Survey Research Methods (2012)
Vol.6, No.3, pp. 145–154
ISSN 1864-3361
http://www.surveymethods.org

Handling Nonresponse in Business Surveys

Riccardo Borgoni, Donata Marasini and Piero Quatto
University of Milan-Bicocca

Business surveys are a valuable indication of the current and the future economic situation. However refusals are very common in this context and may induce bias in the estimates of interest. In this paper the problem of adjusting for nonresponse estimators of relevant economic outcomes is considered. Using a large simulation study, we investigate how weighting adjusted procedures are influenced by the specification of the response model utilised in calculation of the weights. We consider, in particular, how trimming the weights can have an impact on the final estimates and we propose a bootstrap-based procedure to determine an optimal trimming threshold. We illustrate the procedure using a 2009 survey of a large population of Italian firms.
Keywords: regression estimator, nonresponse weighting adjustment, trimming, bootstrap, business surveys

1 Introduction

Business surveys are a valuable indication of the current and future economic situation. Information on business expectations, labour, confidence, profitability and sales, are nowadays collected in various sampling campaigns conducted by public and private institutes in many countries. Although response behaviour in business surveys has been rarely examined in the literature, refusals are very common in this context, and may induce bias in the estimates of interest.

Nonresponse weighting adjustment (NWA) has been proposed as a method to reduce nonresponse bias in the presence of unit nonresponse when the missing mechanism is at random. The main idea of this method is to weight respondents by the inverse of their response probability (Groves et al. 2002). The respondent sample is considered to be the result of a two-phase sampling procedure. In the first phase, a sample of units is drawn from the target population according to a predefined sample design. In the second, a sample of respondents is further obtained by a Poisson design where sampling units are selected according to the probability that a respondent is in the sample. Since the true response probability is usually unknown, it has to be estimated from the data. If auxiliary variables are available throughout the target population, one can take advantage of this by estimating an appropriate generalised linear model for nonresponse occurrence (Beaumont 2005, Kim and Kim 2007).

In this paper, we consider three estimators of the mean of a finite population in the presence of nonresponse once the estimator has been adjusted by weighting. In particular, we compare Horwitz-Thompson, Hájek and the regression estimator in terms of their bias and variance. We also investigate how different specifications of the response model, used to estimate the adjusting weights, impact on the performance of the estimators. We focus, in particular, on the effect of trimming and we propose a bootstrap approach to identify an optimal cut-point. This approach is evaluated by a data-based simulation using a large survey conducted by Assolombarda, an association of Confindustria (the organisation of Italian manufacturing and services firms). Assolombarda covers more than 5000 associated firms with more than 300000 employees (http://www.assolombarda.it/assolombarda.asp). More specifically we consider a survey conducted in 2009 where about two thirds of the association's firms were contacted. The sample was largely affected by refusals, the response rate being 14%. The response rate remained quite stable over the four years preceding 2009, for which we have data. More specifically the response rates over this period were 11.7%, 13.1%, 11.2%, 14.1% respectively. These values are very low. Historically, however, establishment surveys yield low response rates. Dillman (2000:323) mentions an unpublished study of 183 business surveys for which an average response rate of 21% was found. Hence, the aim of this paper is to consider methods to compensate for nonresponse in business surveys. Clearly an extremely cautious stance should be adopted when using different adjustment methods in these situations and analysts should be confident about the missingness mechanisms acting behind the data. Nonetheless, we believe that providing guidance on this topic is of great interest, particularly for those surveys repeated over time for which actions can be taken to reduce the nonresponse to a more acceptable level (Chun 1997).

The plan of the paper is as follows. In section 2, we briefly review various estimators of the mean once nonresponse is accounted for by sampling weights. In section 3, we propose an optimal strategy for selecting the trimming threshold using the bootstrap. The performance of the estimators considered in section 2 is then investigated in section 4, under various scenarios. Finally, in section 5, the preferred procedure is applied to estimate the average number of female employees using the dataset of Italian firms mentioned above. The conclusions set out in Section 6 end the paper.

Contact information: Riccardo Borgoni, Department of Statistics, University of Milan-Bicocca, Milan, Italy,
e-mail: riccardo.borgoni@unimib.it

2 Preliminaries

Let U be a finite population of size N where N is known and $s \subset U$ is a sample of n units selected from U according to a probabilistic sampling design π, which is independent of the variable of interest Y.

To account for nonresponse it is assumed that each unit in U is, potentially, either a respondent or a non-respondent. A response indicator R is therefore defined which takes value 1 if a population unit responds and 0 otherwise. The respondents sample $s_r = \{i \in s : R_i = 1\}$ is considered to be the result of a two-phase sampling procedure. In the first phase a sample of units is drawn from the target population according to a predefined sample design $\{\pi_1\}$. In the second phase, a sample of respondents is further obtained by a Poisson design where sampling units are selected according to the probability that a respondent is in the sample. Hence $R_i, i = 1, \ldots, N$, are Bernoulli variables with $P(R_i = 1|z_i) = p_i = p(z_i)$, where z is a vector of auxiliary variables. The dependence on z induces response probabilities that vary across units and correspond to the propensity scores proposed by Rosenbaum and Rubin (1983, 1985).

In order to estimate the population mean $\bar{Y} = N^{-1} \sum_{i \in U} y_i$ in the presence of unit nonresponse, we consider initially a weighting-adjusted expansion estimator:

$$\hat{\bar{Y}} = N^{-1} \sum_{i \in s_r} \pi_i^{-1} p_i^{-1} y_i.$$

Such an estimator uses response probabilities to weight the respondent data when estimating the mean. In this way, the higher the probability $\pi_i p_i$ the less the corresponding y_i is weighted. Intuitively, if the unit i has $1/m$ chance of being selected and then responding, then it represents m units.

Since p_i is usually unknown the above estimator is useless. A feasible estimator is obtained by replacing p_i by an estimate $\hat{p}_i$ obtained from the data. This leads to a Horvitz-Thompson-like estimator (HT) of the form

$$\hat{\bar{Y}}_{HT} = N^{-1} \sum_{i \in s_r} \pi_i^{-1} \hat{p}_i^{-1} y_i = N^{-1} \sum_{i \in s_r} w_i y_i$$

where $w_i = \pi_i^{-1} \hat{p}_i^{-1}$ i=1,…,n is a modified weight system which adjusts for nonresponse (Little and Rubin 2002). We also consider the well known Hájek ratio-adjusted (H) estimator (Hájek 1964),

$$\hat{\bar{Y}}_H = \sum_{i \in s_r} w_i y_i / \sum_{i \in s_r} w_i$$

which equals HT if

$$\sum_{i \in s_r} w_i = N.$$

If x is a correlated auxiliary variable known for all units in U, we may consider the regression estimator (REG)

$$\hat{\bar{Y}}_{REG} = \hat{\bar{Y}}_{HT} + B\left(\bar{X} - \hat{\bar{X}}_{HT}\right)$$

where

$$\bar{X} = N^{-1} \sum_{i \in U} x_i, \hat{\bar{X}}_{HT} = N^{-1} \sum_{i \in s} w_i x_i \text{ and}$$

$$B = \sum_{i \in s} w_i x_i y_i / \sum_{i \in s} w_i x_i^2$$

(Särndal et al. 1992, Bethlehem 1988). Note that the preceding estimators may be biased given estimated weights.

A popular response model is the uniform response mechanism within subpopulations, where the response probabilities are constant within subpopulation groups. More sophisticated classes of response models have been developed which use an explicit parametric function to relate the response probabilities to the explanatory variables, mostly aiming to reduce bias and to increase the estimator's efficiency (Bethlehem 1988; Beaumont 2005; Little and Vartivarian 2005; Kim and Kim 2007, amongst others). Non-parametric kernel-type smoothing methods have been introduced by Giommi (1984) and recently extended by da Silva and Opsomer (2006, 2009).

3 Weight Trimming via the Bootstrap

The NWA methods discussed above assume that the response mechanism satisfies the assumption that p_i or their sampling estimate $\hat{p}_i$ are bounded away from zero. However, values for these probabilities that are close to zero, although positive, may produce an undue variation of the final estimator. A way to tackle this problem is provided by trimming. Trimming is a well known approach often invoked in order to stabilise the variance in NWA procedures (Potter 1990; Little et al. 1997), although other model-based approaches have been proposed (Elliot and Little 2000). Trimming proceeds by fixing an a priori cutoff, say w_0, for the weights. The cutoff value is assigned to the units that have a weight above the threshold. The sum of the weights exceeding the threshold is then redistributed to the rest of the sample by multiplying the original weight by an adjusting constant in order to guarantee the overall sum of the weights is unchanged. In the following, we considered the adjusting constant

$$\gamma = \left(\sum_{i \in s} w_i - w_0 \sum_{i \in s} I(w_i \geq w_0) \right) / \sum_{i \in s} w_i I(w_i < w_0)$$

where the $I(A)$ is the indicator function of event A. In principle, although this did not happen in the case considered later on in the paper, the adjustment may force a weight that was originally below w_0 to then lie above this threshold. In this case a second stage (or even further stages if necessary) of trimming might be necessary.

The procedure is usually applied to normalised weights i.e. the sum of the weights equals the nominal sample size. Forcing the weights to lie below a threshold guarantees that the probabilities of response all lie above a positive number. Hence trimming reduces the problem connected with extreme low response probabilities that may induce a high weight for some units in the sample and, in turn, may destabilise the final estimate.

Since the threshold has to be selected exogenously, trimming seems to be somewhat arbitrary. As a consequence, a number of different procedures have been suggested. Potter (1990) reviews some of these methods. We consider below a possible way to identify the trimming threshold in order to maximise an appropriately chosen optimality criterion.

Cox and McGrath (1981) suggest basing the trimming procedure on the MSE of an estimator by calculating the MSE for a range of thresholds w_0 and searching for the one that minimises this function. The assumption behind this procedure is that, for a set of weights, a trimming threshold usually exists where the reduction in variance due to trimming is balanced by the increase in the square of the bias. This approach, although extremely appealing when studying the estimator's performance in controlled experimental settings, as for instance in simulation studies, has some complications when applied to real data where the target value one wants to estimate is unknown.

A way around this is provided by adopting a bootstrap procedure to estimate the MSE (Davison and Hinkley 1997) as explained below. Other measures of the estimator's performance, such as the bias or the variance, can be similarly handled within this procedure.

Let T be the estimator of interest calculated using nonresponse weights trimmed at at w_0. Let $\mathrm{MSE}(T|w_0) = v(w_0)$ be the mean squared error of T given the trimming cut-point w_0 and let $w_{01}, w_{02}, \ldots, w_{0R}$ be a set of potential trimming thresholds. Let $t^*_{r1}, \ldots, t^*_{rB}$ be a set of B bootstrap replicates of T obtained using w_{0r}, $r = 1, \ldots, R$. To obtain this set, the following steps were repeated B times:

- a new sample of size n was obtained by simple random sampling with replacement from the sample, n being the sample size,
- T was computed on the sample trimming the weights at w_{0r}, to obtain t^*_r

Given the usual decomposition $v(w_{0r}) = \mathrm{Var}(T|w_{0r}) + (\mathrm{Bias}(T|w_{0r}))^2$, $v(w_{0r})$ can be approximated by estimating the variance as

$$\mathrm{Var}^*(T|w_{0r}) = B^{-1} \sum_{b=1}^{B} \left(t^*_{rb} - \bar{t}^*_r \right)^2$$

where $\bar{t}^*_r = B^{-1} \sum_{b=1}^{B} t^*_{rb}$ and the bias component as $\mathrm{Bias}^*(T|w_{0r}) = \bar{t}^*_r - t_r$, with t_r being the weighted adjusted estimate obtained on the observed sample using the trimming threshold w_{0r}.

A plot of $\hat{v}(w_{0r})$ versus $w_{01}, w_{02}, \ldots, w_{0R}$ can then be used to explore how the MSE of T depends on w_0 and to identify the optimal threshold for trimming as the one that defines the minimum value of this MSE. In order to propagate the uncertainty due to weights estimation, the weights have to be re-estimated within each bootstrap replicate before trimming. Clearly, while the evaluation of $\mathrm{MSE}(T|w_0)$ is difficult, the bootstrap procedure described above is relatively simple to implement.

4 Simulation Study

In this section, we report the results of a large simulation study based on a dataset of firms located in the Lombardy region of Italy. This dataset was introduced in section 1 that will be considered in the next section in more detail.

This simulation exercise had three different goals. Firstly we wanted to compare the performance of the estimators of the mean introduced in section 2 when one has past information on the response behaviour of sampling units that can be used to construct nonresponse weights. It might be expected that this improves the performance of an estimator, although different estimators may benefit from this in different ways. Secondly we considered the impact of a miss-specification of the response model, focussing on the choice of the link function relating the probability of response to a set of explanatory variables under a generalised linear response model. Finally, we wanted to assess the effectiveness of the bootstrap-based optimal trimming method introduced in the previous section.

The simulation design was as follows. We considered a sample of 3657 Assolombarda companies contacted in 2009. For all these firms, the number of employees, the economic sector they operated in and the annual total sales were available. This information is collected annually from the companies when they subscribe to the association. Of this sample, we restricted attention to 3401 companies since it was not possible to get either the information about the annual total sales or the identification code necessary to univocally collect firm information from different files for 256 units in the initial dataset. This dataset then served as a working target population in the simulation study, with average total sales corresponding to the parameter of interest. Note that this value is, in fact, known (the actual figure is not reported here for confidentiality reasons). A questionnaire concerning a number of items, mostly related to the labour conditions, was administered by ordinary mail or email. Out of these, only 407 were returned.

The Monte Carlo experiment was structured as follows:
Step 1: the target population and the actual sample of respondents were used to fit a logistic model for the response indicator variable in the year of interest (2009). This variable is denoted by R_{09} from now on.
Step 2: the model obtained in step 1 was used to simulate missingness in the experiment according to the following procedure:

1. a sample of size n was randomly selected without replacement from the working population assuming, for simplicity, that all units have the same inclusion probability,
2. nonresponse was simulated in the sample by making a Bernoulli draw for each unit according to the probabilities derived by the model obtained in step 1. This produced a simulated response indicator for 2009, denoted by R^*_{09},
3. a logistic regression for R^*_{09} was fitted to estimate the response probability, $P(R^*_{09} = 1)$, of each unit in the simulated sample,

4. a sampling weight was calculated for each unit among the simulated respondents as the reciprocal of the estimated response probability,
5. various NWA estimators of the mean were computed.

Step 3: step 2 was repeated several times to obtain the Monte Carlo relative bias and MSE of each estimator to compare performances.

The simulation study investigated a number of different response models. In all cases, the estimators considered were the HT, H and REG estimators of the mean, with REG using number of employees as the auxiliary variable.

4.1 Assessing the effect of historical information on NWA estimators

We observed that nonresponse in the 2009 survey was highly associated with nonresponse in the 2008 survey. More specifically, the odds ratio defined by the response indicators for the two years, i.e. R_{09} and R_{08}, was nearly 10. Note that all the companies which took part in the survey in 2009 had also been contacted in 2008.

The lagged response indicator, R_{08}, was also positively related to the target variable, with average total sales for respondents being roughly 3.5 times higher than the average total sales of non respondents. Since this variable was related to the probability of response and to the survey outcome, it could be usefully employed, in principle, to reduce the nonresponse bias and variance of the survey estimators (Little and Vartivarian 2005).

Finally, we note that although missingness in the dataset at hand was due to nonreturned questionnaires, it can be considered to be item nonresponse in the sense that some information was available for each firm. This information was not collected by the questionnaire but retrieved from the records of the Association which obtains it annually from firm subscriptions. The variable of interest, average total sales, was, in fact, one of such variables and so had virtually no missingess.

We considered two response models which differ in terms of the covariates included in the linear predictor:

1. Model I included, in addition to R_{08}, another population variable, namely the sector of activity of a firm,
2. Model II included, in addition to the two variables considered in model I, the number of employees i.e. the auxiliary variable used in REG.

Both sector of activity and number of employees were available for the entire population at hand.

We estimated a logistic regression for R_{09} in step 1 using both models. The coefficients of the two logistic regressions were then used to simulate nonresponse in step 2, thus generating two different nonresponse scenarios.

The logistic regression model for R^*_{09} fitted within each iteration of the simulation and then used to derive the adjusting weights for the NWA estimators was specified in terms of four different predictor sets. The variables defining these predictor sets were:

(a) the sector of activity of a company;
(b) the sector of activity of a company and R_{08};
(c) the sector of activity and the number of employees;
(d) the sector of activity, the number of employees and R_{08}.

Comparison between a) and b) when data are simulated according to Model I is aimed at assessing the impact of accounting for the lagged response indicator in the estimated weights. These results are reported in the top part of Table 1.

Comparison of c) and d), when data are simulated according to Model II, addressed a similar task when an additional variable (the number of employees) which correlates both with the nonresponse and the target variable (total sales) was used in the estimated weights. These results are reported in the bottom part of Table 1.

The REG estimator was by far the best performing under the four different scenarios. From the top part of Table 1 it can be noted that both the HT and H estimators performed very poorly in terms of relative bias when the lagged response indicator R_{08} was excluded from the weighting model, whereas the REG estimator did much better. When R_{08} was included in the weighting model, the performance of HT and H improved considerably both in terms of bias and MSE. In contrast, this had a detrimental effect on the REG estimator as far as MSE is concerned, although it still produced a bias reduction. One may argue that this occurred because the information about missingness was already integrated into the REG estimator through the auxiliary variable, which is a good predictor of nonresponse in the population.

The 1000 simulated sets of estimation errors underpinning Table 1 were bootstrapped using the percentile method, and 95% bootstrap confidence intervals for the actual Monte Carlo biases constructed. Asterisks are used in Table 1 to denote where these intervals did not include zero. We see that the Monte Carlo bias is always significant when the weighting model is misspecified due to R_{08} not being included. In contrast, we see that even when the weighting model is correctly specified, with R_{08} included, both HT and H still record significant underestimates of the target parameter when the response model also depends on the auxiliary variable used in REG. However, we also see that the MSE performance of both HT and H improves considerably whenever this auxiliary variable is included in the weighting model. Finally, we note that REG itself becoming more volatile when R_{08} is used in the weighting model. One explanation for this behaviour is that REG already adjusts for nonresponse using the auxiliary variable (employees), which is itself associated with R_{08}. Consequently, addition of R_{08} in the weighting model does not add any more information as far as bias correction for REG is concerned.

The results described above are for a nominal sample size (i.e. before generating nonresponse) of 1000 units taken without replacement from the working target population. The Monte Carlo exercise was repeated for various sample sizes ranging from 250 to 2000 units. We do not report these results in this paper since they were very similar to those shown in Table 1. Interestingly, we found that the sample size scarcely influenced the bias, at least for moderate to large sample sizes. But, as it could have been expected in advance, the sample size heavily affected the variances. In general, ignoring the nonresponse and estimating the parameter of in-

Table 1: Results of the simulation study based on 1000 Monte Carlo replicates. Bias (%) is the modulus of the relative bias expressed as a percentage. An asterisk next to a value for Bias (%) indicates that a 95% bootstrap confidence interval for the underlying Monte Carlo bias does not include zero

Response Model	Activity + R_{08}					
Weighting Model	Activity			Activity + R_{08}		
Estimator	HT	H	REG	HT	H	REG
Variance	809	865	81	174	175	114
Mse	847	908	82	174	175	115
Bias (%)	69	73	10	1	2	3

Response Model	Activity + R_{08} + Employees					
Weighting Model	Activity + Employees			Activity + R_{08} + Employees		
Estimator	HT	H	REG	HT	H	REG
Variance	78	83	69	124	125	114
Mse	79	84	69	125	126	114
Bias (%)	12	9	9	14	12	4

terest just using the sample of respondents produced heavily biased estimates, particularly if there were few respondents.

4.2 Assessing the effect of link mis-specification

In order to assess the impact of a link mis-specification on the estimators' performances, we modified the simulation design by changing the link function of the response Model I used to generate missingness in each iteration, while still using a logistic regression to estimate the weights. In particular a cauchit link, i.e. the cumulative distribution function of the Cauchy distribution, which is much heavier tailed than the logistic, was adopted. The results for a nominal sample size of 1000 units are reported in Table 2, and show how a severe departure from the assumed response model can seriously affect the performance of the final estimator.

Some further sensitivity analyses were conducted by specifying less severe mis-specification of the link function (using a probit or a c-loglog model, see McCullagh and Nelder 1989). These results are not reported in detail here. We found that when the mis-specification is milder than in the simulations underpinning Table 2, then the underperformance of the various estimators is less severe. For instance, when we use a c-loglog link to simulate missingness while still adopting a logistic link to estimate the response probabilities, we obtain MSE values of 867, 941 and 82 for HT, H and REG respectively under a weighting model that only includes Activity.

Our results reinforce the fact that, since the correct response model is never known in advance, analysts should make sure that this aspect is carefully checked in real applications. A possible strategy for this will be considered in the case study presented in the next section.

4.3 Assessing the performance of the bootstrap-based trimming procedure

We finally evaluated the performance of the bootstrap-based optimal trimming procedure described in section 3. In particular, we focussed on the REG estimator with missingness simulated using a logit link. At each iteration of the Monte Carlo study, the bootstrap trimming procedure was applied by bootstrapping the data 500 times and refitting the logistic regression to each bootstrap replicate. The nominal sample size in each iteration was 1000. More specifically, the simulation procedure described at the beginning of this section was modified as follows. After non response generation in the simulated sample (point 2, step 2):

1. take 500 bootstrap samples, and for each of them reestimate the weights by refitting a logistic regression model and then trimming the weights using preselected trimming thresholds,
2. calculate the bootstrap MSE as described in section 3 for each trimming threshold and select the one with the minimum MSE,
3. calculate the REG estimator using weights trimmed to that (minimum MSE) threshold,
4. iterate one thousand times and calculate the Monte Carlo diagnostic.

The MSE of the REG estimator was evaluated over a grid of 22 candidate trimming thresholds, though the actual optimal trimming thesholds used were confined to a subset of 12 of these candidate thresholds. The bottom part of Table 3 shows the averages (across simulations) of both the number of units and their percentage of the total weight for units with weights above the trimming threshold for value in this subset. The trimming level at which the minimum MSE was achieved was used for the actual trimming of the weights used in the REG estimator. The top part of Table 3 shows how the REG estimator based on such trimmed weights compares with the REG estimator based on the original weights.

Table 2: Results of the simulation study based on 1000 Monte Carlo replicates aimed at assessing the effect of the link mis-specification. Bias (%) is the modulus of the relative bias expressed as a percentage

Response Model	logit(Activity + R_{08})					
Weighting Model	logit(Activity)			logit(Activity + R_{08})		
Estimator	HT	H	REG	HT	H	REG
Variance	809	865	81	174	175	114
Mse	847	908	82	174	175	115
Bias (%)	69*	73*	10*	1	2	3
Response Model	cauchit(Activity + R_{08})					
Weighting Model	logit(Activity)			logit(Activity + R_{08})		
Estimator	HT	H	REG	HT	H	REG
Variance	4724	5907	93	919	897	453
Mse	4924	6182	95	919	898	454
Bias (%)	159	186	18	6	4	9

Note: An asterisk next to a value for Bias (%) indicates that a 95% bootstrap confidence interval for the underlying Monte Carlo bias does not include zero.

It appears that, if one is ready to pay a price in terms of an increase in the bias, then the reduction in the MSE obtained by applying the weight trimming procedure proposed in section 3 can be substantial. However, if bias is of particular concern, then the increase of bias due to trimming shown in Table 3 should be considered.

5 Case Study: Estimating the average number of female employees in Lombardy companies

The survey introduced in the previous section is conducted annually, and primarily to measure labour market dynamics. One of the goals is to quantify female labour force participation. In this section our aim is therefore to estimate the average number of female employees per firm in the target population as an indicator of such participation. We consider the sample contacted in 2009 consisting of 3401 out of more than 5000 companies associated with Assolombarda. Out of these only 407 returned the questionnaire. Estimation of the target parameter was carried out using a regression estimator adjusted for nonresponse as discussed in section 2.

We first consider specification of the model used to estimate the response probabilities. Based on the results of the simulation study in section 4, we excluded the lagged response indicator from the model. More specifically, the covariates included in the modelling were: the overall number of employees, the ATECO[1] sector of economic activity and two binary variables, which identify those companies belonging to the top and bottom 10% of the sample, in terms of annual total sales. We collapsed the ATECO classification into twelve categories, which were deemed appropriate for the data considered in this paper.

The simulation results presented in the previous section also suggested that the performance of the regression estimator is sensitive to the link function adopted in the GLM specification. In our data, the logit specification of the link does not pass the Pregibon test at the 10% significance level (p-value 0.09). In order to identify a more appropriate link function for the data at hand, we considered the Pregibon family (Pregibon 1980). This is a flexible family of link functions indexed by two parameters, of the form

$$g(u, a, b) = \frac{u^{a-b} - 1}{a - b} - \frac{(1-u)^{a+b} - 1}{a + b} a, b \in \mathbb{R}, u \in [0, 1].$$

The function g is the quantile function of a version of the generalized Tukey-l distribution. See Freimer et al. (1988) and Koenker and Yoon (2009). The logit link occurs when $a \to 0$ and $b \to 0$ by the usual de L'Hôpital rule; for $b = 0$ one gets a family of symmetric densities with a controlling the heaviness of the tails, whereas b controls for the skewness if $b \neq 0$. Koenker (2006) provides R functions that implement these link functions.

In order to estimate a and b we profiled the likelihood and maximised it numerically. More precisely, indicating by β the parameters of the linear predictor, the procedure consists in iterating the following two steps until convergence:

1. maximise the profile likelihood of (a,b) given $\tilde{\beta}$ to obtain $(\hat{a}, \hat{b})$
2. maximise the profile likelihood of β given $(\hat{a}, \hat{b})$

where $\tilde{\beta}$ is the estimate of β obtained in the previous iteration of the procedure. The algorithm was initialized by setting $\tilde{\beta}$ equal to the estimate provided by ordinary logistic regression in the first iteration and was stopped when the modulus of the relative deviance in two successive iterations, i.e. $\left|D_j - D_{j-1}\right| / D_{j-1}$, was smaller than 10^{-5}. The estimates obtained were $(\hat{a}, \hat{b})$=(4.16, 0.14). We noted that, while a was

[1] ATECO is a nomenclature adopted by ISTAT, the Italian office of national statistics, to translate the EU NACE classification.

Table 3: Results of the simulation study based on 1000 Monte Carlo replicates to assess the performance of the bootstrap-based trimming procedure (500 bootstrap replicates). Results refer to the NWA REG estimator. Bias (%) means the modulus of the relative bias expressed as a percentage

	Response & Weighting Model			
	Model I		Model II	
	No Trimming	Trimming	No Trimming	Trimming
Variance	114	91	114	77
Mse	115	92	114	79
Bias (%)	3	12	4	15

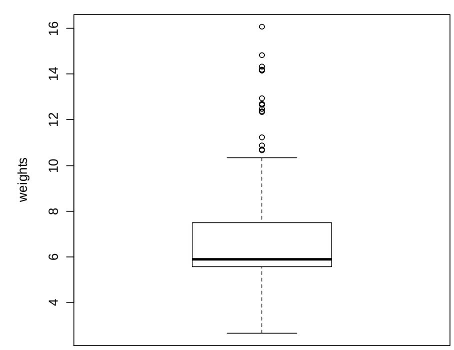
Trimming threshold	8.4	10.1	11.8	13.5	15.2	16.8	18.5	20.2	21.9	23.6	25.3	42.1
Average number of respondents with weights above trimming threshold	153.2	83.1	47.9	28.3	17.9	11.7	7.7	4.9	3.2	2.1	1.5	0.1
Average percentage of total sample weight above trimming threshold	13.9	8.2	5.1	3.3	2.2	1.5	1.0	0.7	0.5	0.4	0.3	0.1

well identified by the likelihood, the likelihood surface was quite flat in the b direction.

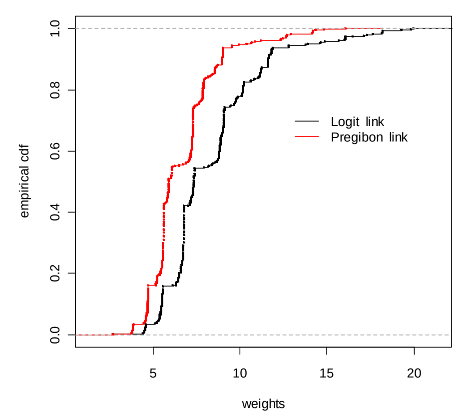
Plugging $(\hat{a}, \hat{b})$ in the Pregibon link, we then fitted a binary regression model which was used to calculate the response probability based on the variables included in the linear predictor. Table 4 shows the estimated regression coefficients of the response model.

From Table 4, it appears that a firm's propensity to respond increases the more employees a firm has, whereas firms that have larger and smaller total sales tended to be less prone to respond (odd ratio 0.61 and 0.54 respectively). This result is not unreasonable since the correlation between the two variables (i.e. total sales and number of employees) is extremely weak in the target population ($R^2 = 0.17$). Interactions amongst the variables in Table 4 were not found to be significant and hence were not considered in the subsequent analysis. The reciprocal of the estimated probabilities generated by this model provided the values for weighting the responding sample units in the regression estimator. Figure 1-(a) shows the distribution of the estimated weights.
Figure 1-(b) compares the empirical cumulative distribution function (cdf) of the weights obtained using the Pregibon link with the cdf of the weights obtained by ordinary logistic regression. The former cdf is stochastically greater than the latter and the difference in the two cdf's was found to be highly significant using a Kolmogorv-Smirnov test. This is due to the fact that the estimated Tukey-l distribution is, although nearly symmetric, lighter tailed than the logistic distribution. This produces a much more concentred sample distribution for the response probabilities than one would have obtained had the logistic regression been used instead. This feature helps in preventing the occurrence of zero or extremely small probabilities of response for the responding sample units.

The bootstrap procedure described in section 3 was applied to determine an appropriate level of trimming for nor-



(a) Distribution of the weights obtained by Predibon binary regression



(b) The cdfs of the sampling weights obtained by different link functions

Figure 1.

Table 4: Estimated parameters of the response model. The reference category for the ATECO classification is "Manufacture of food products"

	Estimate	Std. error	p-value
Intercept	-1.2044	0.3514	0.0006
Number of employees	0.0003	0.0001	0.0435
ATECO: Textiles and apparel manufactures	-0.4635	0.5086	0.3621
ATECO: Chemical/petroleum products manufactures	-0.3065	0.3904	0.4325
ATECO: Rubber/plastics products manufactures	-0.6787	0.4622	0.1420
ATECO: Basic/fabricated metal products manufactures	-0.5616	0.3699	0.1290
ATECO: Electronic manufactures	-0.6306	0.4010	0.1159
ATECO: Other manifacturings	-0.9236	0.4137	0.0256
ATECO: Retail trade services	-1.1779	0.3999	0.0032
ATECO: Transportation and storage activities	-1.0248	0.4498	0.0227
ATECO: Information/communication manufactures	-1.0177	0.3925	0.0095
ATECO: Other services	-0.9189	0.3765	0.0147
ATECO: Unclassified	-1.1182	0.4298	0.0093
Total sales above 90-th sample percentile	-0.4897	0.1949	0.0120
Total sales below 10-th sample percentile	-0.6102	0.1779	0.0006

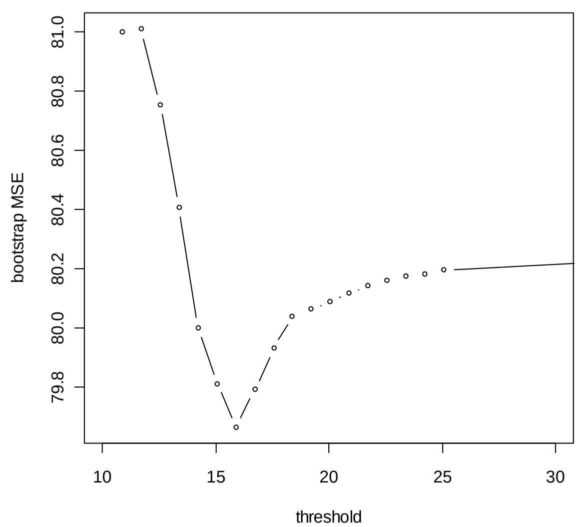


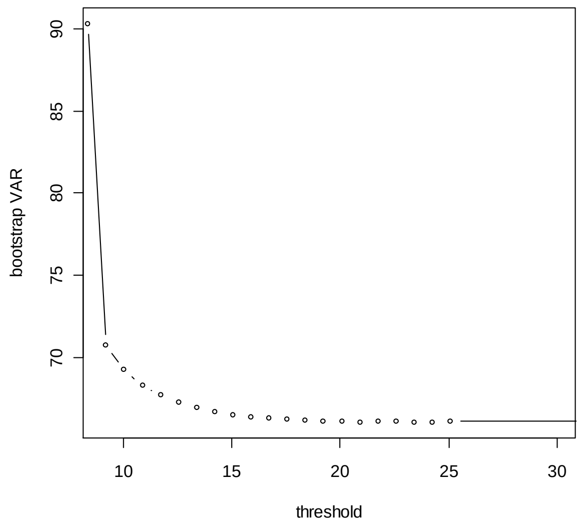
Figure 2. Bootstrap MSE as a function of the trimming threshold

malised weights. Figure 2 shows the shape of the bootstrap MSE of the regression estimator as a function of the trimming threshold. This suggests 15.9 as the optimal trimming level i.e. the one which minimises the bootstrap MSE. Note that the reciprocal of the overall response rate is 8.35 i.e. each firm that returned the questionnaire on average corresponds to, roughly, 8 firms in the sample, and so this procedure suggests that weights be restricted to less than twice the reciprocal response rate. Adopting this threshold resulted in the eight most extreme weights being trimmed down to 15.9.

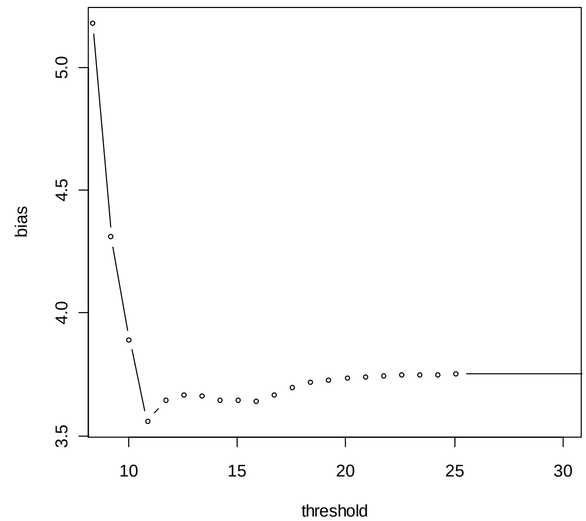
Figure 3 shows how the bias and variance components of the MSE are affected by the trimming. It is clear that the most important component of the MSE is the variance. This component tends to decline monotonically as the threshold increases. The bias, on the other hand, initially decreases, but then tends to increase slowly as the threshold increases.

The average number of women per firm was estimated to be 28 using the respondent sample mean and was even smaller, 21, when estimated using the unadjusted regression estimator. After adjusting for nonresponse using the weights calculated as described above, the regression estimator value was 40. Note that the total number of employees was used as the auxiliary variable in the regression estimator. This variable was found to correlate moderately with the number of females, with the correlation coefficient being 0.7. The 95% confidence interval estimated by the bootstrap was equal to (30.7, 61.3) suggesting that the difference between the respondent mean estimator and the adjusted regression estimator is significant. These 95% confidence limits were obtained by re-sampling the respondents 500 times with replacement.

The average number of respondents in the bootstrap replicates was 407.6, which was basically the same as that observed in the actual sample, ranging from a minimum of 358 firms to a maximum of 478. For each bootstrap sam-



(a) Bootstrap variance



(b) Bootstrap bias

Figure 3. Bootstrap variance (a) and bootstrap bias (b) as a function of the trimming threshold

ple, the response model and the trimmed weights were re-estimated to include variability due to choice of the trimming threshold. Note that the parameters of the Pregibon link function were kept fixed in the bootstrap calculations.

6 Conclusions

Refusals are very common problems in business surveys, although not often discussed in the literature of this field. In this paper, we review some business survey estimation procedures that are adjusted for unit nonresponse under the assumption that this nonresponse is Missing At Random. We consider several aspects which emerge in this context. Business surveys, whether or not structured as a proper panel, are often repeated in time. In these circumstances, information on unit response behaviour in the past is available and can be used in subsequent estimates through the nonresponse weights. By means of an extensive simulation analysis, we found that this information does not necessarily improve the performance of an estimator. In particular, the regression estimator, which may already adjust for nonresponse through the auxiliary variable, benefits much less, if at all, compared with other less complicated estimators of the population mean such as the Horwitz-Thompson and the Hájek estimators.

Adjusting weights are usually estimated from the data, since the probability that a unit remains in the survey is generally unknown in advance. Given the binary nature of the decision a firm takes as to whether the questionnaire is returned or not, it is common practice nowadays to estimate the weights by resorting to logistic regression, particularly when the survey provides secondary variables which may help in differentiating unit response behaviour efficiently. However "binary" does not necessarily mean "logit". We demonstrate that a mis-specification of the link function may have a strong impact on the performance of a weighting adjusted estimator. Using data collected in an annual Italian business survey, we show how a more flexible family of link functions can be usefully employed in this context.

It is known that an estimate of the response probability that is close to zero produces an extremely high weight for the unit and this, in turn, leads to undue variation in the final estimates. Weight trimming is commonly adopted to stabilise the variance in weighting adjusted estimators. In this paper, we propose a procedure to determine a trimming threshold for the weights using the bootstrap which is optimal in the sense of minimising an estimate of the MSE. Our simulations indicate that the reduction in the MSE obtained by applying this procedure can be substantial. In our case study we illustrate how the optimal trimming threshold for the weights depended on the interplay between the variance and the bias of the estimator and how a bad choice of this threshold may induce a large bias.

Acknowledgements

We warmly thank two anonymous referees and the Associated Editor whose comments remarkably improved the contents of this paper. We also thank Assolombarda for providing the data used in this paper and Denise Kilmartin for her help in editing the text.

References

Beaumont, J. F. (2005). Calibrated imputation in surveys under a quasi-model-assisted approach. *Journal of the Royal Statistical Society*, *Series B, 67*, 445-458.

Bethlehem, J. G. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics*, *4*, 251-260.

Chun, Y. I. (1997). *Nonresponse Follow-Up in Establishment Surveys: A Split-Half Experiment.* Proceedings of the Survey Research Methods Section, ASA, http://www.amstat.org/sections/srms/proceedings/papers/1997_170.pdf. Last accessed 13 November 2012.

Cox, B. G., & McGrath, D. S. (1981). *An examination of the effects of sample weight truncation on the mean square error of survey estimates.* Presented at Biometrics Society ENAR Meeting, Richmond, VA.

da Silva, D. N., & Opsomer, J. D. (2006). A kernel smoothing method to adjust for unit nonresponse in sample surveys. *Canadian Journal of Statistics*, *34*, 563-579.

da Silva, D. N., & Opsomer, J. D. (2009). Nonparametric propensity weighting for survey nonresponse through local polynomial regression. *Survey Methodology*, *35*, 165-176.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge University Press.

Dillman, D. A. (2000). *Mail and internet surveys: the tailored design method*. New York: Wiley.

Elliott, M. R., & Little, R. J. A. (2000). Model-Based Alternatives to Trimming Survey Weights. *Journal of Official Statistics*, *16*(3), 191-220.

Freimer, M., Mudholkar, G., Kollia, G., & Lin, C. (1988). A Study of the Generalised Tukey Lambda Family. *Communications in Statistics Theory and Methods*, *17*, 3547-3567.

Giommi, A. (1984). A simple method for estimating individual response probabilities in sampling from finite populations. *Metron*, *42*, 185-200.

Groves, R. M., Dillman, D., Eltinge, J., & Little, R. (2002). *Survey Nonresponse*. New York: John Wiley and Sons.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics*, *35*, 1491-1523.

Kim, J. K., & Kim, J. (2007). Non-response weighting adjustment using estimated response probability. *The Canadian Journal of Statistics*, *35*, 501-514.

Koenker, R. (2006). Parametric links for binary response. *R News*, *6*, 32-34.

Koenker, R., & Yoon, J. (2009). Parametric Links for Binary Choice Models: A Fisherian-Bayesian Colloquy. *Journal of Econometrics*, *152*, 120-130.

Little, R. J., & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means? *Survey Methodology*, *31*, 161-168.

Little, R. J. A., Lewitzky, S., Heeringa, S., Lepkowski, J., & Kessler, R. C. (1997). An assessment of weighting methodology for the national comorbidity study. *American Journal of Epidemiology*, *146*, 439-449.

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall.

Potter, F. (1990). *A Study of Procedures to Identify and Trim Extreme Sampling Weights.* Proceedings of the American Statistical Association, Section on Survey Research Methods, 225-230, http://www.amstat.org/sections/srms/Proceedings/papers/1990_034.pdf. Last accessed 13 November 2012.

Pregibon, D. (1980). Goodness of link tests for generalized linear models. *Applied Statistics*, *29*, 15-24.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*, 33-38.

Särndal, C. E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer.