AIAS

Amsterdam Institute for Advanced labour Studies

Comparing different weighting procedures for volunteer web surveys

Stephanie Steinmetz, Kea Tijdens and Pablo de Pedraza

Working Paper 09-76

November 2009

AIAS

University of Amsterdam

Acknowledgements

The authors would like to thank the Virtual Knowledge Studio Amsterdam and the Erasmus Studio Rotterdam for the funding of this study in the framework of the project 'Improving web survey methodology for social and cultural sciences'. Moreover, the authors are grateful for many helpful comments and suggestions for improvement which they have received from participants of the ESRA conference 2009,the 3rd MESS Workshop, the AIAS Lunch seminar and the EUR-FSW Brown bag seminar.

November 2009

This paper can be downloaded from our website www.uva-aias.net under the section Publications/Working papers.

Comparing different weighting procedures for volunteer web surveys

Lessons to be learned from German and Dutch WageIndicator data

Stephanie Steinmetz

Erasmus University Rotterdam,
Department of Sociology

Kea Tijdens

Erasmus University Rotterdam,
Department of Social Sciences
University of Amsterdam,
Amsterdam Institute for Advanced
Labour Studies AIAS

Pablo de Pedraza

Universidad de Salamanca
Faculty of Law and Department of Applied Economics

WP 09/76

Abstract

The strengths and weaknesses of web surveys have been widely described in the literature. Of particular interest is the question to which degree the obtained results can be generalised for the whole population? To deal with this problem weighting adjustments, like post-stratification and propensity score adjustment (PSA) have been seen as a possible solution. In the scientific community, however, particularly PSA has traditionally not been applied in the field of surveys, and there has been a minimal amount of evidence for its applicability and performance, and the implications are not conclusive. Against this background, the paper attempts to explore the two statistical weighting procedures for the German and Dutch WageIndicator Survey 2006. To evaluate the effectiveness of the weighting techniques in adjusting biases arising from non-randomised sample selection, the existing selection bias has been explored and the efficiency of the weights has be tested by comparing un-weighted and weighted results with those that could be found using data from the German SOEP and the Dutch OSA Panel for the same year. The results reveal that the impact of the applied weights is very limited and that the different weighting methods using balancing variables do not make web survey data more comparable to the general population. This holds for the German as well as for the Dutch sample.

Keywords: web surveys, volunteer web surveys, selection bias, post-stratification weight, propensity score weight, PSA, representativeness

Table of contents

1. Introduction

In the last decades, the web has become a popular tool of data collection not only for commercial marketing agencies but also for scientific purposes. In this context, the introduction of web surveys has triggered a heated debate about their scientific validity (Couper 2000, Fricker and Schonlau 2002, Ilieva et al. 2002, Tingling et al. 2003, Tuten et al. 2002). Arguments in their favour emphasize cost benefits, fast data collection, ease of processing results, flexibility of questionnaire design, and the potential to reach respondents across national borders. In particular, they enable multi-country and multilingual homogenised surveys that are crucial in the current context of globalisation. Arguments against web surveys mainly focus on traditional types of survey errors and related questions of their quality and reliability for scientific use. Particularly non-probability based web surveys are problematic because respondents are not selected at random, and the target population forms a convenience rather than a probability sample. Therefore, very little is known about the degree to which the obtained results can be generalised for the whole population.

To deal with these problems and improve the quality of web survey estimates, different weighting techniques, like *post-stratification* and *propensity score adjustment* (PSA), have been considered. Post-stratification weighting has mainly been applied to correct for socio-demographic differences between the web sample and the population under consideration, whereas PSA aims to correct for differences in socio-demographic and 'webographic' (attitudinal or behavioural) variables regarding individuals' decisions to participate in web surveys (Lee and Vaillant 2009, Loosveldt and Sonck 2008, Schonlau et al. 2009). Although it has been emphasised that for generalising web survey results for the whole population, post-stratification and propensity-based weights are necessary, the implications of the different adjustment procedures are still under discussion (Bethlehem and Stoop 2007, Taylor 2005, Vehovar et al. 1999). As, their application has produced rather diverse results, there is no certainty as to whether the representativeness of web surveys can be improved through weighting.

Against this background, the paper attempts to explore the two above-described weighting procedures in more detail, and evaluate their effectiveness in adjusting biases arising from non-randomised sample selection. Furthermore, the comparison of German and Dutch data will allow us to determine whether selection bias takes on similar patterns across countries.[1] Therefore, the next section will provide an overview of existing knowledge concerning the specific problems of non-probability web survey and the efficiency of post-stratification and PSA techniques. Section three will introduce the different data sets and weighting techniques. In section four, first the biases for the two countries will be described which is essential for exploring the problems of the used data and selecting the variables which might be important for the weights. Second, the efficiency of the different weighting techniques are tested by comparing un-weighted and weighted results from the German and Dutch web survey data with reference data from the German Socio-Economic Panel (SOEP) and the OSA Labour Supply Panel for 2006. Section five, finally, will discuss the findings and the sensitivity of the results cross-nationally and will particularly devote attention to changes in the specification of the PSA.

1 Germany and the Netherlands have been selected because these countries have the highest participation rates in the WageIndicator Survey. Moreover, it was also easy to get access to probability-based reference surveys.

2. Survey quality and sources of error in non-probability web surveys

When the primary purpose of a survey is to gather information about the general population, the information is useless unless it is accurate and representative in this regard. One fundamental element of data quality to be considered in this context is bias[2]. To minimise bias, researchers have traditionally attempted to create samples that provide a reliable cross-section of a given population allowing to draw random, or probability-based samples which produce representative results for the entire population.

2.1. Types of websurveys

With respect to the representativeness of web surveys, a first important clarification is related to the type of recruitment approach. Following Couper (2000) two types of web surveys can be distinguished: *Probability-based* web surveys have the advantage of a proper sample frame which allows the drawing of a probability-based random sample from a population in which every individual has the same probability of being selected. For probability-based web surveys, such as intercept, e-mail request, mixed-mode surveys, and pre-recruited access panels of Internet users[3], that means that all members of the target population are known (the contact or email addresses). Such data can easily be analysed using standard inference procedures and it allows the generalisation of viewpoints across the target population.

In contrast, *non-probability-based* web surveys, like entertainment surveys, self-selected web surveys, and surveys made up of volunteer panels of Internet users, are problematic because not every individual has the same probability of being selected. For, instance, in volunteer web surveys, open invitations on websites are used to select respondents. The probability of receiving such an invitation

2 Defined as the differences between a statistically calculated value and the true population value of the estimate in the target population.

3 Web-based access panels are constructed by wide appeals on well-visited sites and Internet portals. At time of registration, basic demographic variables are asked. In this way a large database of potential respondents is created for future surveys. Only panel members can participate in these web panel surveys. Even though it seems that on this basis a probability-based sample can be drawn, it has to be emphasised that, such access panels also face the problem of self-selection. Also here, the target population is not well defined, and the final sample consists of self-selected online- or offline pre-recruited persons who have agreed to be a member of the panel.

is unknown as well as the probability of accepting it. The probability of being confronted depends on national or regional Internet access rates, and the number of unique visitors of the website. The latter depends on the website's marketing strategies. Due to the absence of an adequate sampling frame and the application of self-selection recruitment methods, data of such surveys form a convenience rather than a probability sample. The degree to which the obtained results can be generalised for the whole population can hardly ever be ascertain.

2.2. Sources of errors for (non-)probability based web surveys

A second aspect of data quality is related to survey errors, such as coverage, sampling, non-response and measurement errors, which are common to all modes of data collection (even a census). Bias is introduced in survey estimates to the extent that those not covered, not recruited, and/or not surveyed are different from those who are covered, are recruited and respond (Groves 2004). For web survey samples, bias typically stems from three main sources:

a) **coverage error** (identifying target population and defining sampling frame: as not all persons have access to the Internet and, those who have differ significantly from those without in terms of socio-demographic and behavioural characteristics)

b) **sampling error** (drawing a sample from a sampling frame: problem particularly for non-probability based web surveys which build upon self-selection recruitment);

c) **non-response error** (contacting respondents: not all selected people are willing or able to complete the survey, those who do differ significantly from those who don't in terms of socio-demographics and behavioural characteristics).

Against this background, it becomes clear that conducting a proper high-quality (web) survey is an ambitious undertaking. Even though some errors can be avoided by taking preventive measures at the design stage, some errors will remain. This applies also to web surveys, and some problems are even more severe for them.

2.2.1. (Under)Coverage error

At present, the (under) coverage error is a serious problem for many web surveys, particularly for those targeting the general population. It occurs when elements in the target population do not appear in the frame population. In order to study a target population, the researcher needs to define a sampling frame from which to draw a sample. In web surveys, the sampling frame would usually be a list of e-mail addresses of the members of the target population. As not every person has Internet access, and 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.[4] Even though Internet penetration rates continue to increase, the possible bias is moreover not only related 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). Recent studies have indicated, for instance, that neither for German nor Dutch people up to 30 years, Internet (under)coverage seems to be a problem (CBS 2004, van Eimeren and Frees 2007). However, the studies also show that elderly people and people with a lower education are hard to reach. In Germany, for example, only 25% of elderly people (60+) reported in 2006 that they use the Internet.

2.2.2. Sampling and self-selection error

Another major difficulty is implementing a probability-based web survey in the absence of an adequate sampling frame (Couper 2000). Problems arise particularly when adopting non-probability and self-selection recruitment methods, like in volunteer web surveys[5]. Horvitz and Thompson (1952) have shown that unbiased estimates of population characteristics can be computed only if a real probability sample has been used, every element in the population has a non-zero probability of selection, and all these probabilities are known to the researcher. Furthermore, only under these conditions, the accuracy of estimates can be computed. In non-probability-based web surveys such a selection does not take place. The survey is simply put on the web and respondents are those people who happen to have Internet, visit the website and decide to participate in the survey. At most, one could say that the target population of such a self-selected survey consist of people who have an Internet-connection and have a non-zero probability of visiting the website and participating in

4 If the target population consists of all people with an Internet connection, there is no problem.

5 As indicated in footnote 4 is it can be questioned whether access panels are not also based on pre-recruited self-selection.

the survey. However, this is not a very well defined population and in most cases, it is not the target population the researcher has in mind. Moreover, previous research has shown that people who self-select into a survey 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).[6] An additional problem of this type of web surveys is that due to the fact that all selection probabilities are unknown it is not possible to compute unbiased estimates for whatever target population.

2.2.3. Nonresponse error

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. Nonresponse is a problem in so far as nonrespondents indeed differ in their answers from respondents and their answers. The extent of bias depends on the rate of non-response as well as on differences between respondents and non-respondents on the variables of interest. When the reasons for nonresponse are linked to the research questions the nonresponse error increases with a declining response. Nonresponse bias is not unique to web surveys but as their response rates tend to be lower when compared to other modes (Lynn 2008, Kaplowitz et al. 2004, Shih and Fan 2008), the problem is quite severe. For instance, Lozar Manfreda et al. (2008) found in a meta-analysis examining 45 published and unpublished experimental comparisons between web and other survey modes on average an 11% lower response rate than in case of other modes.[7] Different reasons, such as inefficiency of response-stimulating efforts (incentives, follow-up contacts), technical difficulties (slow, unreliable connections, low-end browsers), personal problems in using a computer, and privacy and confidentiality concerns could be responsible (Bosnjak and Tuten 2003, Dillman and Bowker 2001, Galesic 2006, Göritz 2006, Heerwegh 2005, Heerwegh and Loosveldt 2002, Kaczmirek 2008, Vehovar et al. 2002). Particularly 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.

6 Coverage and sampling is less of a problem where all members of the target population use the Internet and for whom e-mail addresses are known, like in the case of students, employees, members of organizations, customers, et cetera. Here, the existence of a proper sampling frame allows the drawing of a probability-based sample and the generalisation of conclusions to the whole population using standard inference procedures.

7 In this study the main factors of lower response rates were the sampling frame, the solicitation mode, and the number of contacts. In contrast to other meta-analytical findings from traditional mail surveys, no significant influence of incentives was found (Yammarino et al. 1991).

2.3. Can weighting solve the problem? - Overview of recent findings

To reduce the bias resulting from the inferential problems outlined above, the data can be adjusted to correct coverage, sampling and nonresponse errors. Particularly, weighting adjustments have been seen as a possible solution to improve the quality of web surveys (Bethlehem and Stoop 2007, Dever et al. 2008). In this regard, post-stratification weighting has mainly been applied to correct for socio-demographic differences between the (web) sample and the population under consideration. However, as some variables of interest often do not show a sufficiently strong relationship with the demographic weighting variables, it has been emphasised that post-stratification can correct for proportionality but not necessarily for representativeness (Loosveldt and Sonck 2008). For example, weighting does not solve the problem that Internet users and Non-Internet users may differ substantially in some of their attitudes (Schonlau et al. 2004, Bandilla et al. 2003). As a consequence, researchers have argued that this weighting technique seems to have limited potential for correcting biases in web surveys (Lee 2006, Vehovar et al. 1999).

It is due to these difficulties that another weighting technique called *Propensity Score Adjustment (*PSA*)* has been suggested as an alternative for statistically surmounting inherent problems in web survey data (Lee 2006, Loosveldt and Sonck 2008, Rosenbaum and Rubin 1983, 1984, Schonlau et al. 2009, Schonlau et al. 2002, Varedian and Forsman 2003). This statistical technique aims to correct for differences in socio-demographic and 'webographic' (attitudinal or behavioural) variables regarding individuals' decisions to participate in web surveys. For that purpose, however, a probability-based reference survey is needed in which each member of the population has the same probability of selection and which, particularly, contains 'webographic' questions. The volunteer web sample, then, is adjusted to the probability-based reference sample by estimating the probability of each respondent to participate in the web survey.

Although it has been emphasised that for generalising web survey results for the whole population, post-stratification and propensity-based weights are necessary (Duffy et al. 2005), the implications of the different adjustment procedures are still under discussion. Until now their application in scientific surveys[8] has produced rather diverse results, and there is no certainty as to whether the

8 Particularly, commercial market research agencies (like Harris Interactive) have applied this correction technique for their volunteer web surveys.

representativeness of web surveys can be improved (Taylor 2005). In particular, the statistical theory behind PSA and its implications are not well developed and still need to be studied in more detail. As a consequence, the underlying message of most critiques is that no simple weighting factor or adjustment strategy can make on- and offline samples comparable (e.g. Malhotra and Krosnick 2007, Vehovar et al. 1999). However, it should be emphasised that even though inconsistently applied weights can increase the total survey error, a weighting procedure which would statistically allow to generalise to the whole population (including those without Internet access) would be a major breakthrough (Duffy et al. 2005, Lee and Vaillant 2009, Couper et al. 2007).

3. Data and methods

As indicated at the beginning, in order to study efficiency of the different weighting adjustments that are frequently applied to web surveys to correct for selection biases, an empirical comparison is performed between the data obtained from a continuous volunteer web survey (WIS) and such based on probability-based reference surveys (SOEP and OSA). In this context, it has to be underlined that even though both references surveys are representative for the whole population they may also be subject to all sorts of survey errors. However, as these data sets provide the greatest overlap with the WIS data, particularly with respect to the webographic variables, they seem to be the best available reference surveys.

3.1. Databases

The analysis is based on the German (Lohnspiegel) and Dutch (Loonwijzer) data from the *WageIndicator Survey (*WIS*)* which is a continuous volunteer web survey running now in 48 countries. Since 2004, it has collected information on a wide range of subjects including basic demographics, wages and other work-related topics. Most importantly, the data set also includes variables, such as health and job satisfaction, which can be considered as webographic variables. The WIS dataset has been quite successful in gathering large samples (90.000 in the Netherlands and 70.000 in Germany). However, although in most countries the number of observations of the WIS is larger than in national labour force surveys, the samples seem to fail to be representative of the population because of the above-mentioned methodological problems.

As indicated above, in order to apply different weighting techniques a probability-based reference survey is needed. In case of Germany the *Socio-Economic Panel* (SOEP) 2006 serves as a reference survey. It is a wide-ranging representative longitudinal panel study of private households (occupational biographies, employment, earnings as well as health and job satisfaction indicators.) The panel started in 1984. In 2008, nearly 11,000 households and more than 20,000 persons were sampled.[9] In case of the Netherlands, the OSA *Labour Supply Panel* is used as a reference survey. Since 1985, the Netherlands' Organization of Strategic Labour Market Research (OSA) has conducted this biannual

9 For the recent analyses only information on the personal level has been considered.

survey to collect data about the (potential) labour force in the Netherlands. The panel is a face-to-face survey[10] among a representative sample of about 2000 households which are sampled from the total number of households in the Netherlands.[11] Until 2002 the panel targets members of households between 16 and 65 years of age, who are not following daytime education. Since wave 2004 every member of the household between 16 and 67 years of age is asked to participate in the survey, including those who are following daytime education. The survey includes a large variety of information on labour market positions, educational attainment, and family status. Also here attitudes about job and health satisfaction are covered.

In order to use the data sets and compare the analyses between countries, all data sets had to be harmonised. In this context, several problems evolved because of differences in the data sets. First, a direct comparison of wages between the countries becomes difficult because the OSA reference survey only provides the net hourly income, whereas the SOEP reference survey only contains information on the gross monthly income. Second the variables for health and job satisfaction are based on different item scales. In case of the WIS data sets, a 5 item scale is used, while for the SOEP both variables are measured with a 11 item scale. In case of the Dutch reference sample, job satisfaction is measured with a 4 item scale. As respondents are not directly asked about their health satisfaction, the question concerning the general health condition is used which is also measured with a 5 item scale. This seems justified because a person with a good health condition is likely to be more satisfied with it..

Table 1: Differences between the used data sets

Variables	German and Dutch WIS	SOEP	OSA
Income	Both: gross monthly/ hourly and net hourly	Gross monthly	Net hourly
Health satisfaction	1-5 item scale	0-10 item scale	Question is not really asking about the satisfaction with health but relate to the general condition ("hoe is over het algemeen uw gezondheid? 1-5 item scale (from heel goed tot zeer slecht)
Job satisfaction	1-5 item scale	0-10 item scale	Question is focusing on job satisfaction but uses a 1-4 item scale (zeer tevreden- heelemal niet tevreden

10 Since wave 2004 the face-to-face interview was replaced by a - by choice of the respondent - written or web designed questionnaire.

11 Each new wave of the sample is supplemented with new households due to dropout.

3.2. Methods and model selection

As indicated above, having a representative sample of the population is of paramount importance when conducting a survey. The under- or overrepresentation of certain characteristic (such as age, education, gender, etc.) within the collected sample introduces bias and affects the reliability of the results. By comparing the population distribution of a variable with its sample distribution, it can be assessed whether or not the sample is representative for the population with respect to this variable. If the distributions vary considerably, the sample is selective. To correct this, adjustment weights can be computed to restore in the sample the distribution of the selective variable to the same distribution as observed in the population. There are several methods to do this, in the framework of this paper, two methods, post-stratification weighting and Propensity Score Adjustment (PSA) will be described in more detail.

3.2.1. Post-stratification weighting

Post-stratification weighting is one of the common methods, which is considered to adjust the distribution of characteristics in the sample to the target population. The formula for such weights w_i is:

(1) $w_i = p_p / p_s$

Where p_p is the population proportion, and
p_s is the (web)sample proportion.

The formula can be used for univariate adjustments or based on the cell proportions from bi- or multivariate contingence tables for the target population. Post-stratification assigns identical adjustment weights to all elements in the same stratum. In order to calculate post-stratification weights, a reference data set is needed with which the sample data can be compared.

3.2.2. Propensity Score Adjustment (PSA)

Originally developed for the comparison of populations in the context of experimental designs (Rosenbaum and Rubin 1983, 1984) *Propensity Score Adjustment (*PSA*)* has been suggested as an alternative for statistically surmounting inherent problems in web survey data (Loosveldt and Sonck 2008, Schonlau et al. 2009). It aims to correct differences caused by the varying inclinations of individuals to participate in web surveys (Duffy et al. 2005). As already indicated at the beginning, it adjusts for selection bias due to observed covariates which are demographic as well as 'webographic' (lifestyle/attitudinal) variables measuring general attitudes or behaviour that are hypothesised to differ between the web sample and the general population (Schonlau et al. 2007).

To provide a deeper insight into the underlying logic of this method, a propensity score (ps_i) is the conditional probability that a person will be in one condition rather than in another (e.g., 'being in the web or reference survey') given a set of observed covariates used to predict the person's condition (Rosenbaum and Rubin 1983).

(2) $ps_i = P(I_i = 1/X_i)$

Where I_i is an indicator variable for membership in the web survey, and
Xi contains information that is collected in both surveys.

Like all probabilities, a propensity score ranges from 0 to 1. It is a very convenient method as the propensity score is a single number summarising a person's scores on all the observed covariates and weighting the importance of each background characteristic according to its ability to predict treatment assignment (web survey participation). As randomised experiments yield an equal probability assignment mechanism (e.g. a coin toss), each person has a 50% chance of being in treatment. Thus, each person has a true propensity score of 0.50. With a quasi-experiment, the true propensity score function is not known and must be estimated. As the probabilities of receiving treatment (i.e., propensity scores) are a function of individual characteristics, they are likely to vary from 0.50. For instance, if the researcher dummy codes treatment as 1 and control as 0, then a propensity score above 0.50 would mean the person was more likely to select into treatment than control, and a score below 0.50 would indicate the opposite.

Because propensity scores are derived from observed covariates, a crucial step in designing a quasi-experiment is identifying potentially relevant covariates which are expected to affect treatment selection and outcomes. Researchers are often tempted to use only those covariates for which statistically significant differences between treatment and comparison groups are found. Rosenbaum (2002) offered three cautions against this approach: a) the relationship between the covariate and the outcome is not considered and is just as important in many respects; b) statistical significance is not a prerequisite for practical relevance, especially because the former depends heavily on sample size; and c) the covariates are considered in isolation, whereas adjustments consider them collectively. Rubin and Thomas (1996, p. 253) recommended that *"unless a variable can be excluded because there is a consensus that it is unrelated to outcome or is not a proper covariate, it is advisable to include it in the propensity score model even if it is not statistically significant"*. In practice, however, several procedures are used for covariate selection. For example, a number of papers (Berk and Newton 1985, Lieberman et al. 1996) adopted stepwise regression excluding variables that are not significant in explaining the treatment (the significance level for removing a variable is 0.05). Some choose one-step covariate selection based on theoretical and/or logical relevance (Stone et al. 1995, Duncan and Stasny 2001). However, there are no clear-cut criteria for selecting variables for propensity score models.[12]

Against this background, the central question for the selection of covariates for web surveys is which variables capture the difference between the web respondents and the population of interest. Looking at the various applications of PSA for web surveys, all researchers adjust for differences in the distributions of some socio-demographic variables. Schonlau et al. (2004), for instance, found that a minimum set of demographic variables was needed to adjust for selection bias. Additionally they emphasised that self-assessed health status was a useful variable. Varedian and Forsman (2003) also included lifestyle questions that are meant to capture a respondent's "modernity" (such as knowing cosmetic products etc.) besides age, gender and region. Lee (2006) used self-rated social class, employment status, political party affiliation, having a religion and opinion towards ethnic minorities as variables for propensity scoring. However, her result was that this particular set of non-demographic variables makes little difference. Taylor et al. (2001) used in his election study questions which measured alienation, readership, participation and investment. Comparing an online and telephone survey, they found that weighting by propensity scores using these questions did the most to reduce biases efficiently. In an earlier paper about PSA, Taylor (2001) described the use of questions measuring health status, political party identification and the number of telephone lines as effective in reducing the biases in the used online survey.[13]

Also for the application of PSA a probability-based reference survey is needed in which each member of the population has the same probability of selection and which, particularly, contains the required covariates. After merging both samples using variables common to both data sets, an

12 In her simulation study Drake (1993) showed that it is not very serious if the model for propensity score adjustment is miss-specified, for instance, by mistakenly adding a quadratic term or dropping a covariate.

13 Even though Harris Interactive is one of the first companies which published results based on successful applications of PSA (Danielsson 2004), it surprisingly does not provide insights into their underlying research on the use of this method or valuable webographic questions.

indicator variable (I_i) is defined indicating whether the respondent belongs to the web survey or not. The web sample, then, is adjusted to the reference sample by estimating the probability of each respondent to participate in the web survey using the selected set of covariates (X_i). The most commonly used method for computing propensity scores is the logistic regression, with the observed selected covariates as the predictors and the dummy coded treatment assignment as the dependent variable.[14]

(2) $$\log\left(\frac{\tau(X_k)}{1-\tau(X_k)}\right) = \alpha_k + \beta_k^{'} X_k + \varepsilon_k$$

The logistic regression models to compute the different propensity scores applied in this paper have the following formula:

(3) PS1

$$\alpha_k + \beta_1 * women_{2,k} + \beta_2 * education_{2,k} + \beta_3 * cohort_{2,k} + \beta_4 * nonman_{2,k} + \beta_5 * part_{2,k} + \beta_6 * perm_{2,k} + \beta_7 * nojob_{2,k} + \beta_8 * \log income_k + \varepsilon_k$$

(4) PS2

$$\alpha_k + \beta_1 * women_{2,k} + \beta_2 * education_{2,k} + \beta_3 * cohort_{2,k} + \beta_4 * nonman_{2,k} + \beta_5 * part_{2,k} + \beta_6 * perm_{2,k} + \beta_7 * nojob_{2,k} + \beta_8 * healthsat_{2,k} + \beta_9 * \log income_k + \varepsilon_k$$

(5) PS3

$$\alpha_k + \beta_1 * women_{2,k} + \beta_2 * education_{2,k} + \beta_3 * cohort_{2,k} + \beta_4 * nonman_{2,k} + \beta_5 * part_{2,k} + \beta_6 * perm_{2,k} + \beta_7 * nojob_{2,k} + \beta_8 * jobsat_{2,k} + \beta_9 * \log income_k + \varepsilon_k$$

(6) PS4

$$\alpha_k + \beta_1 * women_{2,k} + \beta_2 * education_{2,k} + \beta_3 * cohort_{2,k} + \beta_4 * nonman_{2,k} + \beta_5 * part_{2,k} + \beta_6 * perm_{2,k} + \beta_7 * nojob_{2,k} + \beta_8 * jobsat_{2,k} + \beta_9 * healthsat_{2,k} + \beta_{10} * \log income_k + \varepsilon_k$$

After calculating the propensity scores, the next step is to balance the non-equivalent groups using matching, stratification, covariance adjustment, or weighting on the estimated propensity score. When applying propensity score weighting, weights (w_i^{ps}) are formed as the inverse of the propensity score (see, Lee 2006, Rosenbaum 1987, Schonlau et al. 2004, 2007). Moreover, since the propensity scores refer to both the web and reference survey respondents, the propensity score weights for the

14 Following the basic guidelines provided in Rosenbaum and Rubin (1984), they suggested to construct one model that uses all the predictors for respondents who have completed data. For respondents with missing data, one or more additional models should be constructed in which only variables with complete data are predictors (more than one model if more than one group is identified with different patterns of missing data).

two samples (I^i=0 and I^i=1) are as follows:

$$(7) \qquad w_i^{ps} \begin{cases} 1/\,ps_i & \text{if } I_i = 1 \text{ (web survey)} \\ 1/(1\text{-}ps_i) & \text{if } I_i = 0 \text{ (reference survey)} \end{cases}$$

4. Results

Before weighting techniques can be implemented, it is important to evaluate the bias comparing the German and Dutch web samples with the reference data sets. As the final aim of the applied weights in this paper is to improve wage estimations, specific selections have been applied to the used data sets. All samples have been restricted to employees and persons aged between 16 and 75 living in Germany and the Netherlands. Furthermore, the monthly gross wage has been limited to 400€-10000€. Particularly for PSA it was also necessary to eliminate missing values which finally led to German samples of N=21914 (Lohnspiegel), and N=7993 (SOEP). For the Netherlands, the Loonwijzer sample contains N=8015, and the OSA sample N=2019.

4.1. Selection bias

As indicated in several studies (Loosveldt and Sonck 2008, de Pedraza et al. 2007), also the German Lohnspiegel (LS) and the Dutch Loonwijzer (LW) are affected by typical selection bias (see descriptive, p.26). In Germany and the Netherlands, women, older persons (45-65+), part-timers, persons living in a region with an unemployment rate above the average, and persons who are satisfied with their health and their job are underrepresented in the web sample compared to the reference survey. Moreover, country-specific patterns can also be observed: In Germany highly educated persons as well as persons in manual occupations are underrepresented, whereas in the Netherlands, low and medium educated persons and persons in nonmanual occupations are underrepresented. These differences might be explained due to the different marketing strategies in the two countries. In Germany, for instance, persons with 'lower' education might be attracted by the homepage which is prominently placed on the DGB - a trade union - homepage. However, in this respect, further research is definitely needed to give insight in how far different entrance homepages might already create selection bias because specific people are attracted by different homepages. Nevertheless, the description of the selection bias reveals that in both countries particularly job satisfaction (LS=24.1% and LW=32.2% differences to the reference surveys), part-time work (LS=14.4% and LW=27.7%) and the oldest age cohort (LS=23.5% and LW=20.9%) differ markedly between the web and the reference surveys. Moreover, it is also obvious that the selection bias for most of these variables is somehow stronger in the Netherlands than in Germany.

German Selection Bias

Figure 1: Socio-demographic variables

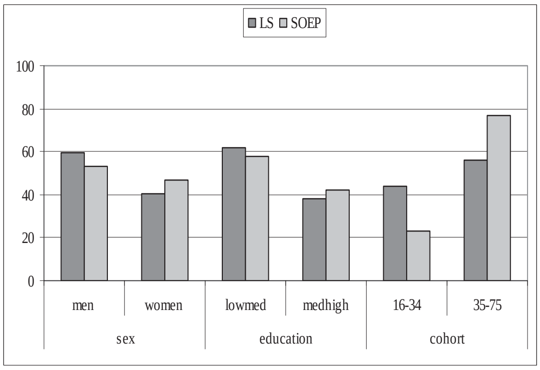


Figure 2: Labour market related variables

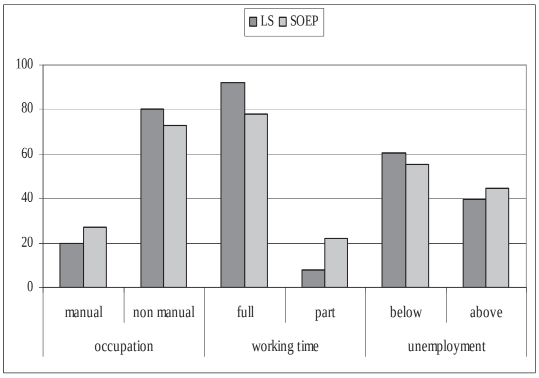
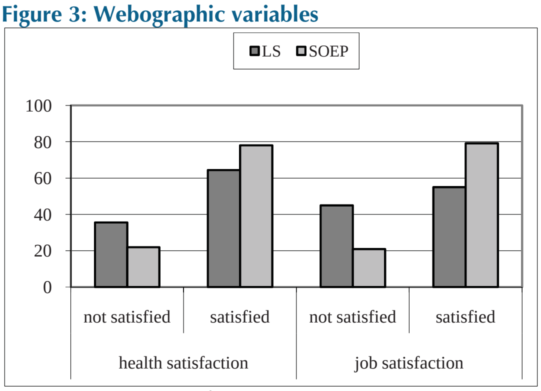


Figure 3: Webographic variables



Source: German LS and SOEP 2006

Dutch Selection Bias

Figure 4: Socio-demographic variables

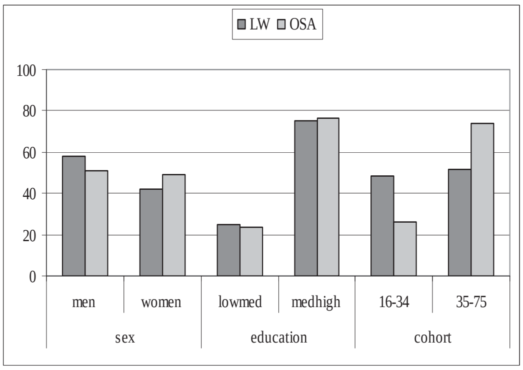


Figure 5: Labour market related variables

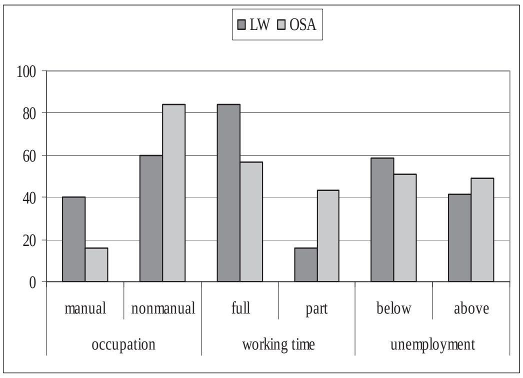
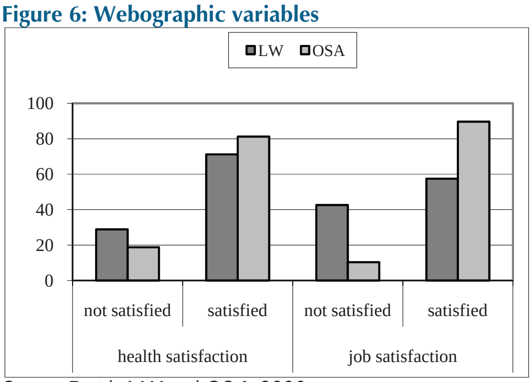


Figure 6: Webographic variables



Source: Dutch LW and OSA 2006

4.2. Correlation analyses

A further step in the analyses of selection bias is to compare correlation matrixes among selected variables between the web and the reference data sets using Pearson's correlation coefficient. As the main purpose of this paper is to evaluate the effectiveness of different weighting techniques for improving wage estimations, the following variables are examined: log wage[15], gender, education, age, nonmanual occupation, part-time, permanent contract, regional unemployment rate, as well as health and job satisfaction. The correlations between these variables are analysed because the comparison between the web and the reference surveys might offer additional insights into the selection bias in the used web surveys.[16]

The results for Germany and the Netherlands (see appendix table A1-A4) show that the basic correlations involving income (logwage) are properly signed. In both countries, income decreases for women, part-timers and for persons living in regions with high unemployment. An increase in income, in contrast, can be observed for higher educated and older persons, for persons working in a nonmanual occupation, for persons with a permanent contract and for persons who are satisfied with their health and their job. With respect to these correlations, no differences can be observed between the Lohnspiegel and the SOEP or between the Loonwijzer and the OSA. This strengthens the argument that the decline of all forms of (non-probability based) web surveys might be exaggerated and that a correction of selection biases, as done in all surveys, might lead to valid and representative results.

4.3. Applying post-stratification weighting and PSA

The selections of the variables which are included in the post-stratification and PSA weights are based on the methodological considerations in section 3.2. and the above-described selection bias. As survey weights, generally, correct for the core demographics (gender, education and age), the first post-stratification weight (W1) simply contains these variables. However, the description has revealed that, additionally, part-timers and persons in nonmanual occupations (Germany), and per-

15 For the comparison of income the logarithmic function of income (logwage), rather than income, is used because income is a nonlinear function with independent variables and has a lognormal shape in most situations.

16 For example, even though the mean of a variable can be biased because of the non randomness of the sample, its product-moment correlation coefficient to a dependent variable could be the same or not so much far off as to the one of the representative sample.

sons in manual occupations (Netherlands) are underrepresented in both web surveys. Therefore, the post-stratification weights W2 and W3 have been constructed. Moreover, as in both surveys a strong selection bias is related to job satisfaction, weights W4 and W5 additionally consider this variable. Finally, weight W6 examines whether a minimum of variables which are mostly affected by selection bias does also serve the purpose. Additionally, for both countries, two country-specific weights have also been constructed (for a detailed description of the different weights, see tables A5 and A6 in the appendix):

General weights
W1= gender (2), education (2) and cohort (2)
W2= gender (2), education (2), cohort (2) and part time (2)
W3= gender (2), education (2), cohort (2) and nonmanual (2)
W4= gender (2), education (2), cohort (2), part time (2) and jobsat
W5= gender (2), education (2), cohort (2), nonmanual (2) and jobsat
W6= part(2) and jobsat(2)

German-specific weights
Wde1= part(2), cohort (2) and jobsat(2)
Wde2= coh(2) and jobsat(2)

Dutch-specific weights
Wnl1= nonmanual(2), part(2) and jobsat(2)
Wnl2= part(2) and nonman(2)

As already indicated, also the propensity score weights have been defined in accordance with the methodology described in section 3.2.. For the logistic regression analysing why people are participating in a web survey, the following variables have been included for the estimation of the propensity scores (see following table 2). The logistic regression coefficients reveal that for Germany, the included covariates in all model specifications have a significant effect on the selective participation in the web survey. In the Netherlands, this holds also except of the variable of education. Furthermore, in the model specification of the first propensity score, the variables permanent contract and log hourly net wage, besides education, have no significant effect on the participation in the web survey (see for more detail tables A7 and A8 in the appendix).

Table 2: Included variables in the PSA models

Variable name	Coding
Gender	Dummy coded (women=1, men=0)
Education	Dummy coded (medium/high=1, low/med=0)
Age cohorts	Dummy coded (16-34=1, 35-65+=0)
Nonmanual occupation	Dummy coded (nonmanual=1, manual=0)
Part-time	Dummy coded (part=1, full=0)
Permanent contract	Dummy coded (perm=1, fixed=0)
Regional Unemployment rate	Dummy coded (above average=1, below average=0)
Health satisfaction	Dummy coded (satisfied=1, unsatisfied=0)
Job satisfaction	Dummy coded (satisfied=1, unsatisfied=0)

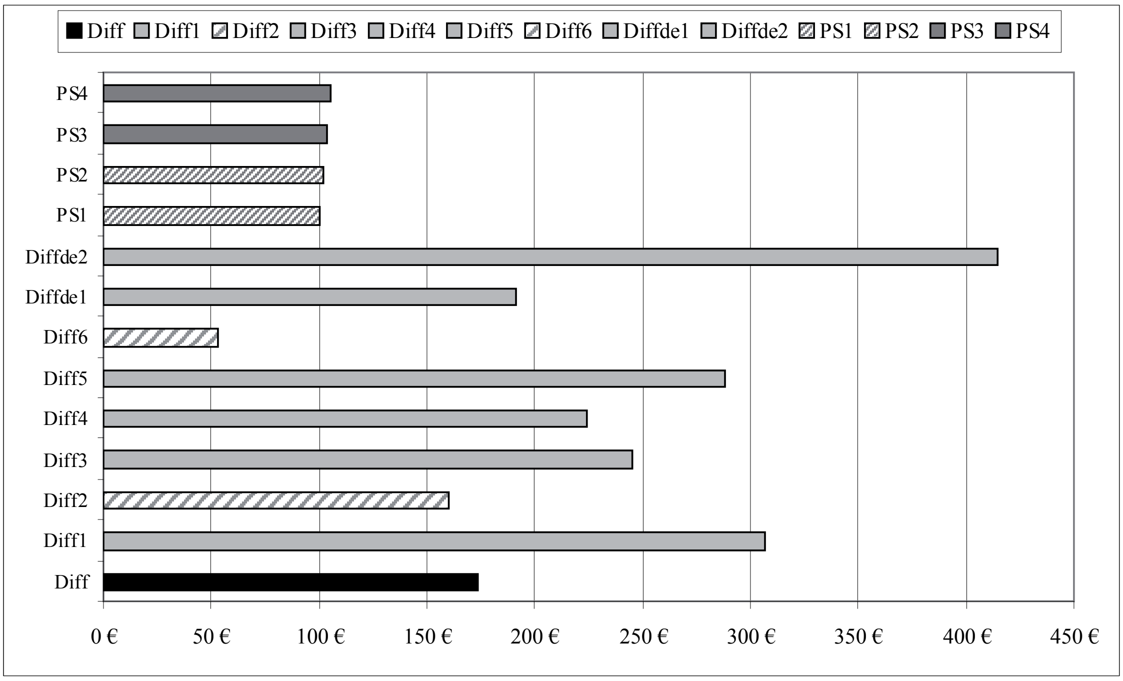
For the creation of propensity score weights also different models have been defined. The first propensity score weight (PS1) captures only the socio-demographic and labour market related variables. However, to test the effect of the 'webographic' variables (health and job satisfaction), the propensity score weights PS2 to PS4 have been defined (details for the different propensity score weights can be found in the appendix table A9).

PS1 = treat women edu2 coh2 nonman part perm nojob logwagemo
PS2 = treat women edu2 coh2 nonman part perm nojob logwagemo + healthsat
PS3 = treat women edu2 coh2 nonman part perm nojob logwagemo + jobsat
PS4 = treat women edu2 coh2 nonman part perm nojob logwagemo + healthsat jobsat

4.3.1. Results for Germany

The following figure 7 describes the differences in the mean wages between the unweighted and weighted Lohnspiegel and the SOEP (see for more detail table A10, appendix).

Figure 7: Differences between gross monthly mean wages, weighted and unweighted LS and SOEP, 2006



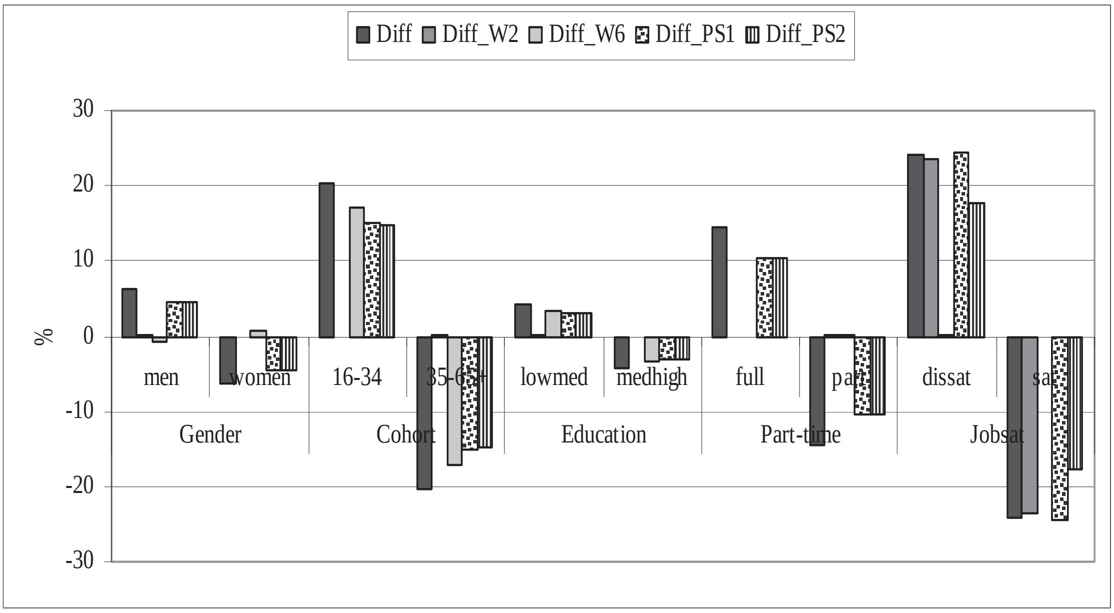
Source: German LS and SOEP 2006, own calculations

Notes: The term Diff refers to the difference between the unweighted LS and SOEP, whereas Diff1-Diffde2 refer to the differences between the unweighted SOEP and the post-stratification weighted LS. PS1-PS4 refer to the difference between the unweighted SOEP and the propensity score adjusted LS.

The black bar (Diff) indicates the differences between the two data sets without weighting. It shows that the mean wage in the Lohnspiegel is around 173€ higher than in the reference survey. With respect to a properly assigned post-stratification and propensity score weight, the expectation would be that this difference between the two data sets diminishes (or, at least, is reduced). The application of the different weights, however, produces rather divergent results. Out of the six post-stratification weights, only W2 (gender, education, cohort and part time) and W6 (part time and jobs satisfaction) are able to adjust the mean income of the web to the reference sample (striped bars) because the difference between the two samples are significantly reduced. With respect to the four defined propensity score weights (PS1-4), the adjustment effect does not differ much. Nevertheless, it seems that PS1 (containing only socio-demographic and labour market related variables) and PS2 (containing socio-demographic, labour market related and the webographic variable of health satisfaction) are more efficient in adjusting the two samples (small striped bars).

In a next step, the aforementioned four 'successful' weights have been implemented for the adjustment of the distribution of selected variables. Looking at the effectiveness of these weights, figure 8 demonstrates differences in the percentage of these variables between the German reference survey (SOEP) and the unweighted (black bar) and weighted web survey (LS). Also here the expectation is that the differences between the two unweighted surveys should diminish with the implementation of the different weights. Starting with the results for the post-stratification weights their application is capable of levelling out the under- and overrepresentation of the variables which have been used for the construction of the weights. For example, for the combined classes of gender, age, education and part-time (W2) or part time and job satisfaction (W6) the differences between the two samples nearly disappear. However, it is also obvious that they are not able to make the web survey and the reference survey respondents comparable, for instance, with regard to their job satisfaction (when applying W2) or their distribution across nonmanual and manual occupations. When turning to the results for the propensity score weights, the results show that the differences between the reference and the web survey became slightly smaller for nearly all variables. Only in case of job satisfaction the results for PS1 (without any webographic variable) indicates that the differences even became larger. However, in comparison to the post-stratification weights, propensity score weights do not totally adjust the two samples with respect to specific variables.

Figure 8: Differences in the effectiveness of weights for selected variables, weighted and unweighted LS and SOEP, 2006



Source: German LS and SOEP 2006, own calculations

Notes: The term Diff refers to the difference between the unweighted LS and SOEP, whereas DiffW2 and DiffW6 refer to the differences between the unweighted SOEP and the post-stratification weighted LS. DiffPS1 and DiffPS2 refer to the difference between the unweighted SOEP and the propensity score adjusted LS.

This rather heterogeneous picture is also mirrored in the results for different wage regressions (monthly log gross wage). The following table 3 presents the unweighted and weigthed regression coefficients for different models including selected explanatory covariates. First of all, it seems important to underline that the findings of the correlation analysis are confirmed when comparing the two unweighted samples (SOEP and LS). Even though the difference in the effects, concerning women and non manual occupations, for instance, is sometimes strong, there is no change in the signs. This supports the argument that data steaming from a continuous volunteer web survey is capable of producing meaningful results. However, as the aim of this paper is to test whether the selected weights are able to improve the representativeness data steaming from a volunteer web survey, for both types of weights, it seems difficult to select a single one which matches best with the results of the regression based on the SOEP data (dark blue).

Table 3: Wage regressions (monthly cross log wage) for the German Lohnspiegel and SOEP, 2006

	SOEP	LS	LS_W2	LS_W6	LS_PS1	LS_PS2
Women	-0.338*** (0.011)	-0.185*** (0.006)	-0.189*** (0.007)	-0.192*** (0.007)	-0.185*** (0.006)	-0.184*** (0.006)
Education	0.321*** (0.010)	0.241*** (0.006)	0.236*** (0.007)	0.240*** (0.008)	0.226*** (0.006)	0.228*** (0.006)
Age cohort	0.295*** (0.012)	0.267*** (0.005)	0.257*** (0.006)	0.255*** (0.007)	0.252*** (0.006)	0.253*** (0.006)
Part-time	-0.602*** (0.013)	-0.601*** (0.010)	-0.622*** (0.012)	-0.602*** (0.013)	-0.638*** (0.012)	-0.637*** (0.012)
Non-manual	0.267*** (0.012)	0.155*** (0.007)	0.177*** (0.008)	0.159*** (0.009)	0.171*** (0.008)	0.168*** (0.008)
Permanent	0.425*** (0.017)	0.231*** (0.008)	0.204*** (0.010)	0.224*** (0.011)	0.225*** (0.009)	0.226*** (0.009)
Unempl.	-0.114*** (0.010)	-0.126*** (0.005)	-0.112*** (0.007)	-0.111*** (0.007)	-0.126*** (0.006)	-0.126*** (0.006)
Jobsat	0.062*** (0.012)	0.084*** (0.005)	0.083*** (0.007)	0.080*** (0.006)	0.084*** (0.006)	0.084*** (0.006)
Healthsat	0.005 (0.012)	-0.002 (0.006)	0.010 (0.007)	0.004 (0.008)	0.005 (0.006)	-0.007 (0.006)
Constant	7.035*** (0.022)	7.374*** (0.010)	7.381*** (0.013)	7.380*** (0.014)	7.359*** (0.012)	7.366*** (0.012)
N	7993	21914	21914	21914	21914	21914

*Source: German LS 2006 (N=21914) and SOEP 2006 (N=7993), * p<0.05, ** p<0.01, *** p<0.001*
Notes: The term LS_W2 and LS_W6 refer to the regression coefficients for the post stratification weighted LS. The term LS_PS1 and LS_P2 indicated the regression results for the application of propensity score weights to the LS

The results show that for different variables, different weights increase the comparability between the SOEP and the Lohnspiegel data. For instance, the variable gender the post-stratification weight (W6) seems the most appropriate, whereas for the variable occupation the weight W2 and in case of health satisfaction the propensity score weight PS1 seem to be the better weighting factors. Finally for some variables, like education and age cohort (light blue) neither post-stratification nor propensity score weights are working.

4.3.2. Results for the Netherlands

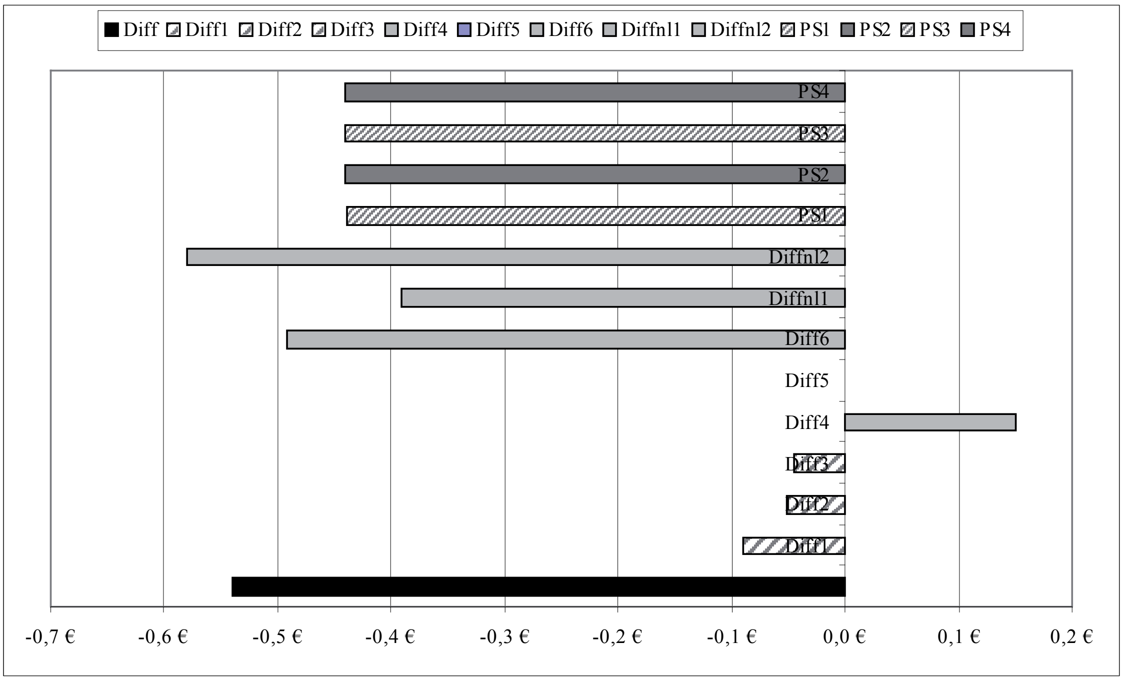
For the Netherlands a somehow different picture emerges. Figure 9 presents the differences in the mean wages between the unweighted and weighted Loonwijzer and OSA survey (for more details, see table A11 in the appendix). Also here the black bar indicates the differences between the two unweighted data sets. For the Netherlands it can be observed that the mean net hourly wage in the Loonwijzer is around 0,5€ lower than in the OSA reference survey. This is a very small difference between the two data sets compared with the bigger income bias in the German case. As above, the implementation of weights should reduce this observed difference between the two data sets.

In this respect, the results for the applied weights are more coherent in case of the Netherlands. For almost all weights, the differences are becoming smaller (except W4 and W2nl)[17] and the adjustment between the two data sets is improved. In this context, particularly the first three post-stratification weights (W1=gender, education, cohort; W2=gender, education, cohort, part; W3=gender, education, cohort, nonman) seem to be effective to adjust the mean income of the web to the reference sample (striped bars).

With respect to the four defined propensity score weights, similar as in the case of Germany the adjustment effect does not differ much between them. Nevertheless, it seems that PS1 (containing only socio-demographic and labour market related variables) and PS3 (containing socio-demographic, labour market related and the webographic variable of job satisfaction) are more successful (small striped bars).

17 The post-stratification weight W5 could not be applied for the Netherlands due to zero cells.

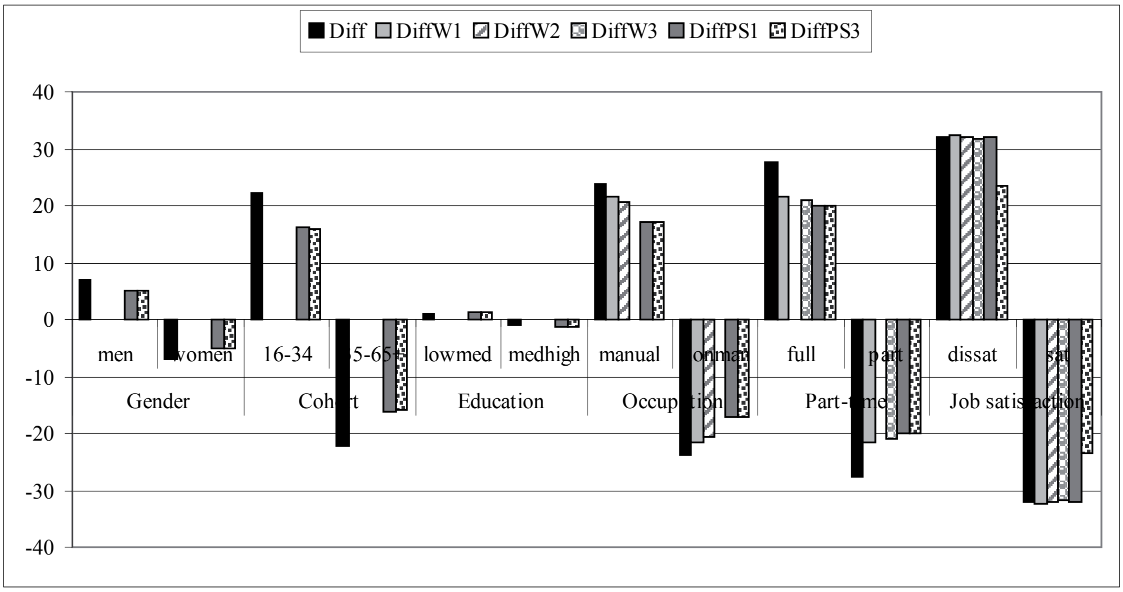
Figure 9: Differences between the net-hourly mean wages, weighted and unweighted LW and OSA, 2006



Source: Dutch LW and OSA 2006, own calculations
Notes: The term Diff refers to the difference between the unweighted LW and OSA, whereas Diff1-Diffnl2 refer to the differences between the unweighted OSA and the post-stratification weighted LW. PS1-PS4 refer to the difference between the unweighted OSA and the propensity score adjusted LW.

Turning to the effectiveness of the above-described 'successful' weights in adjusting the distribution of selected variables, figure 10 demonstrates differences in the percentage of these variables between the Dutch reference survey (OSA) and the unweighted (black bar) and weighted web survey (LW). Also for the Netherlands, the results signal that post-stratification weighting is capable to correct the under- and overrepresentation of the combined classes of gender, age, education (W1), part-time (W2) and non manual occupations (W3). However, they fail with respect to those variables which are not included in the weight (like job satisfaction). When turning to the results for the propensity score weights, similar to Germany, the differences between the reference and the web survey are becoming slightly smaller for all socio-demographic variables. In case of the variable job satisfaction only the propensity score weight PS3 which also includes this variable in its modelling reduces the difference between the to samples significantly. However, also for this example it can be concluded that in comparison with post-stratification weights, propensity score weighting, on average, adjust the distribution of selected variables in two samples only slightly.

Figure 10: Differences between the net-hourly mean wages, weigthed and unweighted LW and OSA, 2006



Source: Dutch LW and OSA 2006, own calculations
Notes: The term Diff refers to the difference between the unweighted LW and OSA, whereas DiffW1-DiffW3 refer to the differences between the unweighted OSA and the post-stratification weighted LW. DiffPS1 and DiffPS3 refer to the difference between the unweighted OSA and the propensity score adjusted LW.

Table 4 presents the unweighted and weighted regression coefficients when implementing these weights to different wage regressions (log net hourly wage). First of all, also here the coefficients for most variables (except health satisfaction) confirm the correlation analysis findings comparing the two unweighted samples (OSA and LW). Although the differences between the effects are sometimes strong, the signs remain the same. This indicates that also in case of the Dutch web sample meaningful results could be produced.

However, similar to Germany, it seems difficult to select a single weight which matches best with the results of the regressions based on the OSA data (dark blue). For different variables, different weights increase the comparability between the Loonwijzer and the OSA data set. For example, for the variable gender the post-stratification weight (W3) seems the most appropriate, whereas for the variable part time the propensity score weight PS3 seem to be the better weighting factors. Moreover, it can be observed that for some variables, like education, age cohort, type of contract and the satisfaction variables (light blue) neither post-stratification nor propensity score weights are working. Given the number of these variables it seems that in comparison to Germany the weights are less effective or even worsen the results for the Netherlands when applied to wage regressions.

Table 4: Wage regressions (hourly log net wage) for the Dutch Loonwijzer and OSA 2006

	OSA	LW	LW_W1	LW_W2	LW_W3	LW_PS1	LW_PS3
Women	-0.186***	-0.114***	-0.126***	-0.139***	-0.146***	-0.120***	-0.119***
	(0.021)	(0.008)	(0.009)	(0.017)	(0.009)	(0.009)	(0.009)
Education	0.277***	0.208***	0.209***	0.206***	0.185***	0.202***	0.207***
	(0.021)	(0.007)	(0.008)	(0.012)	(0.008)	(0.009)	(0.009)
Age cohort	0.264***	0.229***	0.224***	0.203***	0.209***	0.221***	0.222***
	(0.020)	(0.007)	(0.008)	(0.012)	(0.008)	(0.008)	(0.008)
Part-time	-0.595***	-0.012***	-0.029***	-0.013***	-0.013***	-0.020***	-0.030***
	(0.021)	(0.010)	(0.012)	(0.018)	(0.013)	(0.013)	(0.013)
Non-manual	0.256***	0.060***	0.049***	0.066***	0.079***	0.049***	0.059***
	(0.025)	(0.007)	(0.008)	(0.011)	(0.009)	(0.008)	(0.008)
Permanent	0.244***	0.098***	0.090***	0.090***	0.096***	0.096***	0.093***
	(0.025)	(0.009)	(0.011)	(0.015)	(0.011)	(0.010)	(0.011)
Unempl.	-0.006	-0.018**	-0.018*	-0.016	-0.013	-0.018*	-0.021**
	(0.010)	(0.007)	(0.008)	(0.011)	(0.008)	(0.007)	(0.008)
Jobsat	0.032	0.052*	0.050***	0.054***	0.051***	0.050***	0.043***
	(0.027	(0.007)	(0.008)	(0.010	(0.008)	(0.007)	(0.007)
Healthsat	-0.009	0.006	0.010	0.010	0.008	0.008	0.009
	(0.021)	(0.008)	(0.009)	(0.012)	(0.009)	(0.008)	(0.009)
Constant	6.690***	1.959***	1.978***	1.989***	1.982***	1.968***	1.971***
	(0.044)	(0.014)	(0.016)	(0.023)	(0.016)	(0.015)	(0.016)
N	2919	8015	8015	8015	8015	8015	8015

Source: Dutch LW and OSA 2006, own calculations

Notes: The term LW_W1, LW_W2 and LW_W3 refer to the regression coefficients for the post stratification weighted LW. The term LW_PS1 and LW_P3 indicated the regression results for the application of propensity score weights to the LW.

5. Conclusion and discussion

This paper has shown that besides several arguments in favour of web surveys, there are also many disadvantages which might affect the quality and validity of their results. In this context, particularly (non)coverage, self-selection and nonresponse errors have been discussed. As indicated by several authors (Couper et al. 2007, Schonlau et al. 2009) the coverage problem might become less severe because Internet penetration increases steadily. However, a more serious problem is related to the problem of defining a sampling frame from which a probability-based web sample can be drawn.[18]

In this context, the question has been addressed, how results should be treated stemming from volunteer (non-probability based) web surveys? Could they be generalised for the whole population? In how far are weighting techniques capable to help?

In this paper, it could be demonstrated that both selected web samples for Germany and the Netherlands deviated significantly from the reference samples regarding job satisfaction, part time work and age. Smaller selection biases could be found for gender and education. Moreover, country-specific selection bias could be observed with respect to education and nonmanual occupations which might be explained with the different marketing strategies in the countries, and the placement of the web pages.

To correct for these selection biases post-stratification and propensity score weights have been defined. Similar to findings from previous studies (see Lee 2006, Loosveldt and Sonck 2008), the results for post-stratification weights based on different classifications show that, on average, the impact is very limited. However, in this respect country differences could be observed with respect to a comparison of mean wages. Particularly in the Netherlands, post-stratification weights based on very simple models (gender, education, age + part time or nonmanual) are able to adjust the web sample to the reference sample, whereas in Germany this holds only for one weight (part time and job satisfaction). Furthermore, when evaluating the effectiveness of weights to correct for the distributions of different variables of interest, it can be concluded that *'this kind of weighting technique make the proportions*

18 A positive solution has been demonstrated, for instance, by the MESS project of CentERdata (Tilburg), where the Internet is only considered as a 'mode' of data collection and not as the sample frame.

of the variables used comparable, but this does not necessarily make the answers between web respondents and personally interviewed people more comparable with regard to attitude questions' (Loosveldt and Sonck 2008, p. 104).

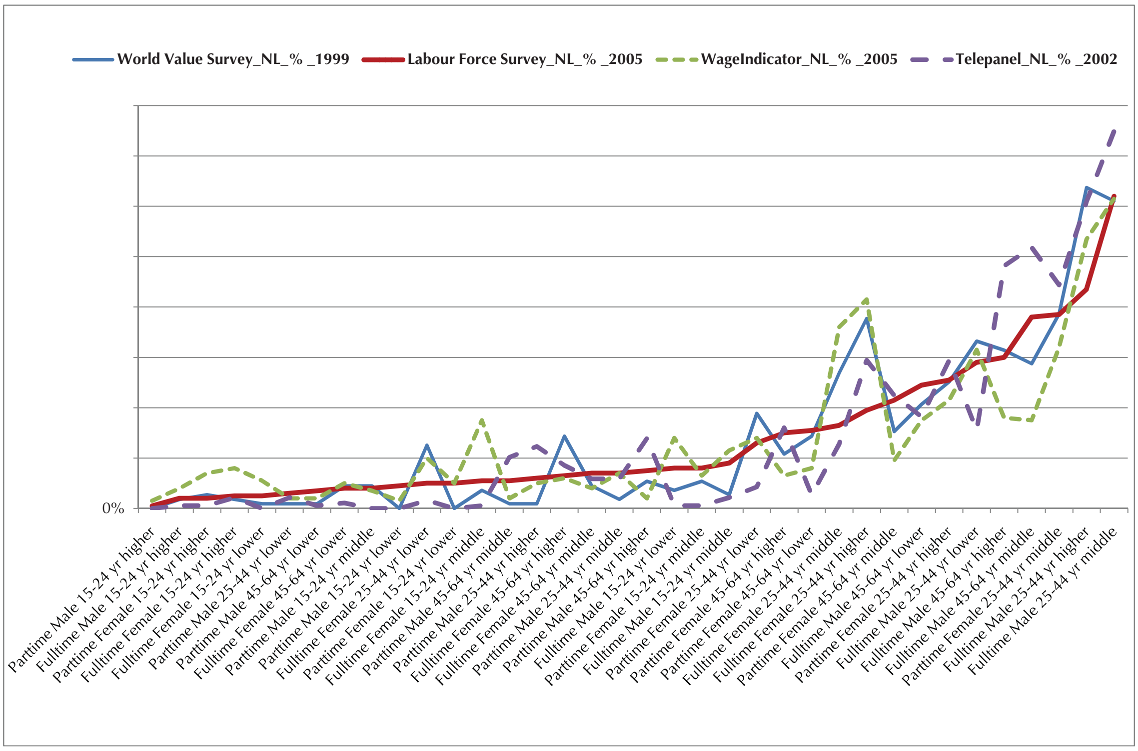
With respect to the findings of propensity score weighting as a possible solution to adjust also for attitudinal questions, in this paper only minimal changes could be observed, particularly for the Dutch sample. Critically, it should be underlined that the difference between the samples with respect to selected variables of interest sometimes even became larger instead of smaller (comparing differences in mean as well as regression coefficients). Moreover, also the inclusion of additional 'webographic' variables did not improve the adjustments considerably.[19]

Against this background it can be summarised that the findings illustrate that the different weighting methods using balancing variables do not make web survey data more comparable to the general population. This holds for the German as well as for the Dutch sample. Moreover, as the unweighted results of the web and reference surveys are quite comparable for both countries (no change in the sign) and none of the applied weights coherently adjusts all coefficients of the web surveys in the appropriate direction it seems to be wiser to use the unweighted web data. In this context, it should be underlined, that even though the applied weighting techniques seem to provide no positive solution with respect to the possibilities to generalise web survey results for the whole population, the collected 'unweighted' web data is not useless. The underlying reasons for the failure of the applied weights could be related to different reasons: for instance, the used reference surveys might be affected by selection bias themselves. Furthermore, it might also be caused by the different mode effects of the web and the reference samples (web questionnaire vs. face to face interviews). With regard to propensity scores it could also be argued that more variables have to be included into the models (problem of unobservables). Moreover, the limited and divergent results for the webographic variables might be related to the fact that the data sets, unfortunately, contain only the two included questions which are not 'classical' webographic variables. To clarify these problems, more analyses and advanced correction techniques are needed.

19 In this respect it should be emphasized that a model has been specified including only the two 'webographic' questions. However, this weight rather increased the difference between the two samples when comparing the mean wages.

Finally, critiques stressing the impossibility to generalise findings of non-probability based web surveys, should reflect on the fact that even probability-based samples might face the problem of self-selection. Persons how are willing to participate in a survey always differ from those who are not participating. This argument can also be supported by the following figure 11 showing the distribution of different variable combinations for selected 'representative' surveys (such as the Dutch Labour Force Survey or the World Value Survey) and the used Dutch WIS data.

Figure 11: Distribution of groups of variables among different surveys



Sources: *Dutch Sample of the World Value Survey 2005, Dutch Labour force survey 2005, Dutch Telepanel 2002 and the Dutch sample of the WageIndicator survey*

Using the Dutch Labour Force Survey (red line) as the 'representative' benchmark, the results show, that not only for the 'unrepresentative' Dutch WIS data (green line) but also for each of the other selected so-called 'representative' surveys selection biases could be observed. In comparison to the other surveys, for most of the variable classifications it would be exaggerated to speak about a fundamental selection bias in case of the WIS data set. In this context, it seems worthwhile to think about the comment of Couper and Miller (2008) to better not treat survey quality as an absolute, but evaluate it relative to other features of the design and the stated goals of the survey.

References

Bandilla, W.; Bosnjak, M.; Altdorfer, P. (2003): Survey administration effects? A comparison of web-based and traditional written self-administered surveys using the ISSP environment module. In: *Social Science Computer Review*, Vol. 21(2), pp. 235-243.

Bandilla, W.; Kaczmirek, L.; Blohm, M.; Neubarth, W. (2009): Coverage und Nonresponse-Effekte bei Online-Bevölkerungsumfragen, pp. 129-143. In: N. Jackob; H. Schoen; T. Zerback (eds.), *Sozialforschung im Internet*: *Methodologie und Praxis der Online-Befragung*. VS Verlag für Sozialwissenschaften.

Berk, R.; Newton, P. (1985): Does arrest really deter wife battery? An effort to replicate the findings of the Minneapolis Spouse Abuse Experiment. In: *American Sociological Review*, Vol. 50, pp. 253-262.

Bethlehem, J.; Stoop, I. (2007): Online panels – a paradigm theft? pp. 113-131. In: M. Trotman et al. (eds.), *The challenges of a changing world*. Southampton: Association for Survey Computing.

Bosnjak, M.; Tuten, T. (2003): Prepaid and promised incentives in web surveys. In: *Social Science Computer Review*, Vol. 21(2), pp. 208-217.

CBS 2004: http://www.cbs.nl/nl-NL/menu/themas/bedrijven/publicaties/artikelen/archief/2004/2004-1506-wm.htm

Couper, M. (2000): Web surveys: a review of issues and approaches. In: *Public Opinion Quarterly*, Vol. 64(4), pp. 464-481.

Couper, M.; Miller, P. (2008): Web survey methods - Introduction. In: *Public Opinion Quarterly*, Vol. 72(5), pp. 831-835.

Couper, M.; Kapteyn, A.; Schonlau, M.; Winter, J. (2007): Noncoverage and nonresponse in an Internet survey. In: *Social Science Research*, Vol. 36(1), pp.131-148.

Danielsson, S. (2004): The propensity score and estimation in nonrandom surveys: an overview. Department of Statistics, University of Linkoeping. Report no. 18 from the project "Modern statistical survey methods". See: http://www.statistics.su.se/modernsurveys/publ/11.pdf

de Pedraza, P.; Tijdens,K.; Muñoz de Bustillo, R. (2007): Simple bias, weights and efficiency of weights in a continuous voluntary web survey. *AIAS Working Paper 07-58*, Amsterdam Institute for Advanced Labour Studies, Amsterdam.

Dever, J.; Rafferty, A.; Valliant, R. (2008): Internet surveys: can statistical adjustments eliminate coverage bias? In: *Survey Research Methods*, Vol. 2(2), pp. 47-62.

Dillman, D.; Bowker, D. (2001): The Web Questionnaire Challenge to Survey Methodologists, pp. 159-178. In: Reips, U.; Bosnjak, M. (eds.): Dimension of Internet Science, Pabst Science Publishers: Lengerich.

Drake, C. (1993): Effects of misspecification of the propensity score on estimators of treatment effect. In: *Biometrics*, Vol. 49(4), pp. 1231-1236.

Duffy, B.; Smith, K.; Terhanian, G.; Bremer, J. (2005): Comparing data from online and face-to-face surveys. In: *International Journal of Market Research*, Vol. 47(6), pp. 615-639.

Duncan, K.; Stasny, E. (2001): Using Propensity Scores to Control Coverage Bias in Telephone Surveys. In: *Survey Methodology*, Vol. 27(2), pp. 121-130.

Fricker, R. (2008): Sampling methods for web and E-mail surveys, pp. 195-216. In: N. Fielding; R. Lee; G. Blank (eds.), *The SAGE Handbook of Online Research Methods*. London: Sage.

Fricker, R.; Schonlau, M. (2002): Advantages and disadvantages of Internet research surveys: Evidence from the literature. In: *Field Methods*, Vol. 14(4), pp. 347-367.

Galesic, M. (2006): Dropouts on the web: effects of Interest and burden experienced during an online survey. *In: Journal of Official Statistics*, Vol. 22(2), pp. 313-328.

Göritz, A. (2006): Cash lotteries as incentives in online panels. In: *Social Science Computer Review*, Vol. 24(4), pp. 445-459.

Groves, R. (2004): Survey Errors and Survey Costs (Wiley Series in Survey Methodology). Wiley - Interscience: Haboken.

Heerwegh, D.; Loosveldt, G. (2002): An evaluation of the effect of response formats on data quality in Web surveys. In: *Social Science Computer Review,* Vol. 20(4), pp.471-484.

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

Haisken-DeNew, J., Frick, J. (Eds.) (2005): Desktop Companion to the German Socio-Economic Panel Study (SOEP), German Institute for Economic Research, Berlin.

Horvitz, D.; Thompson, D. (1952): A generalization of sampling without replacement from finite universe. In: *Journal of the American Statistical Association*, Vol. 47(260), pp. 663-685.

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

Kaplowitz, M.; Hadlock, T.; Levine, R. (2004): A comparison of web and mail survey response rates, In: *Public Opinion Quarterly, Vol. 68*(1), 94-101.

Lee, S.; Vaillant, R. (2009): Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment. In: *Sociological Methods Research,* Vol. 37(3), pp. 319-343.

Lee, S. (2006): Propensity Score Adjustment as a weighting scheme for volunteer panel web surveys. In: *Journal of Official Statistics*, Vol. 22(2), pp. 329-349.

Lieberman, E.; Lang, J.; Cohen, A.; D'Agostino, J.; Datta, S.; Frigoletto, J. (1996): Association of epidural analgesia with caesareans in nulliparous women. In: *Obstetrics and Gynecology*,Vol. 88(6), pp. 993-1000.

Loosveldt, G.; Sonck, N. (2008): An evaluation of the weighting procedures for online access panel surveys. In: *Survey Research Methods*, Vol. 2(2), pp. 93-105.

Lozar Manfreda, K.; Bosnjak, M.; Berzelak, J.; Haas, I.; Vehovar, V. (2008): Web surveys versus other survey modes: A meta-analysis comparing response rates. In: *International Journal of Market Research*, Vol. 50(1), pp. 79-104.

Lynn, P. (2008): The problem of nonresponse, pp. 35-55. In: E. De Leeuw; J. Hox; D. Dillman (eds.), *International Handbook of Survey Methodology*. Lawrence Erlbaum Associates: New York.

Malhotra, N.; Krosnick, J. (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. In: *Political Analysis*, Vol. 15(3), pp. 286-324.

Rosenbaum, P. (1987): Model-based direct adjustment. In: *Journal of the American Statistical Association*, Vol. 82(398), pp. 387-394.

Rosenbaum P. (2002): *Observational Studies*. 2. ed. New York: Springer.

Rosenbaum, P.; Rubin, D. (1983): The central role of the propensity score in observational studies for causal effects. In: *Biometrika*, Vol. 70(1), pp. 41-55.

Rosenbaum, P.; Rubin, D. (1984): Reducing bias in observational studies using sub classification on the propensity score. In: *Journal of the American Statistical Association*, Vol. 79(387), pp. 516-524.

Rubin, D.; Thomas, N. (1996): Matching using estimated propensity scores: relating theory to practice. In: *Biometrics*, Vol. 52(1), pp.249-264.

Schonlau, M.; van Soest, A.; Kapteyn, A.; Couper, M. (2009): Selection bias in web surveys and the use of propensity scores. In: *Sociological Methods Research*, Vol. 37(3), pp. 291-318.

Schonlau, M.; van Soest, A.; Kapteyn, A. (2007): Are ‚Webographic' or attitudinal questions useful for adjusting estimates from web surveys using propensity scoring? In: *Survey Research Methods*, Vol. 1(3), pp. 155-163.

Schonlau, M.; Zapert, K.; Simon L.; Sanstad, K.; Marcus, S.; Adams, J.; Spranca, M.; Kan, H.; Turner, R.; Berry, S. (2004): A comparison between a propensity weighted web survey and an identical RDD survey. In: *Social Science Computer Review*, Vol. 22(1), pp. 128-138.

Schonlau, M.; Fricker, R.; Elliott, M. (2002): *Conducting Research Surveys via E–mail and the Web*. Santa Monica, CA: RAND.

Shih, T.-H.; Fan, X. (2008): Comparing response rates from web and mail surveys: A meta-analysis. In: *Field Methods*, Vol. 20(3), pp. 249-271.

Stone, R.; Obrosky, S.; Singer, D.; Kapoor, W.; Fine, M. (1995): Propensity score adjustment for pretreatment differences between hospitalized and ambulatory patients with community-acquired pneumonia. In: *Medical Care*, Vol. 33(4), pp.AS56-AS66.

Taylor, H. (2005): Does Internet research 'work'? Comparing online survey results with telephone surveys. In: *International Journal of Market Research*, Vol. 42(1), pp. 51-63.

Taylor, H. (2000): Does Internet Research Work? In: *International Journal of Market Research*, Vol. 42(1), pp. 58-63.

Taylor, H.; Bremer, J.; Overmeyer, C.; Siegel, J.; Terhanian, G. (2001): The record of internet-based opinion polls in predicting the results of 72 races in the November 2000 US elections. In: *International Journal of Market Research*, Vol. 43(2), pp.127–135.

Tingling, P.; Parent, M.; Wade, M. (2003): Extending the capabilities of internet-based research: lessons from the field. In: *Internet Research*, Vol. 13(3), pp.223-235.

Tuten, T.; Urban, D.; Bosnjak, M. (2002): Internet surveys and data quality - A review, pp. 7-26. In: B. Batinic, U. Reips and M. Bosnjak (Eds.), *Online Social Science*. Seattle, WA: Hogrefe and Huber Publishers.

Vehovar, V.; Batagelj, Z.; Manfreda, K.; Zaletel; M. (2002): Nonresponse in web surveys, pp. 229-242. In: R. Groves; D. Dillman; J. Eltinge; R. Little (eds.), *Survey Nonresponse*. New York: Wiley.

Vehovar, V.; Batagelj, Z.; Manfreda, K. (1999): Web surveys: can the weighting solve the problem? In: *Proceedings of the Survey Research Methods Section, American Statistical Association (1999)*, pp. 962-967. See: http://www.amstat.org/sections/srms/Proceedings/y1999f.html

Varedian, M.; Forsman, G. (2003): Comparing Propensity Score Weighting with Other Weighting Methods: A Case Study on Web Data. Poster presented at the American Association for Public Opinion Research (AAPOR) 58th Annual Conference, 2003

See: http://www.websm.org/uploadi/editor/Varedian_Forsman_2002_propensity_score.pdf

Yammarino, F.; Skinner, S.; Childers, T. (1991): Understanding mail survey response behavior: a meta-analysis. In: *Public Opinion Quarterly*, Vol. 55(4), pp.613-639.

Appendix

Table A1: Correlations for the German Lohnspiegel 2006

	Wage(m)	Women	Age	Edu	Nonman	Part	Perm	Unempl.	Healthsat	Jobsat
Wage(m)	1									
Women	-0,2798	1								
	0									
Age	0,317	-1,1374	1							
	0	0								
Education	0,3154	0,0192	-0,0819	1						
	0	0,0045	0							
Nonman	0,1484	0,2246	-0,0388	0,3	1					
	0	0	0	0						
Part	-0,2766	0,2849	0,0865	-0,0245	0,0715	1				
	0	0	0	0,0003	0					
Perm	0,1872	-0,0615	0,1613	-0,0392	-0,0055	-0,0415	1			
	0	0	0	0	0,4171	0				
Unempl	-0,1262	0,0123	0,0363	0,0092	-0,0716	-0,0061	-0,0245	1		
	0	0,0677	0	0,1714	0	0,3704	0,0003			
Healthsat	0,0255	-0,006	-0,0923	0,0828	0,0441	-0,0086	-0,0098	-0,0246	1	
	0,0002	0,3754	0	0	0	0,2056	0,1487	0,0003		
Jobsat	0,1375	-0,0421	0,0321	0,0694	0,0585	-0,0053	0,0272	-0,0281	0,2116	1
	0	0	0	0	0	0,4347	0,0001	0	0	

Source: *German LS, 2006*

Note: *Light grey= sign is the same as in the reference survey but the correlation is stronger or weaker, dark grey= sign is different from the reference survey*

Table A2: Correlations for the German SOEP 2006

	Wage(m)	Women	Age	Edu	Nonman	Part	Perm	Unempl.	Healthsat	Jobsat
Wage(m)	1									
Women	-0,3887	1								
	0									
Age	0,2678	-0,0257	1							
	0	0,0215								
Education	0,3944	-0,0305	0,1779	1						
	0	0,0064	0							
Nonman	0,1763	0,3447	0,0359	0,3108	1					
	0	0	0,0013	0						
Part	-0,3865	0,4667	0,09	-0,0138	0,1715	1				
	0	0	0	0,2188	0					
Perm	0,2164	-0,0283	0,3403	0,1171	-0,0068	0,007	1			
	0	0,0113	0	0	0,5425	0,5323				
Unempl	-0,0779	0,0251	0,0426	0,066	-0,0032	-0,0476	-0,0271	1		
	0	0,0249	0,0001	0	0,7778	0	0,0154			
Healthsat	0,0291	-0,0175	-0,1391	0,0326	0,0359	-0,018	-0,0412	-0,0379	1	
	0,0093	0,1182	0	0,0036	0,0013	0,1081	0,0002	0,0007		
Jobsat	0,0814	-0,0174	-0,0217	0,0589	0,0385	0,0079	0,0032	-0,0081	0,3305	1
	0	0,1204	0,0524	0	0,0006	0,4809	0,7746	0,4674	0	

Source: *German SOEP, 2006*

Table A3: Correlations for the Dutch Loohnwijzer 2006

	Wage(m)	Women	Age	Edu	Nonman	Part	Perm	Unempl.	Healthsat	Jobsat
Wage(m)	1									
Women	-0,1888	1								
	0									
Age	0,3071	-0,1743	1							
	0	0								
Education	0,3285	-0,0153	-0,0654	1						
	0	0,1701	0							
Nonman	0,0361	0,2519	0,0365	0,2194	1					
	0,0012	0	0,0011	0						
Part	-0,0469	0,4131	0,0787	-0,0941	0,1119	1				
	0	0	0	0	0					
Perm	0,1408	-0,1504	0,2473	-0,0287	-0,045	-0,0617	1			
	0	0	0	0,0102	0,0001	0				
Unempl	-0,0311	-0,0045	-0,0032	-0,0393	-0,0187	0,001	-0,0068	1		
	0,0053	0,6867	0,7713	0,0004	0,0941	0,9284	0,5445			
Healthsat	0,0099	-0,0008	-0,0537	0,056	0,022	-0,0033	-0,0145	-0,0039	1	
	0,3778	0,9439	0	0	0,0485	0,7681	0,1931	0,7271		
Jobsat	0,0867	-0,024	0,0147	0,0621	0,0319	-0,0086	-0,0035	0,0092	0,2009	1
	0	0,0316	0,1879	0	0,0043	0,4397	0,7519	0,409	0	

Source: *Dutch LW, 2006*

Note: *Light grey= sign is the same as in the reference survey but the correlation is stronger or weaker, dark grey= sign is different from the reference survey*

Table A4: Correlations for the Dutch OSA 2006

	Wage(m)	Women	Age	Edu	Nonman	Part	Perm	Unempl.	Healthsat	Jobsat
Wage(m)	1									
Women	-0,134	1								
	0									
Age	0,2571	-0,1046	1							
	0	0								
Education	0,3359	-0,0065	0,0082	1						
	0	0,7247	0,6593							
Nonman	0,1866	0,2405	0,031	0,3123	1					
	0	0	0,0937	0						
Part	-0,0481	0,6025	-0,0266	-0,028	0,1468	1				
	0,0093	0	0,1506	0,131	0					
Perm	0,1396	-0,0509	0,3544	0,0276	0,0355	-0,0518	1			
	0	0,0059	0	0,1359	0,0552	0,0051				
Unempl	-0,028	-0,0349	-0,019	-0,0571	-0,0339	-0,0456	-0,0019	1		
	0,1298	0,0591	0,305	0,002	0,0667	0,0137	0,9201			
Healthsat	0,0321	-0,0415	-0,0971	0,0603	0,0124	-0,0584	-0,0082	-0,017	1	
	0,0829	0,0249	0	0,0011	0,5037	0,0016	0,6592	0,3596		
Jobsat	0,0166	0,0247	0,042	-0,011	0,0032	-0,0097	0,0097	-0,006	0,119	1
	0,3697	0,182	0,0232	0,5523	0,8608	0,5988	0,599	0,7453	0	

Source: *Dutch OSA, 2006*

Table A5: Description of weights for the German Lohnspiegel, 2006

	Obs	Mean	Std. Dev.	Min	Max
W1	21914	1	.5070705	.4165252	2.083167
W2	21914	1	.6043845	.2030851	3.196147
W3	21914	1	.6927667	.3927108	4.176672
W4	21914	1	.9065323	.1589924	6.523234
W5	21914	1	.8327143	.0415401	4.797885
W6	21914	1	.77857	.3959387	4.195554
Wde1	21914	1	.8686594	.2019751	4.676013
Wde2	21914	1	.6505067	.2230494	1.913719

Source: German LS, 2006

Table A6: Description of weights for the Dutch Loonwijzer, 2006

	Obs	Mean	Std. Dev.	Min	Max
W1	8015	1	.5325757	.4771396	2.115901
W2	8015	1	.9208551	.2277625	6.205515
W3	8015	1	.7655367	.0782483	2.685373
W4	8015	1	1.268722	.0670235	5.376544
W6	8015	1	.9049651	.3328539	3.380586
Wnl1	8015	1	1.120568	.1596509	4.32407
Wnl2	8015	1	1.266734	.0251909	8.070998

Source: Dutch LW, 2006

Table A7: Logistic regression results for the calculation of PS scores for the combined German sample, 2006

	PS_1	PS_2	PS_3c	PS_4
Women	-0.064*	-0.074*	-0.085*	-0.090**
	(0.032)	(0.033)	(0.033)	(0.033)
Education	-0.507***	-0.475***	-0.493***	-0.473***
	(0.031)	(0.031)	(0.032)	(0.032)
Agecohort	-0.942***	-1.008***	-0.970***	-1.009***
	(0.033)	(0.033)	(0.034)	(0.034)
Part-time	-0.815***	-0.801***	-0.728***	-0.727***
	(0.047)	(0.047)	(0.048)	(0.048)
Non-manual	0.589***	0.613***	0.636***	0.646***
	(0.036)	(0.036)	(0.037)	(0.037)
Permanent	-0.417***	-0.422***	-0.414***	-0.417***
	(0.046)	(0.047)	(0.048)	(0.048)
Unempl.	-0.108***	-0.124***	-0.121***	-0.132***
	(0.028)	(0.028)	(0.029)	(0.029)
Logwage.	0.487***	0.508***	0.605***	0.608***
	(0.034)	(0.034)	(0.035)	(0.035)
Healthsat		-0.783***		-0.530***
		(0.032)		(0.033)
Jobsat			-1.196***	-1.076***
			(0.032)	(0.033)
Constant	-1.828***	-1.409***	-1.964***	-1.662***
	(0.251)	(0.254)	(0.256)	(0.258)
N	29907	29907	29907	29907

Source: German LS+SOEP 2006 (N=29907), * p<0.05, ** p<0.01, *** p<0.001

Table A8: Logistic regression results for the calculation of PS scores for the combined Dutch sample, 2006

	PS_1	PS_2	PS_3	PS_4
Women	0.613***	0.605***	0.611***	0.604***
	(0.061)	(0.061)	(0.064)	(0.064)
Education	-0.011	0.002	-0.027	-0.021
	(0.060)	(0.060)	(0.063)	(0.063)
Age cohort	-0.828***	-0.861***	-0.862***	-0.880***
	(0.056)	(0.056)	(0.059)	(0.059)
Part-time	-1.528***	-1.532***	-1.537***	-1.539***
	(0.063)	(0.063)	(0.066)	(0.066)
Non-manual	-1.225***	-1.216***	-1.213***	-1.209***
	(0.061)	(0.061)	(0.063)	(0.063)
Permanent	-0.132	-0.137*	-0.142*	-0.147*
	(0.069)	(0.069)	(0.072)	(0.072)
Unempl.	-0.332***	-0.337***	-0.330***	-0.333***
	(0.047)	(0.048)	(0.050)	(0.050)
Logwage.	0.153	0.175*	0.321***	0.330***
	(0.078)	(0.079)	(0.082)	(0.082)
Healthsat		-0.625***		-0.340***
		(0.058)		(0.062)
Jobsat			-1.858***	-1.803***
			(0.069)	(0.069)
Constant	2.495***	2.936***	3.551***	3.763***
	(0.180)	(0.186)	(0.194)	(0.199)
N	10934	10934	10934	10934

*Source: Dutch LW+OSA 2006 (N=10934), * p<0.05, ** p<0.01, *** p<0.001*

Table A9: Description of PS-Scores for Germany (DE) and the Netherlands (NL)

	Obs	Mean	Std. Dev.	Min	Max
PS1 (DE)	21914	1.365924	.3053881	1.044105	6.157865
PS2 (DE)	21914	1.366487	.3431049	1.02589	6.879322
PS3 (DE)	21914	1.366111	.3952585	1.016708	6.881299
PS4 (DE)	21914	1.364927	.406217	1.013274	7.492218
PS1 (NL)	8015	1.367531	.4139513	1.025386	4.734239
PS2 (NL)	8015	1.36901	.4390628	1.014931	5.04304
PS3 (NL)	8015	1.369435	.5596247	1.00539	6.673388
PS4 (NL)	8015	1.370079	.5680195	1.004527	7.09643

Table A10: Mean comparison of gross monthly income for the unweighted and weighted German Lohnspiegel (LS) data set

	SOEP	LS	LS_W1	LS_W2	LS_W3	LS_W4	LS_W5	LS_W6	LS_ Wde1	LS_ Wde2
Gr.w. (m)	2594,8	2768,1	2901,6	2754,6	2839,7	2819,1	2883,3	2648,2	2786,2	3009,6

	SOEP	LS	PS1	PS2	PS3	PS4
Gr.w. (m)	2594,8	2768,1	2695,2	2696,5	2698,7	2699,7

Source: German LS and SOEP, 2006

Table A11: Mean comparison of net hourly income for the unweighted and weighted Dutch Loonwijzer (LW) data set

	OSA	LW	LW_W1	LW_W2	LW_W3	LW_W4	LW_W5	LW_W6	LW_Wnl1	LW_Wnl2
N. w. (h)	11,46	10,92	11,37	11,40	11,41	11,61		10,96	11,06	10,88

	OSA	LW	PS1	PS2	PS3	PS4
N. w. (h)	11,46	10,92	11,0167	11,0150	11,0151	11,0148

Source: Dutch LW and OSA, 2006

AIAS Working Papers

Recent publications of the Amsterdam Institute for Advanced Labour Studies. They can be downloaded from our website www.uva-aias.net under the subject Publications.

09-75 Welfare reform in the UK, the Netherlands, and Finland. Change within the limits of path dependence.
2009 - Minna van Gerven

09-74 Flexibility and security: an asymmetrical relationship? The uncertain relevance of flexicurity policies for segmented labour markets and residual welfare regimes
2009 - Aliki Mouriki (guest at AIAS from October 2008 - March 2009)

09-73 Education, Inequality, and Active Citizenship Tensions in a Differentiated Schooling System
2009 - Herman van de Werfhorst

09-72 An analysis of firm support for active labor market policies in Denmark, Germany, and the Netherlands
2009 - Moira Nelson

08-71 The Dutch minimum wage radical reduction shifts main focus to part-time jobs
2008 - Wiemer Salverda

08-70 Parallelle innovatie als een vorm van beleidsleren: Het voorbeeld van de keten van werk en inkomen
2008 - Marc van der Meer, Bert Roes

08-69 Balancing roles - bridging the divide between HRM, employee participation and learning in the Dutch knowledge economy
2008 - Marc van der Meer, Wout Buitelaar

08-68 From policy to practice: Assessing sectoral flexicurity in the Netherlands
October 2008 - Hesther Houwing / Trudie Schils

08-67 The first part-time economy in the world. Does it work?
Republication August 2008 - Jelle Visser

08-66 Gender equality in the Netherlands: an example of Europeanisation of social law and policy
May 2008 - Nuria E.Ramos-Martin

07-65 Activating social policy and the preventive approach for the unemployed in the Netherlands
January 2008 - Minna van Gerven

07-64 Struggling for a proper job: Recent immigrants in the Netherlands
January 2008 - Aslan Zorlu

07-63 Marktwerking en arbeidsvoorwaarden – de casus van het openbaar vervoer, de energiebedrijven en de thuiszorg
July 2007 - Marc van der Meer, Marian Schaapman & Monique Aerts

07-62 Vocational education and active citizenship behaviour in cross-national perspective
November 2007 - Herman G. van der Werfhorst

07-61 The state in industrial relations: The politics of the minimum wage in Turkey and the USA
November 2007 - Ruÿa Gökhan Koçer & Jelle Visser

07-60 Sample bias, weights and efficiency of weights in a continuous web voluntary survey
September 2007 - Pablo de Pedraza, Kea Tijdens & Rafael Muñoz de Bustillo

07-59 Globalization and working time: Work-Place hours and flexibility in Germany
October 2007 - Brian Burgoon & Damian Raess

07-58 Determinants of subjective job insecurity in 5 European countries
August 2007 - Rafael Muñoz de Bustillo & Pablo de Pedraza

07-57 Does it matter who takes responsibility?
May 2007 - Paul de Beer & Trudie Schils

07-56 Employement protection in dutch collective labour agreements
April 2007 - Trudie Schils

07-54 Temporary agency work in the Netherlands
February 2007 - Kea Tijdens, Maarten van Klaveren, Hester Houwing, Marc van der Meer & Marieke van Essen

07-53 Distribution of responsibility for social security and labour market policy – Country report: Belgium
January 2007 - Johan de Deken

07-52 Distribution of responsibility for social security and labour market policy – Country report: Germany
January 2007 - Bernard Ebbinghaus & Werner Eichhorst

07-51 Distribution of responsibility for social security and labour market policy – Country report: Denmark
January 2007 - Per Kongshøj Madsen

07-50 Distribution of responsibility for social security and labour market policy – Country report: The United Kingdom
January 2007 - Jochen Clasen

07-49 Distribution of responsibility for social security and labour market policy – Country report: The Netherlands
January 2007 - Trudie Schils

06-48 Population ageing in the Netherlands: demographic and financial arguments for a balanced approach
January 2007 - Wiemer Salverda

06-47 The effects of social and political openness on the welfare state in 18 OECD countries, 1970-2000
January 2007 - Ferry Koster

06-46 Low Pay Incidence and Mobility in the Netherlands- Exploring the Role of Personal, Job and Employer Characteristics
October 2006 - Maite Blázques Cuesta & Wiemer Salverda

06-45 Diversity in Work: The Heterogeneity of Women's Labour Market Participation Patterns"
September 2006 - Mara Yerkes

06-44 Early retirement patterns in Germany, the Netherlands and the United Kingdom
October 2006 - Trudie Schils

06-43 Women's working preferences in the Netherlands, Germany and the UK
August 2006 - Mara Yerkes

05-42 Wage Bargaining Institutions in Europe: a happy Marriage or preparing for Divorce?
December 2005 - Jelle Visser

05-41 The Work-Family Balance on the Union's Agenda
December 2005 - Kilian Schreuder

05-40 Boxing and Dancing: Dutch Trade Union and Works Council Experiences Revisited
November 2005 - Maarten van Klaveren & Wim Sprenger

05-39 Analysing employment practices in Western European Multinationals: coordination, industrial relations and employment flexibility in Poland
October 2005 - Marta Kahancova & Marc van der Meer

05-38 Income distribution in the Netherlands in the 20th century: long-run developments and cyclical properties
September 2005 - Emiel Afman

05-37 Search, Mismatch and Unemployment
July 2005 - Maite Blazques & Marcel Jansen

05-36 Women's Preferences or Delineated Policies? The development of part-time work in the Netherlands, Germany and the United Kingdom
July 2005 - Mara Yerkes & Jelle Visser

05-35 Vissen in een vreemde vijver: Het werven van verpleegkundigen en verzorgenden in het buitenland
May 2005 - Judith Roosblad

05-34 Female part-time employment in the Netherlands and Spain: an analysis of the reasons for taking a part-time job and of the major sectors in which these jobs are performed
May 2005 - Elena Sirvent Garcia del Valle

05-33 Een Functie met Inhoud 2004 - Een enquête naar de taakinhoud van secretaressen 2004, 2000, 1994
April 2005 - Kea Tijdens

04-32 Tax evasive behavior and gender in a transition country
November 2004 - Klarita Gërxhani

04-31 How many hours do you usually work? An analysis of the working hours questions in 17 large-scale surveys in 7 countries
November 2004 - Kea Tijdens

04-30 Why do people work overtime hours? Paid and unpaid overtime working in the Netherlands
August 2004 - Kea Tijdens

04-29 Overcoming Marginalisation? Gender and Ethnic Segregation in the Dutch Construction, Health, IT and Printing Industries
July 2004 - Marc van der Meer

04-28 The Work-Family Balance in Collective agreements. More Female employees, More Provisions?
July 2004 - Killian Schreuder

04-27 Female Income, the Ego Effect and the Divorce Decision: Evidence from Micro Data
March 2004 - Randy Kesselring (Professor of Economics at Arkansas State University, USA) was guest at AIAS in April and May 2003

04-26 Economische effecten van Immigratie – Ontwikkeling van een Databestand en eerste analyses
Januari 2004 - Joop Hartog & Aslan Zorlu

03-25 WageIndicator – Dataset Loonwijzer
Januari 2004 - Kea Tijdens

03-24 Codeboek DUCADAM Dataset
December 2003 - Kilian Schreuder & Kea Tijdens

03-23 Household Consumption and Savings Around the Time of Births and the Role of Education
December 2003 - Adriaan S. Kalwij

03-22 A panel data analysis of the effects of wages, standard hours and unionisation on paid overtime work in Britain
October 2003 - Adriaan S. Kalwij

03-21 A Two-Step First-Difference Estimator for a Panel Data Tobit Model
December 2003 - Adriaan S. Kalwij

03-20 Individuals' Unemployment Durations over the Business Cycle
June 2003 - Adriaan Kalwei

03-19 Een onderzoek naar CAO-afspraken op basis van de FNV cao-databank en de AWVN-database
December 2003 - Kea Tijdens & Maarten van Klaveren

03-18 Permanent and Transitory Wage Inequality of British Men, 1975-2001: Year, Age and Cohort Effects
October 2003 - Adriaan S. Kalwij & Rob Alessie

03-17 Working Women's Choices for Domestic Help
October 2003 - Kea Tijdens, Tanja van der Lippe & Esther de Ruijter

03-16 De invloed van de Wet arbeid en zorg op verlofregelingen in CAO's
October 2003 - Marieke van Essen

03-15 Flexibility and Social Protection
August 2003 - Ton Wilthagen

03-14 Top Incomes in the Netherlands and The United Kingdom over the Twentieth Century
September 2003 - A.B.Atkinson & dr. W. Salverda

03-13 Tax Evasion in Albania: an Institutional Vacuum
April 2003 - Klarita Gërxhani

03-12 Politico-Economic Institutions and the Informal Sector in Albania
May 2003 - Klarita Gërxhani

03-11 Tax Evasion and the Source of Income: An experimental study in Albania and the Netherlands
May 2003 - Klarita Gërxhani

03-10 Chances and limitations of "benchmarking" in the reform of welfare state structures - the case of pension policy
May 2003 - Martin Schludi

03-09 Dealing with the "flexibility-security-nexus: Institutions, strategies, opportunities and barriers
May 2003 - Ton Wilthagen & Frank Tros

03-08 Tax Evasion in Transition: Outcome of an Institutional Clash -Testing Feige's Conjecture
March 2003 - Klarita Gërxhani

03-07 Teleworking Policies of Organisations- The Dutch Experiencee
February 2003 - Kea Tijdens & Maarten van Klaveren

03-06 Flexible Work- Arrangements and the Quality of Life
February 2003 - Cees Nierop

01-05 Employer's and employees' preferences for working time reduction and working time differentiation – A study of the 36 hours working week in the Dutch banking industry
2001 - Kea Tijdens

01-04 Pattern Persistence in Europan Trade Union Density
October 2001 - Danielle Checchi & Jelle Visser

01-03 Negotiated flexibility in working time and labour market transitions – The case of the Netherlands
2001 - Jelle Visser

01-02 Substitution or Segregation: Explaining the Gender Composition in Dutch Manufacturing Industry 1899 – 1998
June 2001 - Maarten van Klaveren & Kea Tijdens

00-01 The first part-time economy in the world. Does it work?
June 2000 - Jelle Visser

Information about AIAS

AIAS is an interdisciplinary institute, established in 1998, aiming to become the leading expert centre in the Netherlands for research on industrial relations, organisation of work, wage formation and labour market inequalities. As a network organisation, AIAS brings together high-level expertise at the University of Amsterdam from five disciplines:

- Law
- Economics
- Sociology
- Psychology
- Health and safety studies

AIAS provides both teaching and research. On the teaching side it offers a Masters in Comparative Labour and Organisation Studies and one in Human Resource Management. In addition, it organizes special courses in co-operation with other organisations such as the Netherlands Centre for Social Innovation (NCSI), the Netherlands Institute for Small and Medium-sized Companies (MKB-Nederland), the National Centre for Industrial Relations 'De Burcht', the National Institute for Co-determination (GBIO), and the Netherlands Institute of International Relations 'Clingendael'. AIAS has an extensive research programme (2004-2008) on Institutions, Inequalities and Internationalisation, building on the research performed by its member scholars. Current research themes effectively include:

- Wage formation, social policy and industrial relations
- The cycles of policy learning and mimicking in labour market reforms in Europe
- The distribution of responsibility between the state and the market in social security
- The wage-indicator and world-wide comparison of employment conditions
- The projects of the LoWER network

Amsterdam Institute for Advanced labour Studies

University of Amsterdam

Plantage Muidergracht 12 • 1018 TV Amsterdam • The Netherlands

Tel +31 20 525 4199 • Fax +31 20 525 4301

aias@uva.nl • www.uva-aias.net