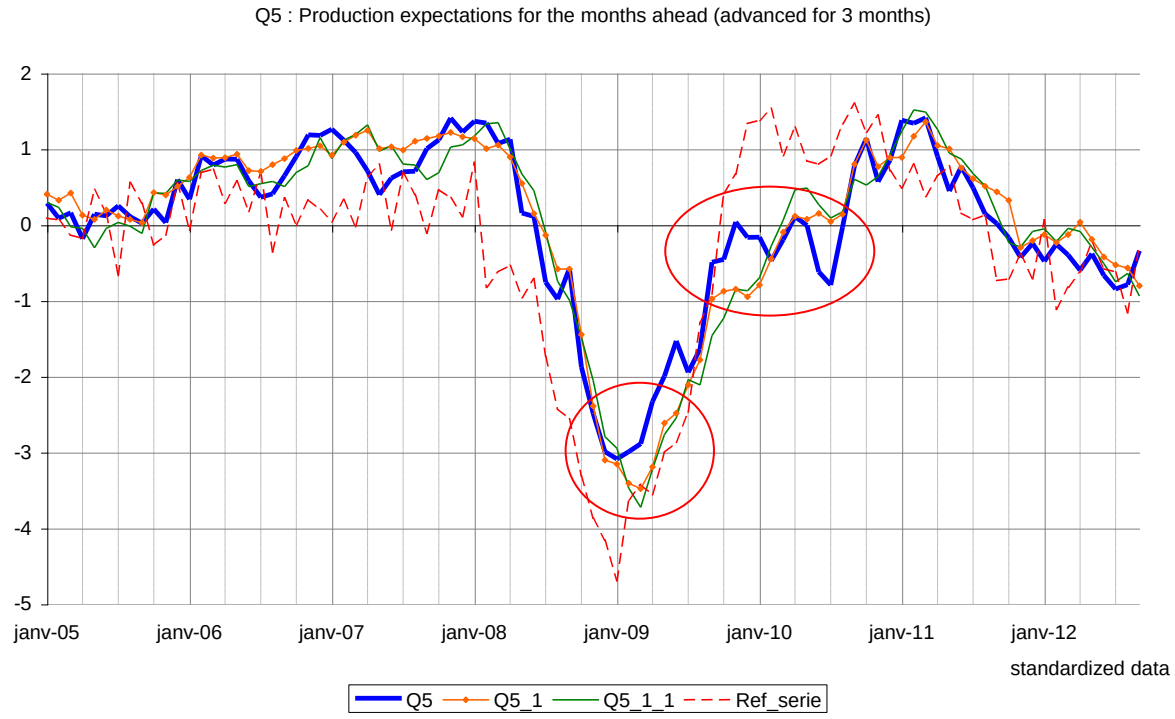
The consequences of a change in weighting scheme on quality indicators needs a qualification. A visual inspection of time series is an other way to assess the relative performance of weighted or unweighted series. Let's take the example of industry survey question Q5: 'Production expectations for the months ahead' and let's point out two periods: end 2008-beginning 2009 and end 2009-beginning 2010 (see graph below).

Period 1: end 2008-beginning 2009

One can observe that the unweighted curves (Q5_1 and Q5_1_1) were much below their long term average (corresponding to zero on the graph) than the weighted one Q5. This could give an advantage to unweighted series, as it is well known that business tendency surveys did not fully measure the intensity of 2008 crisis. However, we can also notice that the weighted series restarted to increase before the unweighted ones, which is more in line with the evolution of the industrial production. That tends to show that biggest enterprises may have anticipated a recovery before the smallest ones. This last result gives a notable advantage to weighted series.

***Period 2*: end 2009-beginning 2010**

The weighted series reached its long-term average in Q4 2009 before stabilising. At the same time, one could observe the same phenomena in the industrial production series. Once again, the weighted series seem to be more suitable to detect the evolution of production in industry.



This paper never intended to give the definite solution to the question of weighting BS results. But it helped to show that the impact of different weighting schemes on time series is not negligible. From a theoretical point of view, it would have been interesting to develop methods that aim at determining the most efficient weights in order to optimize the forecast of the macroeconomic evolutions. However, we have to keep in mind that we need to find weighting schemes easy to implement and above all, that we have to guarantee the availability of long time series, which could not be the case if new weighting schemes had to be implemented.

5. References

DG ECFIN manual (2007), the joint harmonised EU Programme of business and consumer surveys

OECD manual (2003), Business Tendency Surveys: A Handbook

Biau O., Ferrari N. (2006), Théorie de l'opinion, faut-il pondérer les réponses individuelles?

Fansten M. (1976), Introduction à une théorie mathématique de l'opinion.

Mitchell J., Smith RJ., Weale MR. (2011), Efficient aggregation of panel qualitative survey data.

Shiskin J. (1960), How accurate?