CALCULATION OF THE VARIANCE IN SURVEYS OF THE ECONOMIC CLIMATE

By Manuela Alcañiz**§**, Àlex Costa**†**, Montserrat Guillén**‡**, Carme Luna**£** and Cristina Rovira**¥**

§ Grup de Risc en Finances i Assegurances (RFA)-IREA, Universitat de Barcelona, Dept. d'Econometria, Estadística i Economia Espanyola, Av. Diagonal 690, E-08034 Barcelona, Spain. Tel. +34+934021983 / Fax: +34+934021821 / E-mail: malcaniz@ub.edu.
† Institut d'Estadística de Catalunya, Via Laietana 58, E-08003 Barcelona, Spain. Tel. +34+934120088 / Fax: +34+9340123145 / E-mail: acosta@idescat.net.
‡ Grup de Risc en Finances i Assegurances (RFA)-IREA, Universitat de Barcelona, Dept. d'Econometria, Estadística i Economia Espanyola, Av. Diagonal 690, E-08034 Barcelona, Spain. Tel. +34+934037039 / Fax: +34+934021821 / E-mail: mguillen@ub.edu.
£ Institut d'Estadística de Catalunya, Via Laietana 58, E-08003 Barcelona, Spain. Tel. +34+934120088 / Fax:+34+934123145 / E-mail: cluna@idescat.net.
¥ Institut d'Estadística de Catalunya, Via Laietana 58, E-08003 Barcelona, Spain. Tel. +34+934120088 / Fax:+34+934123145 / E-mail: crovira@idescat.net.

Abstract:
Public opinion surveys have become progressively incorporated into systems of official statistics. Surveys of the economic climate are usually qualitative because they collect opinions of businesspeople and/or experts about the long-term indicators described by a number of variables. In such cases the responses are expressed in ordinal numbers, that is, the respondents verbally report, for example, whether during a given trimester the sales or the new orders have increased, decreased or remained the same as in the previous trimester. These data allow to calculate the percent of respondents in the total population (results are extrapolated), who select every one of the three options. Data are often presented in the form of an index calculated as the difference between the percent of those who claim that a given variable has improved in value and of those who claim that it has deteriorated. As in any survey conducted on a sample the question of the measurement of the sample error of the results has to be addressed, since the error influences both the reliability of the results and the calculation of the sample size adequate for a desired confidence interval. The results presented here are based on data from the Survey of the Business Climate (Encuesta de Clima Empresarial) developed through the collaboration of the Statistical Institute of Catalonia (Institut d'Estadística de Catalunya) with the Chambers of Commerce (Cámaras de Comercio) of Sabadell and Terrassa.

Keywords: Economic climate, variances, sampling methods.

1. Introduction

Surveys of public opinion among citizens, whether consumers or businesspeople, have been slowly but progressively incorporated into systems of official statistics. In Spain, the Barometer of the Sociological Research Center (Barómetro del Centro de Investigaciones Sociológicas CIS) has been implemented since the 1970s, however, it has not been part of official statistics until the latest National Statistical Plan (Plan Estadístico Nacional) (2005-2008). In spite of that delay, there is no doubt that opinion pools are important, and the Eurobarometer is a good example of that importance.

The economic statistics of Catalonia contain a significant number of data based on opinion. In fact, we find a number of surveys that rely on individual assessments: the Industrial Climate (Clima Industrial), Export Climate (Clima Exportador), Commercial Urban Climate (Clima Comercial Urbano) and the Economic Climate (Clima Económico) of the Chambers of Commerce of Sabadell and Terrassa. The first two examples are similar to two surveys, in the Spanish statistical system, called the Survey of Industrial Activity and the Survey of Export Activity. They are conducted by the Ministry of Industry, Tourism and Commerce (Ministerio de Industria, Turismo y Comercio). However, the surveys of the Commercial Climate (Clima Comercial) and the Climate of the Chambers of Commerce are unique to the Catalan system of official statistics. With regard to commerce, the Catalan government shows great interest in urban commercial activities in small areas. We believe that the impact which the new Statistical Laws of Catalonia had on the Chambers of Commerce is without precedence in Spain: the Chambers of Commerce can conduct their own official surveys, thus, obeying all the necessary requirements and guarantees. More importantly, these official surveys are becoming obligatory for businesses, and therefore the statistical confidentiality, the principles of impartiality, and the proper quality of official statistics are fully guaranteed.

From the point of view of the methodology and the objectives, it matters whether a survey is official or not. If a business survey is not official, two things happen. First, the response is voluntary, and therefore it is not random. So, this complicates the calculation of variances or the measurement of the confidence level of the conclusions. Secondly, from the point of view of official statistics, business opinion is not only interesting by itself, but as an indirect measure, as a proxy variable of an official objective indicator. These two characteristics cause that the surveys are often viewed as providing instrumental information and what is interesting about them is their capacity to forecast the change in the official indicators. A very good example is the survey of the Industrial Climate (Clima Industrial - ICI) that is considered an approximation of the Index of Industrial Production (Indice de Producción Industrial - IPI).

When the business opinion surveys are becoming official, some important changes occur in the methodology and the objectives, but the predictive power of people's opinions, as compared to more classical indicators used in official statistics, is not compromised. The most important change is that the surveys are becoming obligatory and, therefore, can be conducted on random samples. This fact allows us to use standard sampling techniques to calculate the variances and the error. The second change is that today knowledge of business opinions is by itself an objective of the official system of statistics. Therefore, it is beneficial and interesting to estimate the population value of those opinions as well. For example, this means that instead of being interested only in the predictive power of the ICI in comparison to the IPI, we are also interested in measuring the extent to which our ICI estimate is similar to the population ICI. We believe that the perspective which we are introducing in Catalonia is truly new. For example, when we contacted the authors of the German business climate survey, who use the survey of business opinions that is the best known in Europe, we were

informed that they do not calculate the standard error because their sample is voluntary and therefore, it is not random.[1]

The survey of the Business Climate conducted by the Statistical Institute of Catalonia with the Chambers of Commerce of Sabadell and Terrassa aims at generating official data on the economic trends in the various districts covered by these chambers, and the Catalan district of Valles Occidental. The data are collected every six months to perform the assessment of the current economic situation and of future business outlooks and, thus, to evaluate the strength of the economy and the direction of its future change. The surveys of the economic situation of the Valles Occidental were initiated in January of 2004, by an agreement about collaboration between the Chambers of Commerce of Sabadell and Terrassa and the Statistical Institute of Catalonia. The survey is repeated every six months. 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 slow-down 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, without taking into account the neutral opinions:

Index = % of positive opinions - % of negative opinions

There are a number of organizations in Spain, both public and private, conducting surveys similar to the one we used for our analysis. One example is the survey conducted during the Convention on Collaboration between the Board of Castilla and Leon and the Confederation of the Business Organizations of Castilla and Leon (CECALE). This qualitative survey conducted every trimester, attempts to gather data from the three key sectors constituting the economic base of Castilla and Leon. The survey conducted by the General Office of Statistics (Dirección General de Estadística) gathers the opinions of businesspeople about the current economic climate and trends in key economic indicators such as new orders, inventory levels, production, prices, and the employment environment and investment. Another similar survey is the Survey of the Business Climate (Encuesta de

Clima Empresarial) conducted in order to obtain the barometer of the Urban Economy for the Government Economy Area and Citizen Participation in the Town Council of Madrid (Area de Gobierno de Economía y Participación Ciudadana del Ayuntamiento de Madrid). This is a survey of the 500 largest companies in the City of Madrid and it also provides the Confidence Indicator of Business (Indicator de Confianza Empresarial). In a similar way, the Business Barometer of Andalusia (Barómetro Empresarial de Andalucía) makes it possible to use the indicators to describe business people's perception of the actual economic situation and their short-term expectations. One of the concepts used in the Andalusian survey, that is different from the index analyzed by the majority of comparable surveys, is the *net* index, which reflects the difference between the optimistic and the pessimistic positions in the following way:

$$\text{Net Index} = \frac{\% \text{ positive responses} - \% \text{ negative responses}}{\% \text{ positive responses} + \% \text{ negative responses}} \cdot 100$$

The sign indicates the actual opinion of businesspeople from a given sector or sub-sector. In this case, the index may vary between +100 (fully optimistic situation) and -100 (pessimistic situation). There are also private agencies that conduct surveys by evaluating the business forecasts in their own area. For example, Caja Segovia (2006), conducted a survey of businesses located in the capital of Segovia (60 respondents) and some cities of the province (50 respondents), using a stratified random sample.

Business climate surveys are also carried out on an international level. Perhaps the best known among them is the one conducted in Germany to calculate the so called *Ifo Business Climate Index*. In that research 7,000 businesses operating in all sectors of the economy are surveyed on a monthly basis. The companies are asked to give their assessment of the current business situation and their expectations for the next six months. They can characterize their situation as "good", "satisfactory" or "poor," and their expectations for the

next six months as "more favorable", "unchanged" or "more unfavorable". The "*Ifo*" index is a transformed mean of the balances of the business situation and the expectations. In order to calculate the index values, the two transformed balances are normalized to the average of the year 2000.[2]

2. Obtaining the Data

The Survey of the Economic Climate is conducted every six months by the Statistical Institute of Catalonia and the Chambers of Commerce of Sabadell and Terrassa employing the computer-assisted telephone interviewing (CATI) survey. In the initial phase, the information about the launching of the survey is communicated to the intended respondents by mail, together with the explanation of the survey and a sample questionnaire. After this initial period, the field work begins, and the telephone interviews are conducted by an external agency contracted to perform the service. The agency is also responsible for adjusting the sample, in case it is necessary to correct it (e.g. a firm has discontinued its operation or changed location). Once the results (which are checked for reliability) are obtained, they are extrapolated and analyzed. The results are first stratified by industry sector and adjusted for the size of each sector in terms of the number of businesses represented in the population of each sector. Below is a summary of the industry sectors to which the participating companies belong. They are members of the Chambers of Commerce of Sabadell and Terrassa.

Groups Surveyed	CNAE-93[3]
C01. Food Industry	15-16
C02. Textile Industry	17-19
C03. Metal, Mining and Transport Industries	10-14, 23, 26-35, 40-41

---

[2] www.cesifa-group.de/portal/page?_pageid=36,1899103&_dad=portal&_schema=PORTAL

[3] Clasificación Nacional de Actividades Económicas (1993) (National Classification of Economic Activities). This classification segments companies according to their primary industrial activity.

C04. Chemical Industry	24-25
C05. Wood, Paper and Other Industries	20-22, 36-37
C06. Construction	45
C07. Wholesale and Retail Commerce	50-52
C08. Other Services	55, 60-67, 70-74, 80, 85, 90-93

For the present study, we used a sample of businesses from the Chamber of Commerce of Sabadell directory. We used a sample stratified into the eight sectors as shown above. The sample comprised 371 firms and the data referred to the second half of the year 2005. Table 1 shows the number of employees in the general population and in the sample, stratified by sectors of business activity. The data for the population come from the directory of the Chamber of Commerce of Sabadell, while the sample estimates have been calculated on the basis of answers provided by the companies interviewed in each sector. In order to obtain the total number of employees for the firms in the sample, per sector, the answers obtained from the businesses surveyed in each sector are aggregated. These data vary every six months due to the difference in the reported number of employees.

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
C01	5180	762	6,80
C02	12987	1642	7,91
C03	43912	8806	4,99
C04	12422	1340	9,27
C05	13180	792	16,64
C06	14533	356	40,82
C07	34928	1448	24,12
C08	34095	2767	12,32
TOTAL	171237	17913	9,56

Chamber of Commerce of Sabadell,
2nd half of 2005

The weight assigned to each business, p_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 C01 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.

3. Methodology for Obtaining the Data

Let's consider variable X, on the basis of an 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 variable could be, for example, the "Growth of the Business." For the n businesses in the sample, the observated x_i, i=1,…,n, are obtained. In order to calculate the weighted index, the following 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. Table 2 shows in detail the calculation of the index, for the eight sectors.

Table 2. Obtaining of the Index for the Variable "Growth of the Business"

Sector	n	p_i	$x_i=1$	$x_i=-1$	Σx_i	$p_i\Sigma x_i = \Sigma p_i x_i$	$np_i = \Sigma p_i$	Σp_i^2	Index
C01	17	6.80	3	2	1	6.80	115.56	786.08	**5.88**
C02	35	7.91	8	15	-7	-55.36	276.82	2189.88	**-20.00**
C03	92	4.99	28	24	4	19.95	458.77	2290.81	**4.35**
C04	28	9.27	4	8	-4	-37.08	259.56	2406.12	**-14.29**
C05	26	16.64	9	5	4	66.57	432.68	7199.13	**15.38**
C06	21	40.82	4	4	0	0.00	857.28	34991.72	**0.00**
C07	58	24.12	31	10	21	506.55	1399.05	33742.92	**36.21**
C08	94	12.32	23	21	2	24.64	1158.27	14267.55	**2.13**
TOTAL	371	-	-	-	-	532.06	4958.00	97874.20	**10.73**

Chamber of Commerce of Sabadell,
2nd half of 2005

Since our main objective is to calculate the over-time variances relative to the previous levels of the index, we intend to continue expressing the index in the form of a simple average. First, the following weights are defined:

$$\omega_i = \frac{p_i}{\sum\limits_{i=1}^{n} p_i} \, , \text{ i=1,...,n, this verifies that } \sum\limits_{i=1}^{n} \omega_i = 1.$$

Then, the index can be expressed as:

$$\hat{\theta} = 100 \frac{\sum\limits_{i=1}^{n} p_i x_i}{\sum\limits_{i=1}^{n} p_i} = 100 \sum\limits_{i=1}^{n} \omega_i x_i = \frac{100}{n} \sum\limits_{i=1}^{n} n \omega_i x_i = \frac{1}{n} \sum\limits_{i=1}^{n} y_i = \overline{Y}$$

where $y_i = 100 \, n \, \omega_i \, x_i$, i=1,...,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

$$\overline{Y} = \frac{1}{n} \sum\limits_{h=1}^{L} \sum\limits_{j=1}^{n_h} y_{hj} \text{ , with } n_h = \text{ size of sample in stratum h, h=1,...,L,}$$

when y_{hj} has the value of Y in the "j" business of the "h" stratum.

4. Estimation of the Index Variance

Once it is observed that the opinion index can be expressed as a sample mean, it is necessary to estimate its variance. In this way, it is possible to find out the sample error of the results, to determine the confidence interval, and to calculate the size of the sample necessary to obtain a required precision. 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. Later on, the same calculations are performed, assuming that the sample is a stratified random sample.

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.

4.1. 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:

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

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. Since, as seen 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 to deal with 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.

4.2. 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 $Var(X) = \pi_1 + \pi_2 - (\pi_1 - \pi_2)^2$. 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:

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

Therefore,

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

Based on data from Table 2, we have:

$$\sum_{i=1}^{n} p_i = 4958 \quad y \quad \sum_{i=1}^{n} p_i^2 = 97874.20$$

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

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

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

4.3. Approximation of the Variance under a Stratified Random Sample

Let N be the number of businesses in the population, and N_h the number of the businesses in a stratum (sector) h, h=1,..., L. Then the variable $\overline{Y}_h^*$ is the mean of the observations in the sample of that stratum h:

$$\overline{Y}_h^* = \frac{1}{n_h}\sum_{j=1}^{n_h} y_{hj}^*$$

and, the estimation of the mean for the population, based on a stratified sample design, is:

$$\overline{Y}_{st} = \frac{1}{N}\sum_{h=1}^{L} N_h \overline{Y}_h^*$$

The expression of the variance is known (Cochran, 1977) and it is expressed as:

$$Var(\overline{Y}_{st}) = \sum_{h=1}^{L} W_h^2 \frac{S_h^2}{n_h}(1-f_h)$$

where $W_h = \frac{N_h}{N}$, n_h is the size of the sample of a stratum h, $f_h = \frac{n_h}{N_h}$ is the fraction of the sample in the stratum h, and S_h^2 is the population variance in the stratum h, h=1,...,L. Since this last variance S_h^2 is unknown, it is necessary to use an estimator of the previous expression, so the variance estimator becomes:

$$\hat{V}(\overline{Y}_{st}) = \frac{1}{N^2}\sum_{h=1}^{L} N_h(N_h - n_h)\frac{\hat{S}_h^2}{n_h}$$

where $\hat{S}_h^2 = \frac{1}{n_h - 1}\sum_{j=1}^{n_h}(y_{hj}^* - \overline{Y}_h^*)^2$.

Taking $y_{hj}^* = 100\frac{Nn_h}{N_h}\omega_{hj}x_{hj}$, with $\omega_{hj} = \dfrac{p_{hj}}{\sum_{h=1}^{L}\sum_{j=1}^{n_h}p_{hj}} = \dfrac{p_{hj}}{\sum_{h=1}^{L}n_h p_{hj}}$, we show that:

$$\overline{Y}_{st} = \overline{Y} = \frac{1}{n}\sum_{h=1}^{L}\sum_{j=1}^{n_h} y_{hj}$$

And it is possible to use the expression of $\hat{V}(\overline{Y}_{st})$ to calculate the variance of the opinion index $\overline{Y}$.

On the basis of this methodology, the standard deviation (standard error) of the index has been estimated for the variable "Growth of the Business," for all eight sectors jointly. The result of 4.4107 is slightly lower than under the assumption of a simple random sample. In this case, the 95% confidence interval for the index is:

$$\overline{Y} \pm 1.96 \cdot \hat{\sigma}(\overline{Y}) = (2.09; 19.37)$$

It is observed again that the difference between the percentage of businesspeople who have an optimistic vision and those who express a more pessimistic perspective regarding the growth of the business has a positive value.

If a stratified random sample is used instead of a simple random sample, the variance of the mean estimator is always lower or equal. However, in our case the gain is small. This is caused by the fact that the variance within the strata (sectors of activity) is large, which means that in the same stratum we find businesses sufficiently heterogeneous. This lack of homogeneity within a stratum lowers the precision of the estimations that are obtained by means of a stratified random sample.

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

In the previous section we have assumed that the calculation of the sample weights is not characterized by randomness. Therefore, in the case of a stratified sample, the approximation of the variance in the survey could be performed with the assumption that the statistic of interest is a weighted average. However, 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:

$p_i = \dfrac{O_h}{\sum\limits_{j=1}^{n} \upsilon_j z_i}$, where O_h 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 υ_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,…,n_h and h=1,…,L. If z_{hj} is the number of

employees as stated by business "j" in stratum "h", then:

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

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

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

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 $\hat{\theta}$ 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. In order to perform the estimation of the variance with a re-sampling method, we

should remember that the statistic has been obtained by means of a stratified sample, and that

we need to express it as a combination of indices in separate strata. At the same time, we

need to keep it in mind that the sample weights are estimated in the same survey, and this

complicates excessively the procedure. To summarize, in order to compare the standard

errors obtained by means of different methods, in Table 3 and 4 we show the values for two

of the main variables included in the Survey: "Growth of the Business" and "The New

Orders".

Table 3. Index estimated for the variable "Growth of the Business,"
standard errors and confidence intervals
following different estimation methods

Index=10.7314

Sample	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)
Max. indet.	6.3100	(-1.6362; 23.0990)	(2.6448;18.8180)	(3.4727;17.9901)
Str.r.s..	4.4108	(2.0863; 19.3765)	(5.0787;16.3841)	(5.6574;15.8054)
Re-sampling	4.5305	(1.8516; 19.6112)	(4.9253;16.5375)	(5.5197;15.9431)

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

Table 4. Index estimated for the variable "New Orders,"
standard errors and confidence intervals
following different estimation methods

Index=-8.8498

Sample	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)
Max. indet.	6.3100	(-21.2174; 3.5178)	(-16.9364;-0.7632)	(-16.1085;-1.5911)
Str.r.s..	3.5693	(-15.8457; -1.8540)	(-13.4240;-4.2756)	(-12.9557;-4.7439)
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, Str. r. s. is stratified random sampling

4.5. Approximation to the direct estimation of the index variance and its relation to the variance of a proportion

It is extremely common that in surveys with mainly qualitative questions, the reliability of the answers is evaluated using the standard error of a proportion. It is based on the assumption of maximum indetermination (that is, it is commonly expressed as $p = q = 0,5$), and of the infinite population. In the case of the Survey of the Business Climate, since the number of participating businesses is 371 (and without taking into consideration that sample weights are used), we get an estimator of the variance equal to:

$$\hat{V}(\hat{P}) = \frac{0,25}{371},$$

and this implies a standard error of 2.6%. This value is an underestimate of the actual standard error of an opinion index, because the index is a difference between two estimated proportions. If we want to get a better approximation of the standard error by a simple and direct method, an intuitive and straightforward formula to use is based on the fact that Var

(X)=π_1+π_2−(π_1-π_2)^2 . It requires the assumption of a simple random sample of an infinite

population, where π_1 indicates the proportion of individuals that responded with positive

expectations and π_2 the proportion of individuals who responded with negative expectations.

If $\hat{\pi}_1$ is the proportion estimated for the sample of businesses that gave optimistic

responses, $\hat{\pi}_2$ is the proportion estimated for the sample of businesses that gave pessimistic

responses, $\hat{\theta} = \hat{\pi}_1 - \hat{\pi}_2$ is the estimation of the index, and $\hat{\pi}_3 = 1 - \hat{\pi}_1 - \hat{\pi}_2$ is the proportion

estimated of the indifferent responses, then the standard error of the index is:
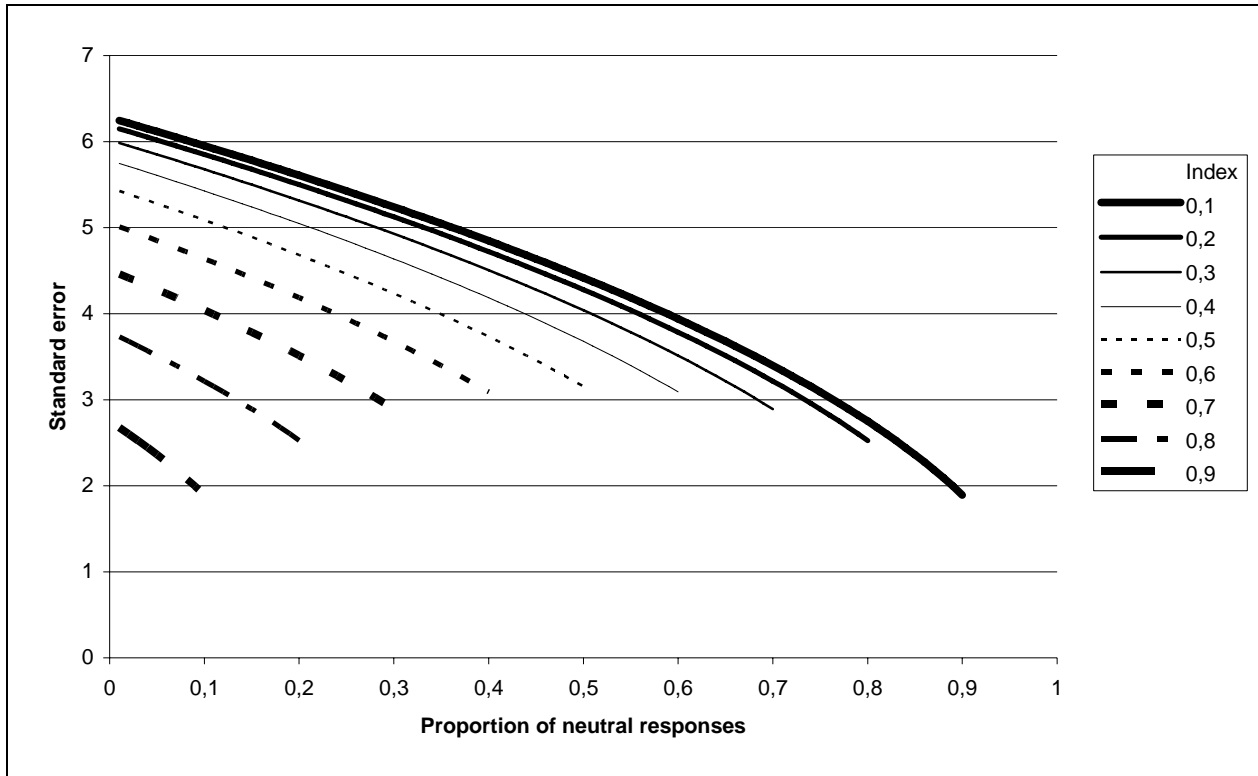
$$\hat{\sigma}(\hat{\theta}) = \frac{\left(\sum_{i=1}^{n} p_i^2\right)^{1/2}}{\sum_{i=1}^{n} p_i} (1 - \hat{\pi}_3 - \hat{\theta}^2)^{1/2} \cdot 100$$

and p_i is the sampling weight of the observation "i". The approximation used here is similar

to the one explained in section 4.2.

This method of approximating the standard error makes it easy to interpret what

occurs for various indices. For example, if the index increases in its absolute value, then the

variance decreases. However, the proportion of businesses that do not give optimistic or

pessimistic responses (that is, those that do not expect any changes) has an impact on the

calculation of the variance. If the proportion of businesses that express a neutral position

increases, then the variance decreases and, therefore, the standard error of the index also

decreases. Obviously, there is a relationship between the estimated index (the difference

between the optimists and pessimists) and the neutral responses because these three

proportions add to the value of one. In Graph 1, the estimated standard error is illustrated by

various values of the estimated index. It has been calculated from the proportion of

businesses that maintain a neutral position.

Graph 1. Approximate standard error for different index values
estimated (in absolute value)

and of the proportion of neutral responses by businesses

When we look carefully at the graph, we notice that if the index is lower than 30% (which is

the most common case), the error corresponds to the curves to the right. If the proportion of

businesses that situate themselves in the neutral zone does not reach 70%, the standard error

that is obtained has a value higher than 2.6%. Moreover, it can be observed on the graph that

for the specific combinations of the index and the percentages of businesses that express a

neutral opinion, the confidence intervals obtained by the traditional approximation of the

standard error of a proportion, have an underestimated value, in excess of 50%.

5. Conclusions and the research program

The presentation of the results of a survey of the economic climate often uses the

difference between the percentage of those who believe that a variable has improved and

those who believe it became worse. In order to obtain an estimation of the variance of such

indices, it is possible to construct confidence intervals both to determine their significance and to calculate the size of the sample adequate for the desired confidence level in subsequent surveys.

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.

In relation to the plan of work, several issues need to be further studied:

1) The explanation of the temporal index stability. We would like to extend over time our analysis to other observations, since the surveys provide us with the most accurate dynamic data profile. In this regard, it is interesting to study the reason why at times that profile is stable, in spite of the relatively large variances of the index. One possible reason is the sample selection method.

2) The determination of the sample size in Climate Surveys. There is no doubt that when the response is compulsory, the samples need to be calculated with great care to make them as small as possible and, thus, to reduce the burden imposed on businesses. The calculation of variances allows us to optimize the sample design (sample allocation) and to determine the minimal sizes required for the desired confidence level. In the Catalan system of statistics this is particularly relevant because in the near future all the business opinion surveys conducted by Chambers of Commerce may become official and therefore, obligatory.

3) The prediction of other objective indicators using business people's opinions. The topic of the relation between opinions of businesspeople and the objective indicators should not be excluded from our range of work. In Catalonia there are at least three surveys that have an objective variable: the index of retail commerce, the index of export and the IPI.

4) The estimation of small areas with and without auxiliary information. One of the domains providing great opportunities for regional official statistics is the calculation of estimations for small areas. The geographic proximity of regional statistical offices evokes our strong interest in the new techniques. In Catalonia there are thirteen Chambers of Commerce. The Council of all Chambers of Commerce conducts one general survey for the entire region of Catalonia, but it does not really capture its thirteen geographic areas. Moreover, some

Chambers (as was illustrated by the example of Sabadell and Terrassa) conduct additional surveys. Therefore, we are able to calculate general estimators for Catalonia and to apply them to small areas with or without auxiliary information, regardless of whether or not every Chamber of Commerce conducts an additional survey.

Acknowledgements

References

CAJA SEGOVIA (2006); Encuesta de Expectativas Empresariales. 1º trimester de 2006. Observatorio Socioeconómico de Segovia.

[http://www.cajasegovia.es/observatorio/observatorio.asp, Consultado:15.7.2006]

COCHRAN, W.G. (1977); *Sampling Techniques.* Wiley. New York.

COSTA, A., GALTER, J., et al (1998); *El clima exportador: un indicador avançat de les exportacions de l'economia catalana.* Nota d'Economia 60. Departament d'Economia i Finances. Generalitat de Catalunya.

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 (2006); Ifo Business Climate Germany, Munich [http://www.ifo-business-climate-index.info], Consultado 15.7.2006

INSTITUTO DE ESTADÍSTICA DE ANDALUCÍA (2006); Barómetro Empresarial de Andalucía. Consejería de Economía y Hacienda.

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, Consultado 15.7.2006]