A Spanish continuous voluntary web survey: sample bias, weighting and efficiency

de Pedraza, P.; Tijdens, K.; Muñoz de Bustillo, R.; Steinmetz, S.

Publication date
2010

Document Version
Final published version

Published in
Revista española de investigaciones sociológicas

Revista Española de Investigaciones Sociológicas (Reis)
N.º 131, 2010, pp. 109-130

A Spanish Continuous Volunteer Web Survey: Sample Bias, Weighting and Efficiency*

Una encuesta voluntaria y continua en la red en España: sesgo, ponderación y eficiencia

Pablo de Pedraza
University of Salamanca
pablodepedraza@usal.es
Kea Tijdens
University of Amsterdam
k.g.tijdens@uva.nl
Rafael Muñoz de Bustillo
University of Salamanca
bustillo@usal.es
Stephanie Steinmetz
University of Amsterdam
steinmetz@fsw.eur.nl

Keywords: Online Surveys, Fieldwork, Sampling, Response Rates, Labour Market.

Palabras clave: Encuestas Online, Trabajo de Campo, Muestreo, Tasas de Respuesta, Mercado de Trabajo.

ABSTRACT

Using micro data from a continuous volunteer web survey (CVWS), the *WageIndicator* (WI), this paper firstly analyses the type of bias that such a survey method produces. Secondly, following a «model-based» approach, two alternative data weighting methodologies are implemented. Thirdly, in order to test whether weighting corrects the bias, thus making it possible to obtain conclusions applicable to the whole labour force, the efficiency of the weighting

RESUMEN

Utilizando los datos obtenidos mediante una encuesta continua en Internet (CVWS), *tusalario.es* (WI), que tiene como población objetivo la población activa española, el trabajo analiza, en primer lugar, el sesgo que dicho método de encuesta produce. En segundo lugar, aplica dos metodologías de ponderación con la intención de corregir los errores de representatividad. En tercer lugar, para saber si la ponderación de los datos corrige el sesgo y hace posible la obtención de

* Acknowledgements: this paper is based on research conducted as part of the WOLIWEB (WOrk LIfe WEB) project, funded by the European Community under the 6th Framework Programme (FP6-2004-506590). For more information see *http://www.wageindicator.org*.

methodologies is evaluated. Since the *WageIndicator* is a labour market oriented survey, weighting efficiency is evaluated by calculating mean salaries, inequality indices and salary regressions before and after applying weights to WI data, and by comparing the results obtained with those achieved using the Structure of Earnings Survey (SES), a wage survey run by the Spanish National Statistics Institute. It is concluded that, after weighting, estimated statistics and parameters move in the right direction.

resultados y conclusiones aplicables al conjunto de la población activa, la eficiencia de ambas metodologías de ponderación es evaluada. La estrategia de evaluación consiste en comparar salarios medios, índices de desigualdad y regresiones salariales obtenidos antes y después de ponderar los datos del WI con aquellos que pueden obtenerse utilizando los datos de la Encuesta de Estructura Salarial (SES), que lleva a cabo el Instituto Nacional de Estadística. Los estadísticos y parámetros estimados se mueven en la dirección correcta después de ponderar.

Pablo de Pedraza
PhD in Applied Economics and BA in Law and Economics from the University of Salamanca. He is currently a non-tenured Assistant Lecturer at the same university.

Doctor en Economía Aplicada y Licenciado en Derecho y Economía por la Universidad de Salamanca. En la actualidad es Profesor Ayudante Doctor en dicha Universidad.

Kea Tijdens
PhD in Sociology from the State University of Groningen. She is currently a research coordinator at the Amsterdam Institute of Labour Studies and a Professor in the Sociology Department at the University of Rotterdam.

Doctora en Sociología por la State University of Groningen. En la actualidad es coordinadora de investigación del Amsterdam Institute of Labor Studies y Profesora en el Departamento de Sociología de la Universidad de Rótterdam.

Rafael Muñoz de Bustillo
PhD in Economics from the Complutense University of Madrid. He is currently a Professor in the Department of Applied Economics at the University of Salamanca.

Doctor en Economía por la Universidad Complutense de Madrid. En la actualidad es Catedrático del Departamento de Economía Aplicada de la Universidad de Salamanca

Stephanie Steinmetz
PhD in Sociology from the University of Rotterdam. She is currently a postdoctoral researcher at the same University.

Doctora en Sociología por la Universidad de Rótterdam. En la actualidad es investigadora postdoctoral en la misma Universidad.

Departamento de Economía Aplicada. Facultad de Derecho (hab. 137). Universidad de Salamanca. Campus Miguel de Unamuno. 37007 Salamanca (Spain).

1. INTRODUCTION: STATE OF THE ART OF THE WEB SURVEYS METHODOLOGY

Using data from the Spanish WageIndicator survey (WI, see *www.tusalario.es*), a Continuous Voluntary Web Survey (CVWS) whose target population is the Spanish labour force, this paper aims to test whether weighting data from a CVWS enables this type of data to be used to obtain conclusions applicable to the whole target population. The paper uses two methods to calculate weights: post-stratification (PS) and propensity score adjustment (PSA). The former method is commonly used to weight traditional survey data; the latter has often been used to weight data from mixed-mode-probability based surveys in which the web was one of the modes used. Both methods have been adapted to be applied to WI data.

The strengths and weaknesses of web surveys have been discussed extensively in the literature (Bronner and Kuijlen, 2007; Coderre et al., 2004; Fricker and Schonlau, 2002; Ilieva et al., 2002; Pratesi et al., 2004; Roster et al., 2004; Schneider et al., 2005; Tingling et al., 2003; Tuten et al., 2002). There remain fundamental concerns related to various sources of survey error, particularly regarding the coverage error, the error associated with the inference from non-probability samples, the non-response error and the measurement error.

At present, the coverage error is a serious problem for many web surveys, particularly those targeting the general population. As not every person has Internet access, and as a list of e-mail addresses covering the whole population does not exist, not everyone has the same probability of being included in the survey. Moreover, though Internet penetration rates continue to increase, the potential bias is related not only to the number of people who have access to the Internet, but also to the differences among them in age, gender, education, and behavioural characteristics (Bandilla et al., 2003; Couper et al., 2007; Dever et al., 2008).

The absence of an e-mail list for the whole population implies the absence of an adequate sampling frame (Couper, 2000). Particularly, problems arise when adopting non-probability and self-selection recruitment methods. In these cases respondents form a convenience rather than a probability sample and little is known about the degree to which the results obtained can be generalised for the whole population. Furthermore, people who self-select into a survey may differ from those who do not in terms of time availability, web skills, or altruism to contribute to the project (Bandilla et al., 2009; Fricker, 2008; Malhotra and Krosnick, 2007).

Coverage and sampling is less of a problem where all members of the target population use the Internet and when e-mail addresses are known, such as students, employees, members of organizations, customers, etcetera. Here, the existence of a proper sampling frame allows the drawing of a probability-based sample, and the conclusions can be generalised to the whole population at stake by using standard inference procedures. However, even once a (probability) sample of potential respondents has been selected, the methodological concerns continue, because not all sample members will be willing or able to complete the survey. The extent of bias depends on the rate of non-response as well as on differences between respondents and non-respondents in the variables of interest. Non-response bias is not unique to web surveys but as their response rates tend to be lower when compared to other modes (Lozar Manfreda et al., 2008; Lynn, 2008; Kaplowitz et al., 2004; Shih and Fan, 2008), the problem is quite severe (but clearly defined) (Pratesi et al., 2004). Different reasons, such as inefficiency of response stimulating efforts (incentives, follow-up contacts), technical difficulties (slow or unreliable connections, low-end browsers), personal problems in using a computer, and privacy and confidentiality concerns could be responsible (Bosnjak and Tuten, 2003; Galesic, 2006; Göritz, 2006; Heerwegh, 2005; Kaczmirek, 2008; Vehovar et al., 2002). For non-probability web surveys, the problem of non-response is hard to define because its evaluation is traceable only in cases where the frame and the chance of selection are known, which is not the case for CVWSs.

Two approaches have been identified to reduce the bias resulting from the inferential problems outlined above and to improve the quality of web surveys (Couper and Miller, 2008). The «design-based» approach attempts to build probability web surveys by using other methods for sampling and recruitment, and, where necessary, providing Internet access to those without it (De Leeuw et al., 2008; Scherpenzeel and Das, 2009). The «model-based» approach attempts to correct for representational biases of non-probability web surveys using different weighting techniques (Bethlehem and Stoop, 2007). Post-stratification (PS) weighting has mainly been applied to correct socio-demographic differences between the sample and the population under consideration. Propensity Score Adjustment (PSA) also aims to correct differences in socio-demographic variables but sometimes has also used «webographic» (attitudinal or behavioural) variables regarding individuals' decisions to participate in web surveys (Lee and Vaillant, 2009; Loosveldt and Snock, 2008; Rosenbaum and Rubin, 1983; Schonlau et al., 2009). As PSA calculation requires two samples —a random sample for calibration and a second calibrated sample—, several research studies have calculated PSA after introducing webographic questions in both surveys. These questions are thought to best capture the differences between the general population and people able and willing to answer Web surveys (Taylor, 2005). Until now, to our knowledge no studies of that kind have ever been conducted in Spain. Although it has been emphasised that in order to generalise web survey results for the whole population, PS and propensity-based weights are necessary (Duffy et al., 2005), the implications of the different adjustment procedures are still under discussion. So far their application has produced rather diverse results, and there is no certainty as to whether the representativeness of web surveys can be improved (Taylor, 2005; Steinmetz et al., 2009; Vehovar et al., 1999). In this respect, this paper aims to contribute to the debate regarding web survey data quality, reliability and validity for scientific use by testing the implementation of the aforementioned adjustment techniques to the WI data.

Although few previous «design-based» measures were implemented during the WI project, the paper uses a «model-based» approach attempting to correct for representational biases of non-probability web surveys using post-stratification and PSA weights. As the WI target population is the Spanish labour force, in the methods developed below it is assumed that the Spanish Labour Force Survey (LFS) sample is representative of the Spanish labour force and that coverage and non-response errors can be captured by comparing WI and LFS data. It is also assumed that statistics and parameters estimated using SES data are population ones. The paper aims to test the extent to which the calculation of weights adjusts the coverage and non-response bias, making it possible to use CVWSs to estimate labour market variables such as mean salaries, inequality indices and salary determinants.

In order to fulfil this goal, the paper is structured as follows. Section 2 sets the WI survey within the context and characteristics of the Spanish internet population by using data from the Media Research Association Survey for Internet Users (AIMC)[1], and it describes the WI bias by using data from the Spanish Labour Force Survey. Section 3 describes the two methods used to calculate weights: PS and PSA. Section 4 evaluates the efficiency of weights by running conventional labour market analyses (mean salaries, inequality indices and salary regressions) using WI data before and after weighting and data from the Spanish Structure of Earnings Survey (SES). Section 5 draws conclusions and section 6 raises discussion regarding the use of web surveys.

2. THE WI SURVEY, INTERNET PENETRATION AND WI SAMPLE BIAS

2.1. *The WI survey and its marketing measures*

The WageIndicator survey (WI) is currently conducted in approximately 50 countries, and it has two main goals, among others. Firstly, it seeks to generate data on labour market issues to increase our knowledge of the socioeconomic determinants of citizens' work life attitudes, preferences and perceptions within an international perspective. The major tool for reaching such a goal is the development of a CVWS based on questionnaires that are

[1] 9.ª Encuesta de la Asociación para la Investigación de los Medios de Comunicación (AIMC) a usuarios de internet.

similar across countries and that are posted on different national web sites[2]. This survey method facilitates the collection of comparable international data on wages and other labour related variables, which are not always available in official surveys, for example data on family life, commuting or job insecurity (see Muñoz de Bustillo and Pedraza, 2009). In addition, because the web survey is continuously posted, it allows for survey modules that change over time. Quick access to homogeneous international data on labour market issues is important, especially in the current context of globalization and rapid technological and social change. Secondly, and as a by-product, the other goal of the WI is to improve web survey methods to facilitate the scientific use of CVWS data. In this respect, although web surveys have many advantages (cost benefits, rapid data collection, ease of processing results, flexibility of questionnaire design and the potential to reach respondents across national borders) there remain unresolved issues having to do with the aforementioned traditional types of survey errors (coverage error, inference from non-probability samples, non-response error and the measurement error). Probably because web surveys are mainly driven by commercial agencies, the dissemination of knowledge of existing web survey methodologies is underdeveloped.

In a CVWS the questionnaire is uploaded on the website and there is no *ex ante* control of the characteristics of the individual (as in a stratified random sample), nor are individuals randomly selected from a universe. The survey is answered in a process of non-controlled self-selection, whereby some individuals complete the questionnaire and others do not. Figure 1 shows the numerous steps (dot line) a person has to take in order to successfully complete the survey.

Firstly, all those without access to the Internet (broadband access) are excluded from the survey. The Spanish internet penetration rate by 2007 was around 44% of the population. Second, Internet attractions direct many surfers to other sites and not every internet user visits the WI homepage. Third, of the small fraction of surfers who visit the site hosting the survey, only a few will be interested in participating in the survey (Porter and Whitcomb, 2003). Finally, only a proportion of those originally willing to answer the questionnaire will go through the whole set of questions. The Spanish WI questionnaire has been available since 2005. Between January 2005 and January 2007 more than 500,000 people visited the web site and 15,000 of them, around 0.08% of the Spanish labour force, completed the questionnaire.

These steps reflect, at the same time, the error sources. However, many of the standard tools for dealing with the problems of underrepresentation in «standard» surveys are not

[2] For more details see *www.wageindicator.org*

FIGURE 1

Stages in the process of filling in a Continuous Voluntary Web Survey such as the WageIndicator, WI

directly applicable in the case of web surveys; in the terms used by Couper (2000): «For surveys where the frame cannot be identified, the problem of non-response is hard to define». Owing to the inherent difficulties in measuring, for example, non-response in open web surveys, researchers have focused on differences in response rates between mail and e-mail questionnaires in order to ascertain, by approximation, whether the non-response rate problem is greater or smaller in web surveys as compared to other types of surveys (Couper et al., 1999; Dillman, 1998; Fricker et al., 2005; Kaplowitz et al., 2004; Kiesler and Sproull, 2001; Llauradó, 2006). The coverage error is clearly present in the first step: people without internet access cannot complete the questionnaire, and the following steps do not have a clear «standard» survey equivalent concept and have been summarized within the ambiguous concept of «self-selection».

Regarding the second step, the error comes from a target population with internet access that does not visit the web site. This non-visiting error cannot be considered either a coverage error, because these people have internet access, or a non-response error, because these people have not been invited to complete the questionnaire.

Although non-response is difficult to define it is related to the third step: every WI homepage visitor is invited to complete the questionnaire; it could be said that CVWS non-response consists of people who after accessing the website do not complete the questionnaire. The fourth step implies non-response to a set of questions.

Together with problems related to sampling, coverage and non-response, web surveys also face problems of measurement errors: differences between the «true» answer and the answer recorded. These measurement errors can be different in web surveys compared to surveys with interviewers who, if properly trained, can explain whatever problems the interviewee might have with the questions.

Some of the bias mentioned can be specifically addressed with the proper resources. Special campaigns aimed at specific underrepresented groups can be developed. But many of the problems remain, especially because the type of action needed to solve specific bias will undoubtedly increase the cost of running web surveys, whereas their low-cost is precisely one of their major attractions. Several measures were taken to increase the number of respondents, but the WI project did not take specific measures to overcome coverage error.

Aiming to increase the number of visits and respondents, the website first gathered useful labour market related information for the public at large. For example, by using the information obtained through the questionnaire and based on salary regressions, web pages with a crowd-pulling Salary Check providing very detailed salary information related to a set of variables such as occupation, age, education, sector of activity, firm size and supervisory position, were constructed. Secondly, nation-wide promotion and publicity consisted of collaborations with trade unions and newspapers. Regarding collaboration with trade unions, there were active links on the web sites of the two main Spanish trade unions (UGT and CCOO). Visitors also had the possibility of drawing up labour market-related questions consisting mainly of personal and very specific situations. They were answered by forwarding visitors to each relevant and specific trade union unit. Consequently, the proportion of affiliated workers in the WI sample was over 25%, whereas this proportion in the Spanish labour force is less than 20% (Visser, 2006). Regarding collaboration with newspapers, there were active links on El País, the main Spanish newspaper. Several news items with conclusions based on the data collected were published. These always increased the number of visits during the week after publication.

To increase women's participation in the survey, which was lower due to their characteristics in the labour market and their lower rate of internet use, specific marketing and targeting measures were taken. Such measures were developed within the framework of WI collaboration with trade unions and newspapers. Trade unions encouraged participation among female members by distributing internal emails. Newspapers published an article on International Women's Day, using WI data to analyse the situation of Spanish women in the labour market.

It is clear that the implemented measures increased the number of visits and encouraged participation, but it is very difficult to monitor, quantify and measure their impact. At the same time, marketing and targeting measures probably introduced other types of bias, for example the participation of more altruistic people who were worried about labour market problems. However, the introduction of behavioural variables in the calculation of weights is very limited because such types of questions are seldom included in the Official Surveys that are used as reference samples.

2.2. *Internet penetration and the WI sample bias*

Internet coverage, one of the main problems of CVWSs, was not tackled within the marketing and targeting measures. It is important to bear in mind that WI data have been collected within the Spanish context of internet penetration that can be studied using the 9th AIMC survey. The 9th AIMC survey was conducted during 2006 in order to collect information regarding the population of Spanish internet users. The following can be concluded according to AIMC data:

a. The Spanish internet population is characterized by a higher level of education than that of the overall population. While 12.5% of the Spanish population have no education and 21% have only completed primary education, only 0.4% and 9.6% of the internet population have these two educational levels. While only 22% of the Spanish population holds a university degree, more than 46% of the internet population have had a university education.

b. Internet use is more widespread among younger people, as 27% of the Spanish population is over 54 years old, compared to only 3.5% of the internet population.

c. Internet use is more widespread among men. While 50.7% of the Spanish population is female, only 30% of internet users are women.

d. Internet users are concentrated in regions such as Madrid and Catalonia. Madrid's population constitutes 13.6% of the Spanish population compared to more than 17% of internet users; and Catalonia's population makes up 15.8% of the Spanish population compared to more than 22% of internet users. In contrast, in some regions the proportion of internet users is far below their share of the Spanish population. For example, almost 18% of the Spanish population live in Andalusia but only 12.7% of internet users do so.

Although the WI target population is the Spanish labour force, and not the whole Spanish population, WI sample demographic characteristics do not differ greatly from the AIMC sample, with the sole exception of the gender balance. Comparing the 2006 Spanish Labour Force Survey (LFS) sample demographic characteristics with those of the WI sample, the following can be concluded:

a. Educational levels are higher in the WI sample as compared to the Spanish labour force. For example, the proportion of Spanish workers who have completed university studies, 22%, is lower than the proportion of university educated workers who have completed the WI questionnaire, 50%. In contrast, the proportion of workers who have finished only secondary education is higher in the Spanish labour force, 60%, than in the WI sample, 33%. Finally, the proportion of workers who have completed primary education is very similar in both samples, 15% in the LFS and 13% in the WI, and the proportion of non-educated workers is, notably, the same 4%.

b. Age has a clear impact on the WI sample. More than 10% of the Spanish labour force is 54 years or older, as compared to only 3% of the WI sample. It is important to note that younger workers, under the age of 24, are also underrepresented in the WI sample, given that they account for more than 11% of the LFS sample but only 7% of the WI one. This underrepresentation can be explained by the fact that although younger people are more familiarized with internet use, most workers joining the labour force before reaching the age of 24 are low-educated workers.

c. Although internet use is more widespread among men, women are not underrepresented in the WI sample: the proportion of women is around 40% in both the LFS and the WI samples. The difference between both samples is less than 1%, which is comparable to the difference between the Structure of Earnings Survey (SES), commonly assumed to be a representative sample of Spanish labour force, and the LFS. However, during the early stages of the Spanish WI survey, before specific marketing and targeting measures for women were implemented, the proportion of females in the sample was much lower, around 33%.

d. Regarding the regional distribution, the proportion of workers in the WI sample from regions such as Madrid, 23.5%, and Catalonia, 19.2%, where the internet is more widespread, is higher than the proportion of workers from those regions in LFS, 14.7% and 16.9%, respectively. In contrast, regions such as Andalusia, where the proportion of internet users is below its share of the Spanish population, are underrepresented in the WI sample: while 16.5% of the Spanish labour force is concentrated in Andalusia, only 13% of the WI sample work in this region.

It can be concluded that the WI is quite successful in terms of the number of visits and visitors that it attracts, as well as in terms of responses to its questionnaire: almost 15,000 completed questionnaires at a much lower cost than any other sampling method. Nevertheless, it is clear that important problems remain to be solved with regard to the representativeness of the sample. Access to and use of the Internet is biased by educational levels, age, and geographical units, which clearly influences the demographic characteristics of a CVWS, making the coverage problem the most important one at present. In fact, unless specific marketing and targeting measures are taken, such as those implemented to overcome the WI gender bias, internet access, in this case internet population characteristics described by the AIMC survey, is able to predict CVWS bias according to a set of demographic variables, in this case the WI sample bias.

At the same time, the coverage error might be expected to become progressively less important in the future as the internet population becomes more similar to the whole population due to the increase in access to the internet (Chen et al., 2002; López, 2008; López, 2006). For example, in The Netherlands, a country where internet penetration is much higher, the Dutch WI survey, launched in 2001, is already accessed by gardeners and other workers in low-wage occupations. Furthermore, in the Dutch WI sample, underrepresented groups in terms of age have also decreased their underrepresentation over the years. While underrepresentation of older workers in Spain starts at age 45, the equivalent age in the Dutch WageIndicator is 55. This seems to be a matter of the age-technology gap and Internet use by older workers. For this reason, in the absence of marketing measures, the age-technology gap implications of web surveys might be solved with the passing of time as workers in the overrepresented age intervals —those under 45 in Spain— grow older and reach the underrepresented age groups of 45-59 and 60 and over. In that case, while the Dutch age-technology gap might be solved within a decade this process will take more than 15 years in Spain.

In the next section we use the above set of demographic variables (education, age and region) to weight the data obtained by CVWSs. We calculate two types of weights to attempt to correct for representational biases.

3. THE «MODEL-BASED» APPROACH AND THE WI SAMPLE

CVWSs enter uncharted territory in terms of the state of the art of sampling and surveying methods. A random sample —the standard procedure followed by surveys— aimed at collecting data from a population in which every individual has the same probability of being selected, can be analysed by using the standard inference procedures. In contrast, as mentioned in the introduction, CVWSs face several problems that make a proper analysis and interpretation of the results much more difficult[3]. As shown in the previous section, the probability of being selected for a CVWS differs among Spanish regions, increases with educational level and decreases with age. This section follows a «model-based» approach that attempts to correct for representational biases of non-probability web surveys by using two weighting techniques: Post-stratification (PS) and Propensity Score Adjustment (PSA). In both cases the conditional variables are: regions, age and educational level.

3.1. *The «post-stratification technique» (PS)*

Conventional PS weights have been calculated by using age, education and Spanish regions as conditional variables. The following steps have been taken. Firstly, the WI indicator sample and the LFS were divided into 24 groups, using four age intervals (15-24, 25-54, 55-64, 65 and over), three educational levels (low, medium, high) and two types of regions: regions overrepresented in the WI sample (Catalonia, Madrid, Aragón, La Rioja and Castile-Leon) and regions underrepresented (Andalusia, Asturias, Balearic Islands, Canary Islands, Cantabria, Castile-La Mancha, Extremadura, Galicia, Murcia, Navarre, Basque Country, Ceuta, Melilla and Valencia). Secondly, the proportion of each group in both samples was calculated. For example, 2.9% of the WI indicator sample are workers between the ages 35 and 44 low-educated and living in an overrepresented region. At the same time, according to the LFS, 2.8% of the Spanish labour force are workers aged between 35 and 44 years old, highly educated and living in an overrepresented region.

[3] A good introduction to the specificities of web surveys can be found in Couper (2000) and, from a different, more practical perspective, in Dillman and Bowker (2001).

Thirdly, a *weight variable* was obtained by dividing the population proportions by the sample proportions:

$$weight = \frac{n_{population}}{n_{sample}} = \frac{\%LFS}{\%WI}$$

The resulting weight in the example above is 0.9597. This methodology allows us to weight, by a number greater than 1, groups of age, educational level and region whose representation in the *WageIndicator* sample is lower than their proportion in the population and, to weight by a number less than 1, groups whose representation in the *WageIndicator dataset* sample is higher than their proportion of the population.

Fourthly, in order to avoid excessively high weight values, whenever the difference between the WI proportion and the LFS proportion involved a weight of more than two, two or more groups were merged together. Two merging principles were followed when merging groups. Two groups with different educational levels were never merged together and, whenever possible, neither were two groups from different age intervals. There are labour market related reasons behind these merging principles (it is important to bear in mind that the data are collected for labour market analysis). Although working in one region or another might be an important determinant of an individual situation within the labour market, age and education have a greater impact on variables such as job stability and type of contract.

3.2. *The Propensity Score Adjustment (PSA)*

In order to compare the efficiency of the PS and PSA techniques, we calculated PSA weights by using the same conditional variables: age, education and Spanish regions. To calculate PSA weights the following steps were taken.

Firstly, the WI sample was merged with the Spanish Structure of Earnings Survey (SES). Secondly, a probit regression was calculated, using the abovementioned conditional variables as explanatory variables, to estimate individual probabilities of participating in the WI sample. The dependent variable of the probit regression takes a value of 1 when the observation comes from the WI and 0 when it comes from SES. Finally, by using the probit estimated probabilities of participating in the WI survey, the inverse of that probability was calculated for weighting the WI data. The calculated weight takes a higher value when the probabilty of participating in the WI, according to the selected conditional variables, is lower.

It is important to note that the PSA calculation is limited by the lack of attitudinal or «webographic» questions in the SES survey.

4. DOES WEIGHTING SOLVE THE PROBLEM?

Mean salaries, inequality indices and salary regressions were calculated before and after implementing weights to the WI data. The results were compared with those obtained using the SES.

Starting with mean annual salaries, according to the Spanish SES the mean annual salary in 2006 was 18,888.18€. According to the WI it was 22,902.81€. The difference can be easily explained by using the conditional variables that were used to calculate weights. Firstly, in the WI sample educational levels are higher than in the SES, which implies higher salaries. Secondly, the overrepresented age intervals in the WI are those in which salaries rise faster and the underrepresented age intervals are either those in which salaries are very low (16-24) or those in which they start rising more slowly (over 55). Finally, salaries in cities such as Madrid and Barcelona are higher.

The estimated mean salary after adjusting the WI sample by using PS weights was 21,903.06€, whereas by using PSA weights it was 21,288.67€. Both are closer to the population value; however, both are still far from it.

TABLE 1

Mean salaries and inequality indices according to SES and WI weighted and non-weighted data

	SES	WI	PS WI	PSA WI
Mean salary (standard error)	€18,888.18 (33.46)	€22,902.81 (212.63)	€21,903.06 (251.95)	€21,288.67 (351.81)
Wage-Gini-index	0.3687	0.3596	0.3593	0.3645

(standard errors betwen brackets)

Due to the high number of observations available for the SES sample —almost 200,000—, the standard error is the lowest among the four estimations. The implementation of both types of weights implies an increase in standard errors. It is important to note that the higher the number of variables used to calculate weights, the closer the estimated mean is to the population mean, but the higher the increase in standard errors. Weights including a more detailed division of Spanish regions were also calculated. Here, the estimated mean salary decreased to 20,511.23€, which is closer to the SES mean, but the standard error increased to 757.11. A potential solution to account for more variables without increasing the standard errors could be the implementation of «design-based» measures such as marketing and targeting measures for specific underrepresented groups. Comparing PSA and PS weights, PSA is more efficient in moving the estimated mean closer to that of the population but PS involves a lower impact on standard errors.

Regarding the inequality indices, the WI Gini index is around one point higher than that of the SES. After implementing PSA weights, the WI Gini index is almost the same as that derived from the SES. However, PS did not change the estimated Gini index.

Regarding the salary regressions, the estimations using SES data show that being a woman implies a lower salary (38% lower), that being one year older increases a salary by 6% and that wage increases due to age are lower the older the worker, as is captured by the negative impact of the variable age square. These three impacts are also captured in the regression using WI data, although WI regression captures lower impacts (being a woman only implies a 24% wage penalty and being a year older only a 5% wage increase). The estimated coefficients after weighting WI data using PSA, although still lower, are closer to the SES ones. At the same time, significant values after weighting are lower than those for the SES and the WI estimations due to the increase in standard errors that weighting produces. The same applies to the estimated coefficients for educational levels: the impact is lower in the WI regression than in the SES one, and PSA weights are successful in moving estimated parameters in the right direction, but significance values decrease considerably after weighting (table 2).

A researcher using only WI data would be able to correctly predict the sign of the impact of each explanatory variable on wages. With the exception of working in construction, which is not significant in the WI and the PSA WI regressions but is significant according to the SES, almost every explanatory variable considered a significant explanatory variable of wages according to the SES would be considered a significant variable using both WI, PS WI and PSA WI regressions.

However, a researcher using only WI data would not be able to correctly predict the size of the impact of each explanatory variable on wages. WI estimations predict lower impacts for every explanatory variable and PSA weights are successful in correcting estimated parameters by moving them in the right direction. At the same time, although the significance values are lower after weighting, with the exception of age squared and working in construction, no erroneous conclusion regarding the identification of significant

TABLE 2

Salary Regressions (Dependent variable defined as gross logarithm)

	SES	WI	Post-strat. WI	PSA WI
Women	–.38***	–.2369***	–.2618***	–.3237***
Age	.0623***	.0511***	.04***	.0549*
Age squared	–.0006***	–.0004***	–.0003***	–.00042
Educational level				
No education	–.84***	–.5674***	–.6113***	–.5786**
Primary	–.7015***	–.4856***	–.4962***	–.5692***
Vocational training I	–.4491***	–.4051***	–.396***	–.413***
Secondary I	–.6657***	–.3568***	–.3593***	–.4417***
Secondary II	–.4292***	–.324***	–.3242***	–.3721***
Vocational training II	–.3795***	–.292***	–.3185***	–.3049***
University I	–.1332***	–.1233***	–.1296***	–.1446***
Postgraduate	.106***	.1207***	–.1145**	.1184***
Activity sector				
Construction	.0718***	.0082	.0124	.0587
Services	–.1452***	–.0939***	–.0934***	–.1057**
Adj R-squared	0.3875	0.2624	0.2767	0.3177
Number of obs.	197 626	14 240	14 240	14 240

*, **, *** significant at 0.05, 0.01 and 0.001 respectively.

variables could be obtained from the PSA WI regression. In other words, the reduction in significance levels is not big enough to lead to the (erroneous) conclusion that a variable is not significant when it actually is. As in the calculation of mean salaries, estimated parameters using PS weights are farther from population parameters than those using PSA. However, the impact on significance values is much lower.

Therefore, it can be concluded that although weighting procedures correct estimated parameters in the right direction, they increase standard errors and reduce significance values. PSA estimated parameters are closer to population ones. Although it is not the case in our case study, it is important to note that the reduction in the levels of significance may lead to the erroneous conclusion that an explanatory variable is not significant when it actually is. Such a risk is lower when using PS. In this respect, some authors argue that weighting is useful when calculating population characteristics but not when modelling (Pudney, 2009; Escobar and Fernández, 2010; Angrist and Pischke, 2009) while others argue that weight should be used in both cases (Deaton, 1997).

5. CONCLUSIONS

Although both weighted methods analysed improve the indices obtained by the web-based survey in terms of reducing their differences with the indices generated by the SES, adjusted means, inequality indices and salary regression estimated parameters still differ from population magnitudes. Therefore, it is not possible to conclude that the Spanish labour market magnitudes can be reliably estimated by using a CVWS, in this case the WI data. However, neither can it be concluded that the «model-based» approach is useless. The two types of adjustment measures evaluated in this paper corrected estimated statistics in the right direction. It should be tested whether several complementary measures (marketing and targeting measures, incentives or use of mixed modes in Official Surveys) could improve these results. Future efforts should focus on developing joint methodologies for the «design-base» and the «model-base».

Firstly, it should be recalled that the Spanish WI project, with very meagre resources, was able to overcome the gender bias. Marketing measures have also been proven to be effective in the Dutch WI, where the project has received much more public support. As personalization increases response rates (Heerwegh, 2005), the use of marketing measures in the Spanish WI, focusing on other underrepresented groups such as older and younger workers, would improve the efficiency of the «model-based» adjustment techniques. In fact, complementary measures such as incentives or marketing and targeting strategies would facilitate the calculation of weights using less conditional variables, involving a lower impact on standard errors and significant values. However, as a drawback, this type of measure increases web survey costs.

Secondly, taking into account the considerable amount of savings that web surveys may involve for Statistical Institutes, projects such as the WI should be encouraged and supported. This may consist of economic support but should also involve collaboration with web survey experts. Although expensive, a good starting point for such collaboration could be the inclusion of «webographics» and attitudinal questions in Official Surveys questionnaires.

Thirdly, in parallel to Official Surveys such as the SES, Statistical Institutes could start making use of CVWSs and/or mixed-mode web surveys with the same target population. Mixed-mode surveys have a clear advantage with respect to CVWSs because, if well designed, they are able to isolate different types of bias, which at the same time makes it easier to calculate the weights. The use of CVWSs and/or mixed-mode web surveys would not necessarily imply giving up the current sample design but would make it easier to monitor, study and facilitate the transition to the use of new technologies in data collection. In fact, in the USA the use of new technologies in Official Surveys is much more widespread.

6. DISCUSSION

Most European Statistical Institutes are not yet paying attention to web surveys, nor are they supporting them. Many survey specialists are constantly neglecting the need to increase research activities to give web surveys scientific validity. As a result, the European Survey scientific community is moving slowly and sceptically towards the adaptation of survey methodologies to the constantly and rapidly improving technological and communication changes. What is more, national and European calls for proposals are not clearly supporting this research stream. Although many citizens can already, for example, make their yearly income tax returns on line, almost no Official European Survey uses the web as a mode of data collection. Notable exceptions in this respect are countries such as Norway and The Netherlands, where academic driven projects such as the LISS panel are supported. The LISS panel reduces web non-coverage bias by giving internet access to certain parts of the population and facilitates the calculation of weights by including «webographic» questions, which is paving the way towards a scientific-based web survey methodology. Although on the frontiers of knowledge, the current state-of-the-art on the use of the web as a means for data collection is sufficiently advanced for Statistical Institutes to, at least, start testing and exploring new modes. However, under the excuse of non-coverage and non-response bias, European social scientists are forced to constantly lag behind their counterparts in the USA in terms of profiting from web survey advantages. In this respect, suffice it to say that although the Spanish SES data used here refer to 2006 wages, they were not available to researchers until 2008. Within the current context of globalization and rapid changes, such a slow mode of data collection cannot be considered optimal for rigorous political advice. The option of waiting until the technological device is solved with the passing of time in order to take advantage of cheaper and quicker new technologies does not seem to be a very intelligent one, since nobody knows what the state of technology will be in ten years' time. New technological changes will most likely produce new technological devices.

To sum up, the «model based» approach has good qualities in the sense that estimated statistics after weighting WI data are closer to population statistics. The results can be criticized, on the one hand, because estimations still differ from population values and, on the other hand, because the implementation of weights increases standard errors. Future research should test firstly whether results could improve in the presence of more active «design-based» measures focusing on demographic characteristics such as education and age; and secondly, whether the inclusion of «webographics» and attitudinal variables as conditional variables in the calculations will improve the results found here. Resources are needed for both purposes because although the commercial sector very frequently uses web surveys, more effort is needed to give the web survey methodology credibility and validity for scientific use, which can only be done and disseminated by academics in the name of public interest.

REFERENCES

Angrist, Joshua D. and Jörn-Steffen Pischke (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton: Princeton University Press.

Bandilla, Wolfgang, Michael Bosnjak and Patrick Altdorf (2003): «Survey administration effects? A comparison of web-based and traditional written self-administered surveys using the ISSP environment module», *Social Science Computer Review*, 21 (2): 235-243.

Bandilla, Wolfgang et al. (2009): «Coverage und Nonresponse-Effekte bei Online-Bevölkerungsumfragen», in *Sozialforschung im Internet*: *Methodologie und Praxis der Online-Befragung,* eds. N. Jackob, H. Schoen and T. Zerback, VS Verlag für Sozialwissenschaften.

Bethlehem, J. and I. Stoop (2006): «Online Panels – A Paradigm Theft?», in *Annual Meeting of the Association for Survey Computing.*

Birnbaum, Michael H. (2004): «Human Research and Data Collection Via the Internet», *Annual Review of Psychology,* 55: 803-832.

Bosnjak, Michael and Tracy L. Tuten (2003): «Prepaid and promised incentives in web surveys», *Social Science Computer Review*, 21 (2): 208-217.

Bronner, Fraser and Ton Kuijlen (2007): «The live or digital Interviewer: A comparison between CASI, CAPI and CATI with respect to differences in response behaviour», *International Journal of Market Research,* 49 (2): 167-190.

Chen Wenhong, Jeffrey Boase and Barry Wellman (2002): «The Global Villagers: Comparing Internet Users and Uses around the World», in *The Internet in Everyday Life*, ed. Wellman, Haythornthwaite, Oxford: Blackwell.

Coderre, François, Anne Mathieu and Natalie St-Laurent (2004): «Comparison of the quality of qualitative data obtained through telephone, postal and email surveys», *International Journal of Market Research,* 46 (3): 347-357.

Couper, Mick P., Johnny Blair and Timothy Triplett (1999): «A Comparison of Mail and E-Mail for a Survey of Employees in Federal Statistical Agencies», *Journal of Official Statistics,* 15 (1): 39-56.

Couper, Mick P. (2000): «Web surveys: a review of issues and approaches», *Public Opinion Quarterly*, 64 (4): 464-481.

Couper, Mick P. and Peter Miller (2008): «Web survey methods – Introduction», *Public Opinion Quarterly*, 72 (5): 831-835.

Couper, Mick P. et al. (2008): «Risk of disclosure, perceptions of risk, and concerns about privacy and confidentiality as factors in survey participation», *Journal of Official Statistics*, 24 (2): 255-275.

Couper, Mick P. et al. (2007): «Noncoverage and nonresponse in an Internet survey», *Social Science Research*, 36 (1): 131-148.

Couper, Mick P., Eleanor Singer and Roger Tourangeau (2004): «Does voice matter? An interactive voice response (IVR) experiment», *Journal of Official Statistics*, 20 (3): 551-570.

De Leeuw, Edith D., Joop J. Hox and Don A. Dillman (2008): *International Handbook of Survey Methodology,* New York, NY: Lawrence Erlbaum Associates.

Deaton, Angus (1997): *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. ISBN 0-8018-5254-4.

Dever, Jill A., Ann Rafferty and Richard Valliant (2008): «Internet surveys: can statistical adjustments eliminate coverage bias?», *Survey Research Methods*, 2 (2): 47-62.

Dillman, Don A. and Denise Bowker (2001): «The Web Questionnaire Challenge to Survey Methodologists», in *Dimensions of Internet Science,* eds. Ulf-Dietrich Reips and Michael Bosnjak, Lengerich: Pabst Science Publishers.

Dillman, Don A., Jolene Smyth and Christian Leah (2009): *Internet, Mail and Mixed-Mode Surveys: the tailored design method*, ed. John Wiley, Holboken, 3th Edition.

Duffy, Bobby et al. (2005): «Comparing Data from online and Face-to-Face surveys», *International Journal of Market Research*, 47 (6): 615-639.

Escobar, Modesto, Enrique Fernández and Fabrizio Bernardi (2009): *Análisis de datos con Stata,* Madrid: Centro de Investigaciones Sociológicas.

Fricker, Scott et al. (2005): «An Experimental Comparison of Web and Telephone Surveys», *Public Opinion Quarterly*, 69 (3): 370-392.

Fricker, Ronald D. (2008): «Sampling methods for web and E-mail surveys», in *The SAGE Handbook of Online Research Methods*, eds. N. Fielding, R. M. Lee and G. Blank, London: Sage.

Fricker, Ronald and Mathias Schonlau (2002): «Advantages and disadvantages of Internet research surveys: Evidence from the literature», *Field Methods*, 14: 347-367.

Galesic, Mirta (2006): «Dropouts on the web: effects of interest and burden experienced during an online survey», *Journal of Official Statistics*, 22 (2): 313-328.

Göritz, Anja S. (2006): «Cash lotteries as incentives in online panels», *Social Science Computer Review*, 24 (4): 445-459.

Healey, Benjamin (2007): «Drop downs and scroll mice: the effect of response option format and input mechanism employed on data quality in web surveys», *Social Science Computer Review*, 25 (1): 111-128.

Heerwegh, Dirk (2005): «Effects of personal salutations in E-mail invitations to participate in a web survey», *Public Opinion Quarterly,* 69 (4): 588-598.

Ilieva, Janet, Steven Baron and Nigel M. Healey (2002): «Online surveys in marketing research: pros and cons», *International Journal of Marketing Research*, 44 (3): 361-376.

Kaczmirek, Lars (2008): *Human-Survey Interaction: Usability and Nonresponse in Online Surveys*, ed. Mannheim: MADOC.

Kaplowitz, Michael, Timothy Hadlock and Ralph Levine (2004): «A comparison of web and mail survey response rates», *Public Opinion Quarterly*, 68 (1): 94-101.

Kiesler, Sara and Lee Sproull (2001): *Human Values and the Design of Computer Technology*, ed. Centre for the Study of Language and Information.

Lee, Sunghee and Richard Vaillant (2009): «Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment», *Sociological Methods Research*, 37 (3): 319-343.

Llauradó, Oriol (2006): «El trabajo de campo online: qué hemos aprendido en los últimos 10 años», *Investigación y Marketing* (Monográfico sobre la investigación online; el impacto de la nuevas tecnologías), 91 (junio): 25-40.

López, Antoni (2006): «Que las escasas diferencias no nos impidan aprovechar las ventajas de Access Panel Online», *Investigación y Marketing,* 91 (junio): 20-24.

López, Antoni (2008): «¿Seguimos hablando del on-line?», *Investigación y Marketing,* 98 (marzo): 28-30.

Loosveldt, Geer and Nathalie Snock (2008): «An evaluation of the weighting procedures for online access panel surveys», *Survey Research Methods*, 2 (2): 93-105.

Lozar Manfreda, Katja et al. (2008): «Web surveys versus other survey modes: A meta-analysis comparing response rates», *International Journal of Market Research*, 50 (1): 79-104.

Lynn, Peter (2008): «The problem of nonresponse», in *International Handbook of Survey Methodology*, eds. E. D. De Leeuw, J. Hox and D. A. Dillman, New York: Lawrence Erlbaum Associates.

Malhotra, Neil and Jon A. Krosnick (2007): «The effect of survey mode and sampling on inferences about political attitudes and behaviour: comparing the 2000 and 2004 ANES to Internet surveys with nonprobability samples», *Political Analysis*, 15 (3): 286-324.

Muñoz de Bustillo, Rafael and Pablo de Pedraza (2010): «Determinants of Job Insecurity in 5 European Countries», *European Journal of Industrial Relations,* 16 (1): 5-20.

Porter, Stephen R. and Michael E. Whitcomb (2003): «The Impact of Contact Type on Web Survey Response Rates», *Public Opinion Quarterly*, 67 (4): 579-588.

Pratesi, Monica et al. (2004): «List-based web surveys: Quality, timeliness, and nonresponse in the steps of the participation flow», *Journal of Official Statistics,* 20 (3): 451-465.

Rosenbaum, Paul R. and Donald B. Rubin (1983): «The central role of the propensity score in observational studies for causal effects», *Biometrika*, 70: 41-55.

Roster, Catherine et al. (2004): «A comparison of response characteristics from web and telephone surveys», *International Journal of Market Research,* 446 (3): 359-373.

Schaefer, David R. and Don A. Dillman (1998): «Development of a Standard E-Mail Methodology: Results from an Experiment», *Public Opinion Quarterly*, 62 (3): 378-397.

Scherpenzeel, A. and M. Das (2009): «True' Internet panels: evidence from the Netherlands», in *Social Research and the Internet: Advances in Applied Methods,* eds. M. Das, P. Ester and L. Kaczmirek, Boca Raton: Taylor & Francis.

Schneider, Sid J. et al. (2005): «Telephone, internet, and paper data collection modes for census 2000 short form», *Journal of Official Statistics,* 21 (1): 89-101.

Schonlau, Mathias et al. (2009): «Selection bias in web surveys and the use of propensity scores», *Sociological Methods Research*, 37 (3): 291-318.

Shih, Tse-Hua and Xitao Fan (2008): «Comparing response rates from web and mail surveys: A meta-analysis», *Field Methods*, 20 (3): 249-271.

Steinmetz, Stephanie, Kea Tijdens and Pablo de Pedraza (2009): «Comparing different weighting procedures for volunteer on line panels – lessons to be learned from German, Dutch and Spanish Wage Indicator data». Presented at *ESRA Conference* 2009 Warsaw.

Sparrow, Nick (2006): «Developing Reliable On Line Polls», *International Journal of Market Research,* 48.

Taylor, Humpfrey (2005): «Does Internet research "work"? Comparing online survey results with telephone surveys», *International Journal of Market Research*, 42 (1): 51-63.

Tingling, Peter, Michael Parent and Michael Wade (2003): «Extending the capabilities of internet-based research: lessons from the field», *Internet Research*, 13 (3): 223-235.

Tourangeau, Roger, Mick Couper and Frederick Conrad (2006): «Color, labels, and interpretive heuristics for response scales», *Public Opinion Quarterly*, 71 (1): 91-112.

Tuten, Tracy L., David J. Urban and Michael Bosnjak (2002): «Internet Surveys and Data Quality - A Review», in *Online Social Science*, eds. B. Batinic, U. D. Reips and M. Bosnjak, Seattle, WA: Hogrefe and Huber Publishers.

Vehovar, Vasja and Katia Manfreda (1999): «Web surveys: Can the weighting solve the problem?», *Research Methods*, pp. 962-967.

Visser, Jelle (2006): «Union membership statistics in 24 countries», *Monthly Labour Review* (January): 38-49.

RECEPCIÓN: 26/10/2009
REVISIÓN: 21/01/2010
APROBACIÓN: 16/02/2010