

THE STATISTICAL ACCURACY OF SURVEYS ON BUSINESS AND ECONOMIC PERSPECTIVES: A CASE STUDY

Ramon Alemany, Manuela Alcañiz, Montserrat Guillén

Department of Econometrics, Statistics and Spanish Economy
Riskcenter-IREA, University of Barcelona
Av. Diagonal 690, 08034 Barcelona, Spain
email: ralemany@ub.edu, malcaniz@ub.edu, mguillen@ub.edu

Abstract

In this paper, we develop a new decision making model and apply it in political Surveys of economic climate collect opinions of managers about the short-term future evolution of their business. Interviews are carried out on a regular basis and responses measure optimistic, neutral or pessimistic views about the economic perspectives. We propose a method to evaluate the sampling error of the average opinion derived from a particular type of survey data. Our variance estimate is useful to interpret historical trends and to decide whether changes in the index from one period to another are due to a structural change or whether ups and downs can be attributed to sampling randomness. An illustration using real data from a survey of business managers opinions is discussed.

Keywords: business perspectives, climate surveys, error estimation.

Acknowledgments: We thank the Spanish Ministry of Economy and Knowledge grant number ECO2011-21787-C01-03 and Generalitat de Catalunya AGAUR SGR2009-1328.

1. Introduction

Public opinion surveys have progressively become part of official statistics. Among those, surveys of economic perspectives can be used to measure expectations and they can anticipate the evolution of macroeconomic indicators. Responses to questions in surveys of the economic climate are usually qualitative because they collect information on the beliefs of managers and experts about current and long-term prospects. For instance, a typical question in a survey interview on business climate may ask a manager of a selected company whether

during a given semester sales or new orders increased, decreased or remained the same as in the previous semester. He or she can then be asked if sales will increase, decrease or remain the same in the next subsequent semester.

Data from surveys of economic perspectives allow to calculate the percent of respondents in the total population who have an optimistic, a neutral or a pessimistic view on the status and future trends of their businesses. Data are often presented in the form of an index calculated as the difference between the percent of those who claim that a given business climate trend has improved and of those who claim that it has deteriorated. The higher the index is, the more optimistic is the general opinion. When the index is negative, then it means that the prevailing feeling is pessimistic.

As in any survey, measurement of the sampling error of the results has to be addressed, in order to evaluate the accuracy of the results. Estimating the variance for some key variables in the data set can also be helpful to reconsider the sample size which is required for a desired width of confidence intervals.

When addressing the problem of estimating the variance of indexes of business climate, we encountered the problem of having to estimate the variance of a ratio estimator. Let us be more specific. Opinions given by managers of business climate opinion surveys are usually weighted. This seems natural if one wants to give an overall view. So, in practice more weight is given to respondents from larger firms than to opinions from managers of small firms that participate in the survey. The way these weights are defined complicates the estimation of the variance of the business climate index. For instance, in our application, more weight is given to responses from opinions by managers of firms in sectors with a larger number of employees. Moreover, the number of employees is not known in advance, but it is part of the basic information that has to be provided by the manager participating in the pool. Therefore, his or her opinion will be proportional to the number of employees that work in his or her company sector. Finally, in order to combine different sectors of the economy, there is also a combination of the opinions given by managers in each of the economic sectors that conform the population.

The article is organized as follows. In the next sections we present the notation and general framework to address the estimation of the sampling error of indexes of the economic climate. Then we show the case study and a discussion in the last section concludes.

2. Background and data framework

Our case study is about the Survey of the Business Climate (Encuesta de Clima Empresarial) developed by the Statistical Institute of Catalonia (Institut d'Estadística de Catalunya) and the Chambers of Commerce of Sabadell and Terrassa (Catalonia, Spain) in 2005. The survey aimed at generating official data

on the economic trends in the various districts covered by these chambers. The data were collected every six months to perform the assessment of the future business outlooks and, thus, to evaluate the strength of the economy and the direction of its future change.

The first survey, conducted in January 2004, focused on the second half of 2003 and on the forecasts for the first half of 2004. Compared to the final results for the year 2003, the data for the first half of 2004 suggested an economic expansion. During the second half of 2004, however, a certain moderate slowdown of the economy was recorded. There was a recession in various sectors of industry; the indicators showed negative values, with a decline especially noticeable in textile and tailoring industries. The retailing sector as well as the other service sectors exhibited more stable index values. The construction sector also remained unchanged, but during the first half of 2004 it showed a mild upswing.

During the first half of 2005, the results of the survey indicated a steady overall economic activity with some symptoms of improvement in comparison to the second half of the year 2004. Construction continued to play the role of the economic engine, while the textile industry was in recession. Retailing, after a relatively dynamic period, also experienced a slowdown. Overall, the results for the second half of 2005 confirmed a mild improvement across the economy most clearly observable in retailing, construction, and the service sectors. These sectors helped to maintain the dynamics of the economic activity in related areas, compensating for the persistent stagnation of other branches of industry.

In addition to the evaluation of economic activity, the Survey asked about expectations for the following year in four different areas: a) price trends, b) employment trends, c) the labor force utilization, and d) the expected trends in net revenues. It also addressed the factors that affected the business operating environment. In that regard, the increase in competition, the weakening of the demand, the shortage of the qualified labor force, and the rise of the production costs have been considered the most relevant factors that restricted the business operating environment. In general, the answers to each question offer a choice among one of three characteristics: growth, stagnation, and a decline. As a result, the main concept used in most business climate surveys is the index reflecting the difference between the proportion of optimistic and pessimistic opinions about the changes observed in each of the variables of interest.

Table 1 presents a summary of the industry sectors and the number of companies participating in the Survey. They are members of the Chambers of Commerce of Sabadell and Terrassa.

The weight assigned to each business, π_i , is calculated as the number of employees in the population of each sector divided by the number of employees in the firms that belong to the sample. For example, every employee in the sample sector "Food Industry" represents 6.80 employees of that sector. The assigned weight is the same for all businesses surveyed in a given sector, regardless of their size.

Table 1. Number of Employees in the Population and in the Sample by Sector of Activity

Sector	Employees in Population	Employees in Sample	Factor of Extrapolation
Food Industry	5180	762	6.80
Textile Industry	12987	1642	7.91
Metal, Mining and Transport Industries	43912	8806	4.99
Chemical Industry	12422	1340	9.27
Wood, Paper and Other Industries	13180	792	16.64
Construction	14533	356	40.82
Wholesale and Retail Commerce	34928	1448	24.12
Other Services	34095	2767	12.32
TOTAL	171237	17913	9.56

Chamber of Commerce of Sabadell (2005)

In general, the opinion of managers in business climate surveys is measured by the option chosen among three characteristics: growth, stagnation, and decline. As a result, the main concept used in most business barometer surveys is the index reflecting the difference between the proportion of optimistic and pessimistic opinions about the changes observed in each of the variables of interest, without taking into account the neutral responses.

3. Methodology

Let us consider variable X, which informs about the opinion expressed by a businessperson, we assign (1) when a positive change is reported, (0) when no change is reported, and (-1) when a negative change is reported. One possible topic could be, for example, the “Growth of the Business.” For the n businesses in the sample, the observed responses x_i , $i=1, \dots, n$, are obtained. In order to calculate the opinion index, the following weighted estimator is used:

$$\hat{\theta} = 100 \frac{\sum_{i=1}^n p_i x_i}{\sum_{i=1}^n p_i}$$

which means, that the mean is weighted using factors p_i . That formula is valid for all sectors jointly and for every sector separately.

Since our main objective is to calculate the variance of the opinion index, we intend to continue expressing the index in the form of a simple average. First, the following weights are defined:

$$\omega_i = \frac{p_i}{\sum_{i=1}^n p_i}, \quad i=1, \dots, n, \quad \text{this verifies that } \sum_{i=1}^n \omega_i = 1.$$

Then, the index can be expressed as:

$$\hat{\theta} = 100 \frac{\sum_{i=1}^n p_i x_i}{\sum_{i=1}^n p_i} = 100 \sum_{i=1}^n \omega_i x_i = \frac{100}{n} \sum_{i=1}^n n \omega_i x_i = \frac{1}{n} \sum_{i=1}^n y_i = \bar{Y}$$

where $y_i = 100 n \omega_i x_i$, $i=1, \dots, n$. This means, the index is the sample mean of the variable Y . If we want to adopt a notation where the sector (stratum) is taken into account, then

$$\bar{Y} = \frac{1}{n} \sum_{h=1}^L \sum_{j=1}^{n_h} y_{hj},$$

with n_h the sample size in stratum h , $h=1, \dots, L$, and where y_{hj} has the value of Y in the “ j ” business of the “ h ” stratum.

Various approximations are used in the calculation of the variance. First, a simplified situation is considered, and it is assumed that the sample is a simple random sample. The sample error for the climate index is calculated for every variable of the survey, and an upper bound is computed assuming the maximum possible variance. The sample error will not exceed that bound regardless of the proportion of the optimistic and pessimistic responses to a given question.

Beside the analytical approximations, the calculations are performed by means of re-sampling methods. The advantage of the first method is its simplicity; however it is based on the assumption that the sample weights are fixed and known. Having calculated the sample weights by the number of employees declared in the survey, it is better to consider those factors as estimates based on survey data. Consequently, the opinion index is a more complex estimator, and it suggests the use of the methods of variance estimation for non-linear estimators.

3.1. Approximation of the Variance under the Assumption of a Simple Random Sample

Following the expressions of the simple random sample (Cochran, 1977), the variance estimator of the opinion index is:

$$\hat{V}(\bar{Y}) = \left(1 - \frac{n}{N}\right) \frac{s^2}{n}, \quad \text{with } s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{Y})^2$$

This expression allows us to estimate the standard deviation (standard error) for the index of the variable “Growth of the Business” for all eight sectors jointly (Chamber of Commerce of Sabadell, 2nd half of 2005), and the result is 4.53. As shown in Table 2, the index for that variable is 10.73, the 95% confidence interval for that index (assuming normality) is $\bar{Y} \pm 1.96 \cdot \hat{\sigma}(\bar{Y}) = (1.85; 19.61)$. Given that we have an interval with two positive extremes, we can say that when we assume a simple random sample, the percentage of businesspeople who perceive the change in the “Growth of the Business” as positive is higher than that of businesspeople that have less optimistic expectations.

3.2. Maximum Indetermination under a Simple Random Sample

Next, we calculate the upper bound of the standard error for the previous estimator. Therefore, we keep it in mind that the random variable X may have three possible values: 1, 0 and -1, with probabilities of π_1 , $1-\pi_1-\pi_2$ and π_2 . Its mathematical expectation is equal to $E(X)=\pi_1-\pi_2$ and its variance is $\text{Var}(X)=\pi_1+\pi_2-(\pi_1-\pi_2)^2$. Since π_1 and π_2 are positive and their sum is lower than or equal to one, the variance of X is less than or equal to 1. On the basis of the expression of the index \bar{Y} , we obtain:

$$\text{Var}(\bar{Y}) = \frac{\sum_{i=1}^n p_i^2 \text{Var}(x_i)}{\left(\sum_{i=1}^n p_i\right)^2} \cdot 100^2 \leq \frac{\sum_{i=1}^n p_i^2}{\left(\sum_{i=1}^n p_i\right)^2} \cdot 100^2$$

Therefore,

$$\sigma(\bar{Y}) \leq \frac{\left(\sum_{i=1}^n p_i^2\right)^{1/2}}{\sum_{i=1}^n p_i} \cdot 100$$

Based on data from Table 2, we have $\sum_{i=1}^n p_i = 4958$ and $\sum_{i=1}^n p_i^2 = 97874.20$.

Therefore, $\left(\sum_{i=1}^n p_i^2\right)^{1/2} = 312.8485$ and, consequently,

$$\sigma(\bar{Y}) \leq \frac{312.8485}{4958} 100 = 6.3100$$

This value represents an upper bound for the standard error for any variable in the survey when all sectors are taken together.

The procedure that we have just finished presenting may provide a simple tool for estimating the standard error for specific indices calculated in this Survey. Instead of limiting the variance of the opinion index, the variance can be estimated by using the percentage of businesses that answer “the change is positive” and replace π_1 and, respectively, π_2 can be replaced by the percentage of businesses that answer “the change is negative”. For example, in the case of the index for the variable “Growth of the Business,” we have: $\hat{\pi}_1 = 0.324$ and $\hat{\pi}_2 = 0.216$, and so,

$$\hat{V}(X) = \hat{\pi}_1 + \hat{\pi}_2 - (\hat{\pi}_1 - \hat{\pi}_2)^2 = 0.324 + 0.216 - (0.324 - 0.216)^2 = 0.528$$

Therefore, $\hat{\sigma}(\bar{Y}) = \frac{312.8485}{4958} 100 \cdot \sqrt{0.528} = 4.59$, which is a very straightforward estimate of the standard error.

3.3. Approximation of the Variance Using a Re-sampling Method

In order to obtain the weighted values that are used in the estimation of an index, it is necessary to use the number of units (employees) declared by businesses participating in the survey. Therefore, the statistic cannot be just a weighted average. We also need a more complex estimator, if the sample weights p_i are estimated on the basis of a survey, such as the one we used in our study. If z_i is the number of employees declared by the business “i,” then we know that:

$$p_i = \frac{O_h}{\sum_{j=1}^n v_j z_j}, \text{ where } O_h \text{ is the number (known) of units in the sector (h), to which}$$

the observation “i” belongs. If “j” belongs to the same stratum as “i”, then v_j has the value of 1, and if it does not, then its value is 0.

Using a notation with two sub-indices, we could say that p_{hj} is the weight of the observation “j” of the stratum “h”, $j=1, \dots, n_h$ and $h=1, \dots, L$. If z_{hj} is the number of employees as stated by business “j” in stratum “h”, then:

$$p_{hj} = \frac{O_h}{\sum_{l=1}^{n_h} z_{hl}}.$$

Next, the opinion index can be expressed as a ratio between two estimators

$$\hat{\theta} = \frac{\sum_{i=1}^n p_i x_i}{\sum_{i=1}^n p_i} 100 = \frac{\sum_{i=1}^n 100 p_i x_i}{\sum_{i=1}^n p_i} = \frac{\sum_{h=1}^L \sum_{j=1}^{n_h} 100 p_{hj} x_{hj}}{\sum_{h=1}^L \sum_{j=1}^{n_h} p_{hj}} = \frac{\sum_{h=1}^L \sum_{j=1}^{n_h} 100 \frac{O_h}{\sum_{l=1}^{n_h} z_{hl}} x_{hj}}{\sum_{h=1}^L \sum_{j=1}^{n_h} \frac{O_h}{\sum_{l=1}^{n_h} z_{hl}}} .$$

In order to estimate the variance in this case we will use a re-sampling method that allows us to calculate that estimation by extracting sub-samples and to calculate the same estimator on those sub-samples. We have implemented here the Jackknife method. First, we assumed a simple random sample to calculate an approximation of the variance, and the value of $\hat{\theta}$ has not yet been expressed as a combination of the estimators within the strata. In order to obtain the estimator of θ that excludes an observation, we opt for one that reproduces the procedure of calculation of $\hat{\theta}$ as a complex ratio estimator. When the observation j^* in the stratum h^* is excluded, the estimator $\hat{\theta}_{(h^*, j^*)}$ needs to be calculated on the basis of new sample weights. They exclude the information provided by the business that is not taken into consideration. This means, the factor weight does not change with observations that do not belong to the same stratum (business sector) as the excluded observation. At the same time, all those that belong to the same sector have one observation less and the total number of employees of the sub-sample excludes the ones from that business. In that case, $\hat{\theta}_{(h^*, j^*)}$ is the estimator obtained by means of new sampling weights in the sub-sample that excludes the observation j^* of the stratum h^* . The expression for the leave-one-out statistic is:

$$\hat{\theta}_{(h^*, j^*)} = \frac{\sum_{\substack{h=1 \\ h \neq h^*}}^L \left(\sum_{j=1}^{n_h} 100 \frac{O_h}{\sum_{l=1}^{n_h} z_{hl}} x_{hj} \right) + \sum_{\substack{j=1 \\ j \neq j^*}}^{n_{h^*}} 100 \frac{O_{h^*}}{\sum_{\substack{l=1 \\ l \neq j^*}}^{n_{h^*}} z_{hl}} x_{h^*j}}{\sum_{\substack{h=1 \\ h \neq h^*}}^L \sum_{j=1}^{n_h} \frac{O_h}{\sum_{l=1}^{n_h} z_{hl}} + \sum_{\substack{j=1 \\ j \neq j^*}}^{n_{h^*}} \frac{O_{h^*}}{\sum_{\substack{l=1 \\ l \neq j^*}}^{n_{h^*}} z_{hl}}}$$

Then the estimation of the variance of $\hat{\theta}$ is obtained as (Efron, 1986):

$$\hat{V}_{JK}(\hat{\theta}) = \frac{n-1}{n} \sum_{h^*=1}^L \sum_{j^*=1}^{n_{h^*}} (\hat{\theta}_{(h^*, j^*)} - \hat{\theta})^2$$

We note that the variance is larger than the variance obtained for a simple random sample design.

4. Results

Table 2 shows in detail the calculation of the opinion index for one of the variables in the survey.

Table 2. Obtaining the Index for the Variable “Growth of the Business”

Sector	Sample size	p_i	Freq. $x_i=1$	Freq. $x_i=-1$	$\Sigma p_i x_i$	Σp_i	Σp_i^2	Index
Food Industry	17	6.80	3	2	6.80	115.56	786.08	5.88
Textile Industry	35	7.91	8	15	-55.36	276.82	2189.88	-20.00
Metal, Mining and Transport Industries	92	4.99	28	24	19.95	458.77	2290.81	4.35
Chemical Industry	28	9.27	4	8	-37.08	259.56	2406.12	-14.29
Wood, Paper and Other Industries	26	16.64	9	5	66.57	432.68	7199.13	15.38
Construction	21	40.82	4	4	0.00	857.28	34991.72	0.00
Wholesale and Retail Commerce	58	24.12	31	10	506.55	1399.05	33742.92	36.21
Other Services	94	12.32	23	21	24.64	1158.27	14267.55	2.13
TOTAL	371	-	-	-	532.06	4958.00	97874.20	10.73

Chamber of Commerce of Sabadell (2005)

In Table 3 we show the values for two of the main variables included in the Survey: “Growth of the Business” and “The New Orders”.

Table 3. Estimated index standard errors and CI with different estimation methods

Growth of Business Index=10.7314				
Method	Standard Error	CI 95%	CI 80%	CI 75%
S.r.s.	4.5299	(1.8528; 19.6100)	(4.9261;16.5367)	(5.5204;15.9424)
S.r.s.max. var.	6.3100	(-1.6362; 23.0990)	(2.6448;18.8180)	(3.4727;17.9901)
Re-sampling	4.5305	(1.8516; 19.6112)	(4.9253;16.5375)	(5.5197;15.9431)
New Orders Index=-8.849				
Method	Standard Error	CI 95%	CI 80%	CI 75%
S.r.s.	3.5759	(-15.8586; -1.8410)	(-13.4325;-4.2671)	(-12.9633;-4.7363)
S.r.s max. var.	6.3100	(-21.2174; 3.5178)	(-16.9364;-0.7632)	(-16.1085;-1.5911)
Re-sampling	3.6592	(-16.0218; -1.6778)	(-13.5393;-4.1603)	(-13.0592;-4.6404)

CI (confidence interval), S.r.s. is simple random sample

5. Conclusions

We have obtained results by the estimation of standard errors for indices supplied by the Survey of the Business Climate conducted by the Statistical Institute of Catalonia with the Chambers of Commerce of Sabadell and Terrassa. The estimation presented here is both analytical and empirical, and re-sampling methods were used together with the data from the above mentioned business survey. Moreover, a methodology was presented for the approximation of the calculation of standard errors in situations of maximum indetermination of the answer.

In this work, we claim that an error can occur if we use the traditional method of calculation of the variance for a proportion based on the assumption of maximum indetermination in Surveys of the Economic Climate. When we focus our interest on the estimation of opinion indices that calculate the difference between positive and negative opinion proportions, there are three factors that are not accounted for: a) the variable of interest is not a proportion but it can be expressed as a weighted average, b) the sample is neither a simple nor a stratified random sample, and c) sample weights are used. It has been demonstrated that the approximation based on the traditional formula that assumes maximum indetermination in the estimation of the variance of a proportion, may imply that the confidence intervals for the opinion indices become too narrow. Moreover, (in case of a simple random sample and with the sample weights used in the survey of the Chamber of Commerce of Sabadell) we have focused on the relation that exists between the standard error of an index, its value and the proportion of businesses that have neither expressed optimistic nor pessimistic opinions (that is, expressed a neutral position). We have concluded that the highest standard error can be found in those lowest indices and that when the proportion of those indifferent responses increases, the standard error decreases.

References

- Cochran, W.G. (1977); *Sampling Techniques*. Wiley. New York.
- Efron, B. (1986); *The Jackknife, the Bootstrap and Other Resampling Plans*. CBMS-NSF. Regional Conference Series in Applied Mathematics, Philadelphia.
- Ifo Institute for Economic Research (2011); *Ifo Business Climate Germany*, Munich [<http://www.ifo-business-climate-index.info>], Consulted 15.4.2011
- Lohr, S.L. (2000); *Muestreo: diseño y análisis*. International Thomson Editores. México.
- Sinn, H.W., Abberger, K. (2006); *The forecasting quality of the Ifo Index* [http://www.cesifo.de/link/ifosd_2006_4_6.pdf], Consulted 15.4.2011]