Nonparametric Estimation of Non-Response Distribution in the Israeli Social Survey

Yury Gubman[1], Dmitri Romanov[2]

Abstract

Non-response adjustment in the Israeli Social Survey (ISS) is based on the MAR assumption. Association of non-response with key socio-economic characteristics (individual's economic status and degree of religiosity) which do not correlate strongly with standard survey design and calibration variables may corrupt the MAR assumption validity. We analyze survey and item non-response in ISS by estimating non-parametric sharp bounds for conditional means of key ISS variables. Statistical tests for checking validity of MCAR and MAR assumptions are proposed, where the test statistics are based on the width of the interval between the estimated bounds. We find significant departures from MAR assumption in the ISS data. Non-response propensity varies significantly between population groups assumed to be homogenous according to the survey design. We propose to utilize information about income and religiosity, available on individual or neighborhood level, for improving the ISS design.

Key Words: survey non-response, item non-response, MAR assumption, sharp bounds.

1. Introduction

At the Israeli Central Bureau of Statistics (ICBS), imputation, post-stratification and weighting calibration procedures are used for adjusting the obtained estimates to survey non-response. The Missing Completely at Random (MCAR) or Missing at Random (MAR) assumptions lie behind these procedures. In most cases, the validity of these assumptions can not be checked. However, given significant association of non-response with key socio-economic characteristics (individual's economic status, labor force participation, family status and degree of religiosity), which do not correlate strongly with standard survey design and calibration variables, and availability of reliable proxy variables for these traits from administrative sources, may provide a unique possibility to check the MAR assumption validity. An attractive way to study non-response distribution and non-response impact to the survey estimates is estimation of sharp bounds for conditional mean, using nonparametric procedure proposed by Horowitz and Manski (1998, 2000). This method does not require any assumption about the true non-response characteristics, and allows estimation of the impact of ambiguity caused by all types of non-response.

In the current study, non-response distribution and its influence on survey estimators are investigated on the Israeli Social Survey (ISS) 2006 data. For this purpose, Horowitz and Manski sharp bounds and their standard deviation are estimated, for a set of variables, with and without conditioning on the known socio-demographic and geographic variables from the administrative sources. Using the sharp bounds methodology, we propose statistical tests for checking MCAR and MAR assumptions, where the width of the interval between the lower and the upper bounds is used as an estimator of non-response influence.

Using the survey data only, MAR assumption can not be tested statistically, because the non-respondents data, which is needed for such a test, is unavailable. We overcome this difficulty by linking the survey data with the administrative databases and conditioning

[1] Corresponding author, E-mail: yuryg@cbs.gov.il, Israeli Central Bureau of Statistics and Hebrew University of Jerusalem
[2] Israeli Central Bureau of Statistics

on the administrative covariates which are strongly correlated with the survey theme and/or with the key survey variables.
We find significant departures from MAR assumption in the ISS. The non-response propensity differs between various population groups assumed to be homogenous according to the survey design. We propose to utilize information about incomes, labor market status and degree of religiosity of Jews, available on the individual or neighborhood level from auxiliary sources, for improving the ISS design.
The rest of the paper is organized as follows: in section 2 the issue of survey non-response is discussed. Section 3 explains methodological issues of sharp bound's estimation procedure proposed by Horowitz and Manski, including estimation of standard deviation for bounds. Statistical test for checking MCAR assumption using the sharp bounds approach is developed, and methodology for detecting departures from MAR assumption using administrative data is proposed. The ISS methodology is outlined briefly in section 4. Section 5 presents sharp bounds estimates and their standard deviations for selected ISS variables, conditional on administrative covariates of interest. Given the ISS sample design, MAR assumption is tested. The main conclusions and discussion are presented in Section 6.

2. Non-response in surveys and its influences

2.1 Missing data generating mechanisms

Every survey is subject to non-response. We distinguish between interview non-response, where the selected person did not participate in the survey, and item non-response, where the selected person participated in the survey but did not answer certain question. Both types of non-response result in missing values in the survey data, unless imputations are made.. The impact of non-response may be different, according to the missing values generating mechanism. Three mechanisms of non-response are known in the literature.

1. Missing Completely At Random (MCAR): Non-respondent's data is ignorable. The probability that observation *i* is missing $P_i(m)$ does not depend on observed and unobserved characteristics:

$$P_i(m \mid Y_r, Y_n) = P_i(m), \tag{1}$$

where Y_r is known respondent's data, and Y_n is unobserved non-respondent's data. Equation (1) means that the probability of a person to be survey non-respondent does not depend on his characteristics or survey's theme. For item non-response, it means the probability does not depend on the specific question. If data was generated by MCAR, the valid results could be obtained by performing the analysis only on the complete data, with no bias being introduced in the survey estimators.

2. Missing At Random (MAR): Conditionally on some set of covariates, the non-respondent's data is ignorable. The probability of observation *i* being missing does not depend on unobserved traits:

$$P_i(m \mid Y_r, Y_n) = P_i(m \mid Y_r). \tag{2}$$

Equation (2) means, that the respondent's data is sufficient for obtaining valid survey estimators, and missing value generating mechanism can be expressed solely in terms of the observed data. Let *y* be a survey variable. If the non-respondent's values of *y* are ignorable, they do not provide any additional information about the population distribution of *y*. Hence, conditioning on survey covariate *x* the MAR assumption is given by:

$$P(y \mid x, z=0) = P(y \mid x, z=1) \tag{2a}$$

where $z = 1$ for respondents and 0 otherwise. Equation (2a) implies that $E(y \mid x, z = 0) = E(y \mid x, z = 1)$, for each survey variable *y*.

Denote a set of observed covariates by $X = \{x_1, ... x_k\}$. Suppose *X* controls non-response bias in *y*: (a) conditionally on *X*, the probability of response does not depend on *y*, (b) conditionally on *X*, distributions of *y* for respondents and non-respondents are equal. It follows that given *X*, the probability of *y* to be missing does not depend on *y*. Then, MAR assumption (2) can be rewritten as:

$$P_i(m \mid y_r, y_n, X) = P_i(m \mid y_r, X) = P_i(m \mid X). \tag{2b}$$

Definition (2b) is widely used in survey design practice (see Vartivarian and Little, 2003). By choosing an appropriate set of observed covariates which fulfill conditions (a) and (b) above, one can create stratas which are homogenous with respect to non-response probabilities. Equation (2b) means, that in each strata survey non-response is generated at random and non-response probabilities are independent of the survey variables or/and survey topic.

3. Not Missing At Random (NMAR): The probability of observation to be missing depends on observed and unobserved data. Under NMAR, the unknown non-respondents data is non-ignorable and the sampling process does not identify $E(y \mid x)$ without a bias.

Assuming MCAR or MAR provides validity to a number of widespread statistical methods for non-response adjustment, such as standard procedures for imputations and weight's calibration. In recent years, estimating methods under NMAR missing values generating mechanism were developed, but they are rather complex and do not used in the statistics bureaus.

2.2 Identification and characterization of survey and item non-response

MCAR assumption can be tested using the survey data. In order to reject MCAR, it is sufficient to show that the non-response probability depends on observed characteristics.

Detecting violations of the MAR assumption, however, is more difficult and testing it requires knowing non-respondents characteristics. Obviously, this information is not available from the survey itslf. In some cases, however, other sources of information about non-respondents might be analyzed, such as non-respondents' survey or administrative records.

Groves and Couper (1998) concluded that participation in surveys depends on the socio-economic and demographic characteristics of person and/or household, such as income, household's size and age. The authors pointed out, that a higher response rate among minority groups could be caused by the household's socio-economic status. Their study was based on the analysis of social surveys, income surveys and labor force surveys from different countries.

Sherman (2000) proposed statistical tests for checking MAR and MCAR assumptions for the item non-response case. The proposed test is based on comparison, conditioning on a chosen covariate, of the estimated non-response rate for the variable of interest, with the expected rate. The odds ratio statistics were calculated. The key assumptions of the study were (a) survey has a simple random sample; (b) survey non-response is random. Applying this methodology to the National Election Study showed significant departures from the MCAR and MAR. However, the proposed tests can not be used universally, especially when in presence of complex survey design. Furthermore, generalization to the multi-covariate case is not straightforward.

Analysis of characteristics of "late" and "difficult" interviews (interviewed after 9 or more contacts) in the American National Health Interview Survey (NHIS), Chiu and Hardy (2001) found that the estimates from selected health items are quite different among late/difficult interviews compared to all other respondents. The authors assumed that late or difficult respondents have approximately the same characteristics as non-respondents. They found that the probability to be late or difficult respondent depends on economic status and household's size.

Dixon (2002, 2004) examined item non-response rates in the American Current Population Survey, using factor analysis of the refusal pattern (2002), and logistic regression models with a set of Census 2000 covariates (2004). It was found, that non-response for the specific questions depends not only on race, age and sex, but also on the labor force participation variables.

Wang (2004) compared interview non-response rates in the Norwegian Business Tendency Survey, and found that these rates, conditioning on the number of employees in establishment, distribute approximately uniform. He concluded that MCAR assumption is valid in this case.

Riphahn and Serfling (2005) studied item non-response in the German Socio-Economic Panel. The item non-response rates for selected economic questions in the survey depend on the research variables, such as income and wealth indicators.

Abraham, Maitland and Bianchi (2006) proposed non-parametric weights correction procedure to mitigate a non-response bias. By comparison of sample weights (inverse of sampling probability) and the proposed weights in the American Time Use Survey, the authors found that the non-response rates in this survey depend on socio-economic characteristics of persons: for example, people who are weakly integrated into their communities are less likely to respond. Note that a-priori identification of such population groups, in order to improve sample design, may be difficult.

Schechtman, Yitzhaki and Artsev (2005) studied interview non-response characteristics in the Israeli Household Expenditure Survey (IHES) 1997-2001, using extended Gini regression methodology. Significant departures from the MAR assumption were found: the non-response propensity was proven to be associated with person's income, work status, household size, religion and degree of religiosity (for Jewish population). Weights adjusted for non-response were used as a dependent variable, under the assumption that without interview non-response, there is no need for weight's adjustment. Hence, the higher is the rate of non-response in a population group, the higher weight should be assigned to the units in this group. Consequently, the difference between original (inverse of sampling probability) and adjusted weights in this group increases. However, weights adjustment procedure in the ICBS is used for two goals - non-response adjustment and minimization of discrepancies between several data sources, and it is quit difficult do distinguish between the two influences. Furthermore, the adjusted weights were calculated using MAR assumption, which was not checked.

Pfeffermann and Sikov (2008) analyzed the IHES 2005 data. They estimated the probability of survey non-response, given a set of survey covariates, by logistic regression. It was found that the probability to be non-respondent depends on household's income, contacted person sex, district and household size. They concluded that one cannot ignore the non-respondents' data in this survey, and the true missing value generating mechanism is NMAR. This conclusion was used as a motivation for developing a new imputation procedure for the NMAR case.

In recent years, statistical offices gain access to ever increasing number of administrative data sources. These new "external" data are a potent source of information for non-response studies, since they cover a wide spectrum of population's socio-economic characteristics, whose values are therefore available for all the sampled persons, including non-respondents. In the current study we use three covariates from administrative data source, which do not correlate strongly with standard survey design and calibration variables - work status, income and degree of religiosity of Jews. Possible generalization of the proposed methodology to surveys in which sample unit is household (Labor Force Survey, Household Expenditure Survey) will be discussed in section 6.

3. Non-response study and sharp bounds

3.1 Sharp bounds for conditional mean

The general theory of sharp bounds for conditional means was developed in Horowitz and Manski (1998 and 2000). In particular, they developed sharp bounds for the case in which covariate X is subject to censoring. In our case, where a set of administrative covariates is used, there is no missingness in X, and tighter bounds can be obtained.

Let $s \in S$ be the strata used in the sampling process. Let $z = 1$ if a unit is interviewed and $z = 0$ otherwise. First, all survey variables are bounded, and without loss of generality we normalize survey variable $\tilde{y}$ to take values in the unite interval. If $\tilde{y} \in [a,b]$, the appropriate transformation is: $y = (\tilde{y} - a)/(b - a)$. We derive sharp bounds on $E(y \mid x)$ that holds without any assumption on the missingness process.

By the Law of Iterated Expectations:

$$\begin{aligned} E(y \mid x) &= E(y \mid x, z=1)P(z=1 \mid x) + E(y \mid x, z=0)P(z=0 \mid x) \\ &= \{\sum_{s \in S} E(y \mid x, s, z=1)P(s \mid x, z=1)\}P(z=1 \mid x) + E(y \mid x, z=0)P(z=0 \mid x) \end{aligned} \tag{3}$$

where S is a set of survey stratas. By Bayes theorem:

$$P(z=1 \mid x) = P(x, z=1)/P(x) = \{\sum_{s \in S} P(x, z=1 \mid s)P(s)\}/P(x) \tag{4}$$

Using covariates from the administrative sources allows us to assume (a) there is no item non-response on the covariates *x*, and (b) *P(x)*, the overall population distribution *x*, is known. These assumptions are appropriate because the administrative sources we use (Population Register, Tax Authority records, Ministry of Education databases) are of a census type, and some of them are used in ICBS for providing benchmarks for survey estimation procedures (in particular, in weighting adjustment procedure), in numerous household surveys. In addition, these sources are used as a sampling frame (Population Register is used as the sample frame for ISS).

The survey reveals $P(s \mid x, z=1), s \in S$. Item non-response prevents full revelation of $E(y \mid x, s, z=1)$, but the survey does reveal that:

$$\begin{aligned} E(y \mid x, s, z=1) \in [&E(y \mid x, s, z=1, w=1)P(w=1 \mid x, s, z=1), \\ &E(y \mid x, s, z=1, w=1)P(w=1 \mid x, s, z=1) + P(w=0 \mid x, s, z=1)] \end{aligned} \tag{5}$$

where $w = 1$ if a unit reports y, and $w = 0$ otherwise. In (5), the lower bound is obtained by assuming that all unobserved values of *y* equal zero and the upper bound by assuming that they all equal one. Finally, the data reveals nothing at all about $E(y \mid x, z=0)$. Hence, this quantity can take any value in the unit interval. It follows that $LB \le E(y \mid x) \le UB$, where:

$$\begin{aligned} LB_f &= \left\{\sum_{s \in S} E(y \mid x, s, z=1, w=1)P(w=1 \mid x, s, z=1)P(s \mid x, z=1)\right\}P(z=1 \mid x), \\ UB_f &= \left\{\sum_{s \in S} [E(y \mid x, s, z=1, w=1)P(w=1 \mid x, s, z=1) + P(w=0 \mid x, s, z=1)]P(s \mid x, z=1)\right\} \times \\ &\times P(z=1 \mid x) + P(z=0 \mid x) \end{aligned} \tag{6}$$

where LB_f and UB_f are lower and upper bounds, respectively, calculated for full sample (respondents and non-respondents). LB_f and UB_f in (6) are conditional means

of y given x. The width of the interval between the bounds (6) reflects both item ($w = 0$) and survey ($z = 0$) non-response.
For respondent's data, formula (6) for sharp bounds may be simplified:

$$\begin{aligned} LB_r &= \sum_{s \in S} E(y \mid x,s,w=1) P(w=1 \mid x,s) P(s \mid x) \\ UB_r &= \sum_{s \in S} [E(y \mid x,s,w=1) P(w=1 \mid x,s) + P(w=0 \mid x,s)] P(s \mid x) \end{aligned} \tag{7}$$

where LB_r and UB_r are lower and upper bounds, respectively, calculated for respondents only. The width of the interval between the bounds (7) reflects item ($w = 0$) non-response only, and this allows carrying out item non-response analysis. This interval is tighter than the interval obtained by (6). Note that the width of the interval between the bounds (6) and (7) does not depend on the sample size. We return the estimated bounds to the original variable scale after the estimation.
Upper and lower sharp bounds are asymptotically normal, because they are sample means (see Imbens and Manski, 2004, for details about asymptotic properties of the estimated bounds). It follows, that for the full sample, with probability 0.95:

$$LB_f - 1.96\sigma_{LB_f} \le E(Y \mid X) \le UB_f + 1.96\sigma_{UB_f} \tag{8}$$

where σ_s is standard deviation of sample statistic *S*. The confidence interval (8) expresses both identification uncertainty, which rises from the fact that we do not know the answers of non-respondents, and statistical uncertainty, which results from drawing a finite sample. It follows, that every point estimator for $E(y \mid x)$ should lie in interval (8); otherwise, such estimator is unacceptable because it lies outside logically possible values of the conditional mean. We estimate the standard deviations for bounds (6) and (7) using bootstrap methodology. The algorithm is as follows (for a stratified survey):
Step 1: Consider each strata of the sample design separately and draw a pseudo-sample of the correct size by random sampling with replacement from the available strata data (respondents and non-respondents).
Step 2: Combine the strata-specific pseudo-samples to form the overall pseudo survey. Compute upper and lower bounds for full data and for respondent's data only.
Step 3: Repeat steps 1 and 2 *B* (sufficiently large) times to build up a bootstrap distribution of the above statistics, and calculate an estimate for standard deviation:

$$\hat{\sigma}_S = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(S - \bar{S})^2} \,. \tag{9}$$

where $\bar{S}$ is a mean of a statistic *S*, and $\hat{\sigma}_S$ is an estimated standard deviation of *S*. Now we can substitute upper and lower bounds and their standard deviations by the appropriate estimators.
The initial sample weights are used for calculation of $E(y \mid x,s,z=1,w=1)$ in (6) and $E(y \mid x,s,w=1)$ in (7). Clearly, in this case a bias may be introduced in upper and lower bounds for respondent data only (7). However, using the adjusted for non-response weights may be misleading, because weights calibration procedures applied in ICBS require MAR assumption (weight's adjustment procedure due to non-response in ISS will be described in section 4). In addition, in the current study we are interesting in a width of the interval between the bounds; its value does not depend on the conditional mean (see (11) and (12) below).
Estimated sharp bounds for survey data are of special interest for economic research. Firstly, the bounds provide information about the distribution of conditional mean in the

case where MCAR and MAR assumptions are implausible. Secondly, they allow "best-case" - "worst case" analysis of socio-economic data for policy implications.

3.2 Statistical tests for MCAR and MAR assumptions

3.2.1 Testing MCAR assumption

Define:

$$\begin{aligned} I_f(y \mid x) &= UB_f - LB_f \\ I_r(y \mid x) &= UB_r - LB_r \end{aligned} \tag{10}$$

The explicit expression for I_f is given by:

$$I_f(y \mid x) = \left\{ \sum_{s \in S} P(w=0 \mid x,s,z=1) P(s \mid x,z=1) \right\} P(z=1 \mid x) + P(z=0 \mid x) \tag{11}$$

and for I_r by:

$$I_r(y \mid x) = \sum_{s \in S} P(w=0 \mid x,s) P(s \mid x) \cdot \tag{12}$$

Let $\hat{I}_f$ and $\hat{I}_r$ be survey estimators for width between the bounds in (10) and (11). $\hat{I}_f$ and $\hat{I}_r$ are sample means (because they are differences between two sample means over the same sampled units). Hence, I_f and I_r are also asymptotically normal. Their standard deviations can also be estimated by using bootstrap technique, equation (9).

Let *x* be categorical covariate, getting values $x_1,...,x_k$. Let the null hypothesis be: The probability to be non-respondent does not depend on *x*. In this case, the widths of the intervals as defined in (10), conditional on *x*, will be the same.

Hence, we can formalize: $H_0 : I_f(y \mid x = x_i) = I_f(y \mid x = x_j), \forall i,j \in [1,k]$. The intervals are calculated over two independent sub-samples of size n_i and n_j (for example, men and women), and the fact that I_f is a sample mean allows us to apply two-sample t-test for means with unequal variances (Snedecor and Cochran, 1989). For our case, the test statistic is given by:

$$t = \frac{\hat{I}_f(y \mid x = x_i) - \hat{I}_f(y \mid x = x_j)}{\sqrt{s^2(1/n_i + 1/n_j)}} , \tag{13}$$

where

$$s^2 = [((n_i - 1)\hat{\sigma}^2_{I_f(y|x=x_i)} + (n_j - 1)\hat{\sigma}^2_{I_f(y|x=x_j)})/(n_i + n_j - 2)] .$$

The null hypothesis is two-sided, and it will be rejected with significance level α if $|t| > t_{\alpha/2,\nu}$, where $t_{\alpha/2,\nu}$ is the critical value of the t-distribution with ν degrees of freedom. The value of parameter ν is given by:

$$\nu = \frac{(\hat{\sigma}^2_{I_f(y|x=x_i)}/n_i + \hat{\sigma}^2_{I_f(y|x=x_j)}/n_j)^2}{(\hat{\sigma}^2_{I_f(y|x=x_i)}/n_i)^2/(n_i - 1) + (\hat{\sigma}^2_{I_f(y|x=x_j)}/n_j)^2/(n_j - 1)} \tag{14}$$

If the null hypothesis is rejected, for some *i j*, we conclude that the probability to be non-respondent depends on *x*. In particular, MCAR assumption is violated in the researched survey. In order to test independence of the probability to item non-respondent on *x*, substitute I_f by I_r in (13). In this case, the null hypothesis is given by: $H_0 : I_r(y \mid x = x_i) = I_r(y \mid x = x_j), \forall i,j \in [1,k]$. To check interview non-response only, we will chose a variable without item non-response, and apply (13).

3.2.2 Testing MAR assumption

I. Testing (2a) MAR assumption can not be checked using the survey data alone, because survey reported data reveal nothing about the distribution of missing data $P(y \mid x, z=0)$. Following Dorsett (2004) we propose to check the MAR assumption by linking the survey data with available administrative data.

Let y^* be a variable from the administrative database, which is strongly correlated with key survey variable *y*, and/or with survey topic. Hence, y^* may be treated like a survey variable with no missing values, since it's values are known for all sampled units. Under the MAR assumption, the respondent's data is sufficient to estimate conditional population distributions for all survey variables, and in particular of y^*. This statement justifies linking administrative data that is relevant for survey topic to survey respondent's data, in order to decrease survey costs and response burden (see Jenkins, Lynn, Jackle and Sala, 2005 and Buck, Jenkins and Laurie, 2006). It follows that under MAR assumption, conditional on covariate *x*, the non-respondent's data provides no additional information about the population distribution of y^*. The null hypothesis is given by: $E(y^* \mid x, z=0) = E(y^* \mid x, z=1)$, and the t-test statistic is:

$$t = \frac{\hat{E}(y^* \mid x, z=1) - \hat{E}(y^* \mid x, z=0)}{\sqrt{s^2(1/n_r + 1/n_n)}} \tag{15}$$

$$s^2 = [((n_r - 1)\hat{\sigma}_r^2 + (n_n - 1)\hat{\sigma}_n^2)/(n_r + n_n - 2)],$$

where $\hat{\sigma}_r^2$ and $\hat{\sigma}_n^2$ are estimated variances of $y^* \mid x$, for respondents and non-respondents respectively, n_r is a number of respondents, and n_n is a number of non-respondents. The null hypothesis will be rejected with significance level α if $|t| > t_{\alpha/2,\nu}$, where $t_{\alpha/2,\nu}$ is the critical value of the t-distribution with ν degrees of freedom. Calculation of ν is similar to (14). Conditional on survey strata *s*, statistic (14) allows checking the MAR assumption. If the null hypothesis is rejected, survey non-response distribution turns out to be depending on survey topic and on the survey variable, which contradicts the MAR condition (2a). In this case, we have to conclude that true missing value generating mechanism is NMAR.

II. Testing (2b) Let $X = \{x_1, \ldots x_k\}$ be a set of observed survey design covariates. By (2b), conditioning on *X* will create stratas homogenous with respect to survey non-response. In particular, *X* "undoes" bias in survey variables, because the survey and item non-response in each strata assumed to be random. Let x^* be categorical observed covariate, getting values $x_1^*, \ldots, x_m^*$, when x^* is known for both respondents and non-respondents. Suppose also that *x** is orthogonal to *X*. Assuming MAR and conditional on *X*, the survey non-response rates should be independent of *x**. It follows, that under MAR, given the survey design based on *X,* the null hypothesis for testing MAR assumption is : $I_f(y \mid X, x^* = x_i^*) = I_f(y \mid X, x^* = x_j^*)$, for overall non-response, and $I_r(y \mid X, x^* = x_i^*) = I_r(y \mid X, x^* = x_j^*)$ for item non-response. The two-sided t-test should be applied, where the test statistic is calculated by (13). Rejection of the null hypothesis means, that MAR assumption in (2b) is violated, conditionally on the specific set of survey design covariates, and utilizing x^*, instead of X, in the survey design will be compatible with the MAR assumption.

We use the width of interval between the sharp bounds as a test statistic for checking MCAR and MAR assumptions, due to additional information provided by these statistics. The values of $\hat{I}_f$ and $\hat{I}_r$ represent identification uncertainty introduced to conditional mean due to both types of non-response. Notably, this test does not require any assumption regarding the true missing data generating mechanism. In addition, $\hat{I}_f$ and $\hat{I}_r$ have desirable asymptotic properties.

4. Israeli Social Survey methodology

4.1 Israeli Social Survey

The Israeli Social Survey (ISS) has been conducted annually since 2002 on a sample of persons aged 20 and older. The main purpose of the ISS is to provide up-to-date information on welfare of Israelis and on their life conditions. The ISS is the first survey conducted by the ICBS using national Population Register as a sampling frame. For localities with more than 7,500 persons aged 20 or older, one-stage systematic random sample is drawn. For localities having fewer then 7,500 listed residents aged 20 or older, two-staged systematic random sample is drawn, where the first stage involves sampling of localities, whereas on the second stage a sample of persons from the frame is drawn. This sample design is based on combination of three demographic variables: five population groups (Arabs in East Jerusalem, Arabs outside East Jerusalem, immigrants who arrived since 1990 , immigrants who arrived before 1989, Israeli-born Jews), seven age groups (20-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+), men and women. This gives a total of 70 design groups. No imputations of the survey design variables are done. In order to adjust the ISS data to the non-response influences, the following actions are carried out.

4.2 Weights and weighting calibration

The initial weights are calculated as an inverse of sample probability, for each sampled person. The weights' adjustment due to non-response is carried out as follows:

1. Strata with high interview non-response rate (more then 40%) is merged with similar strata (in sense of sample design covariate's distribution within the strata). The initial weights in the high non-response strata are multiplied by a coefficient of excessive non-response.
2. Post-stratification is carried out, by some key parameters: demographic (sex, age group, population group, immigrants), geographic groups and localities types.
3. Final weights are computed using the raking method (Deville, Sarndal and Sautory, 1993), in which the distributions of basic ISS variables are adjusted to three external distributions of demographic variables in the Population Register, and distribution of the key labor status variables estimated from the Labor Force Survey.

The raking procedure changes the weights obtained in step (1) iteratively, adjusting to each of the three mentioned above marginal distributions (but not to their intersection) until the convergence criterion is achieved. The above procedure minimizes the bias caused by survey non-response and the discrepancies between the ISS estimators on the one hand, and the Labor Force Survey estimators and the demographic figures published by the ICBS on the other.

Sampling errors are calculated using Tailor's method, after additional post-stratification using locality's size.

5. Empirical results

Our study is based on the 2006' ISS data. The sample size was 9,499 persons. 562 persons were found to be not belonging to the survey population, so the sample of eligible units included 8,937 persons. Of those, 1,648 did not respond (18.4 percent of the eligible sample).

Four key ISS variables were chosen for our analysis:

1. Worked last week (coded as '1' - worked last week, '2' - serving regular army service, '3' - serving reserve army service, '4' - did not work last week). There was no item non-response n this variable;
2. Expectations on general life conditions in near future, dubbed as 'General Optimism' – (coded as '1' - my life will get better, '2' - my life won't change, '3' - my life will get worse). Item non-response rate is 11.0 percent;
3. Gross (monthly) salary from all jobs (10 income categories, from '1' - NIS 2,000 or less, to '10' - more then NIS 14,000). Item non-response rate is 5.3 percent;

In current study, we use three administrative covariates which are highly correlated with survey topic or/and some important survey variables:

1. Gross (monthly) income from work (5 categories, by income quintiles);
2. Work status ('1' - salaried employee, '2' self-employed, '3' unemployed or not participating in labor force);
3. Degree of religiosity, defined only for Jews ('1' - secular, '2' - religious, '3' - Ultra-orthodox).

The values of these variables are linked to the survey sample by the national PIN, which is straightforward in the ISS case because it is framed by the Population Register where these PIN exist.

The first two variables, work status and income, are derived from individuals' tax returns filed with the Tax Authority. It has been proven they are highly correlated with variables reported in the surveys (see Furman, 2005), after controlling for the differences in definitions. Figure 1 shows point estimates and estimated bounds for the survey variable: Gross salary from all jobs (Y axis) conditioning on the administrative covariate: Gross income from work (X axis).

Degree of religiosity, derived from the information on type of school attended by person's children or person himself, based on administrative School and Students Registers, is highly correlated with the self-reported degree of religiosity (see Portnoi, 2007). Figure 2 shows point estimators and estimated bounds for the survey variable "Degree of religiosity " (Y axis), where: '1' - Ultra-Orthodox (Haredi), '2' - Religious, '3' – Traditional-religious, '4' – Traditional- secular, '5' Secular, conditioning on the administrative variable Degree of religiosity (X axis).

We calculate the point estimators using ISS calibrated weights. Figures 1 and 2 show that the survey variables "Gross salary from all jobs" and "Degree of religiosity " are strongly correlated with their administrative counterparts. Comparing the width of the intervals between lower and upper bounds, we can see (on Figure 1) that identification uncertainty due to overall non-response is somewhat higher for persons with low income At the same time, we observe (on Figure 2) that the interval widens dramatically with an increase in the degree of religiosity, i.e. overall non-response uncertainty is much higher among the Ultra-Orthodox Jews. The obtained sharp bounds for conditional mean are informative in all cases.

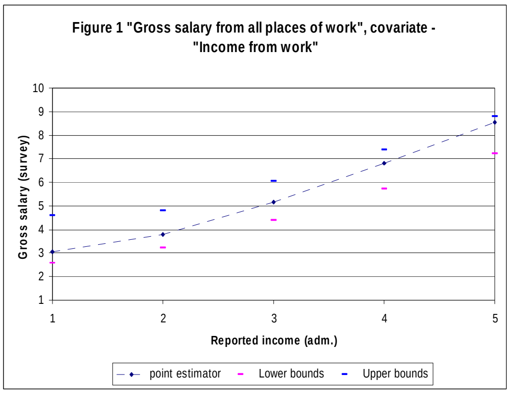

Figure 1 "Gross salary from all places of work", covariate - "Income from work"

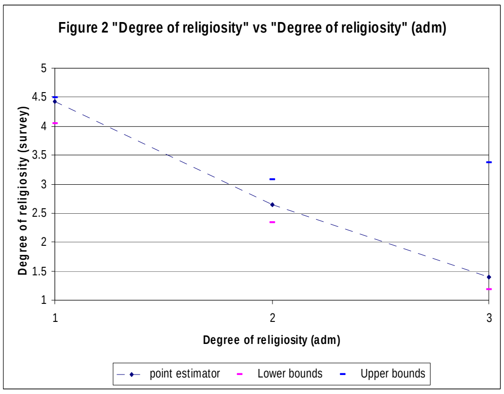

Figure 2 "Degree of religiosity" vs "Degree of religiosity" (adm)

5.1 Testing MCAR

In Table 1, we check survey and item non-response rates conditional on the administrative covariates, using the test described in section 3.2.2. Table 1 presents test statistics I_f for and I_r for survey and item non-response, respectively, and their standard deviations. In the last column, Table 1 presents minimum p-value for test (13), and the test for which it was obtained. For example, '<0.0001, 1 vs. 5' means that $H_0 : I_f(y\,|\,x_i = 1) = I_f(y\,|\,x_j = 5)$ is rejected with significance level <0.0001, and testing every other combination x_i, x_j will provide lower significance level (higher *I* type error probability). If the minimum p-value<0.05, we will conclude that the conditional non-response distribution significantly depends on the covariate *x*. Values of I_f and I_r indicate the level of non-response rates: higher value of the test statistics indicates higher non-response rates in the appropriate cell.

Table I shows, that there is a significant dependence of interview and item non-response distributions on three researched administrative covariates. The interview non-response probabilities dramatically increase with increase in degree of religiosity, with very high non-response rates for Ultra-orthodox Jews (about 50% of all interview non-respondents). Persons who did not work in 2006 according to the Tax Authority records (either unemployed or non-participants in labor force) have significantly higher probability to be non-respondent. Persons in the 3-rd income quintile have the highest interview non-response rate, and persons in the 5-th quintile - the lowest. Item non-response rates for survey variable "Gross salary from all jobs" significantly increase with an increase in income. Conditional on work status and degree of religiosity,

Table 1: Testing MCAR assumption in ISS 2006 data

Adm. Covariate	Value	Interview non-response				Item non-response						
		N_f	Working last week			N_r	Gross salary from all places of work			Optimism - general		
			I_f	Std(I_f)	Min p-value		I_r	Std(I_r)	Min p-value	I_r	Std(I_r)	Min p-value
Income from work	1	1019	0.3867	0.0354	<0.0001 3 vs 5	888	0.3517	0.0820	<0.0001 1 vs 5	0.2040	0.0231	<0.0001 1 vs 3
	2	1020	0.3588	0.0327		898	0.3624	0.0645		0.1737	0.0184	
	3	1019	0.4122	0.0321		879	0.3889	0.0638		0.1525	0.0181	
	4	1018	0.3566	0.0312		897	0.5761	0.0814		0.1628	0.0182	
	5	1020	0.3294	0.0291		908	0.5740	0.0778		0.1586	0.0181	
Work Status	1	4466	0.3701	0.0138	<0.0001 1 vs 3	3915	0.3826	0.0329	<0.0001 1 vs 3	0.1630	0.0088	<0.0001 1 vs 3
	2	410	0.3439	0.0460		363	1.2741	0.1763		0.2314	0.0353	
	3	4061	0.7757	0.0196		3011	0.5576	0.0850		0.2929	0.0117	
Degree of Religiosity	1	5028	0.3156	0.0126	<0.0001 1 vs 3	4499	0.5254	0.0413	<0.0001 1 vs 3	0.2098	0.0088	<0.0001 1 vs 3
	2	1217	0.5374	0.0322		999	0.4568	0.0819		0.3043	0.0220	
	3	1152	1.4948	0.0425		578	0.5471	0.1156		0.3287	0.0296	

the highest non-response rates are obtained for the self-employed persons and Ultra-Orthodox Jews.

Analysis of item non-response for survey variable "General Optimism" shows that the non-response rates decrease with an increase in income. Persons not working in 2006 exhibit the highest non-response rates for this question, whereas the employees have the lowest non-response propensity. As in interview non-response case, the non-response rates increase with an increase in the degree of religiosity. We conclude, that the probability to be interview or item non-respondent significantly depends on person's income, work status and degree of religiosity. Conditional on chosen administrative covariates, item non-response distributions differ between the analyzed survey questions. Thus, the MCAR assumption hypothesis should be rejected in this case.

5.2 Testing MAR

In the ISS data, MAR assumption is in place, conditional on population group, age group and sex. In Table 2 we check this assumption by applying the test described in section 3.2.2.

Conditional on the sample design covariates, means of three administrative variables were calculated for the full sample. We select 5 stratas from original survey design: 112 - Israeli-born Jews, men, 25-34 years old; 225 - Jews - immigrants who arrived by 1989, women, 55-64 years old; 314 - Jews - immigrants who arrived in 1990 and later, men, 45-54 years old; 413 - Arabs outside East Jerusalem, men, 35-44 years old and 422 - Arabs outside East Jerusalem, women, 25-34 years old. Table 2 shows that there are significant differences between the respondents' and non-respondents' distributions of the three administrative covariates.

The ISS non-respondents are characterized by a lower income and a higher rate of non-workers, relatively to the respondents. For Jewish population (stratas 112, 225, 314), the percent of religious and Ultra-Orthodox Jews among the non-respondents is higher than among the respondents. As mentioned above, these administrative covariates are strongly correlated with key ISS variables and with survey topic.

Applying test 3.2.2 (II) to ISS data allows checking validity of the MAR assumption, according to which the sample design variables listed in 4.1 should control for non-response bias in the key survey variables. The results are shown in Table 3. Statistical tests were not carried out for cells with less then 10 observations.

For survey non-response (variable "Working last week"), the null hypothesis $I_f(y \mid X, x^* = x_i^*) = I_f(y \mid X, x^* = x_j^*)$ is rejected in all cases, with the significance level for Jews higher than for Arabs. Conditional on income from work, no clear trend in the test statistics (and, consequently, in the non-response rates) can be detected. In strata 112, persons with low and high income tend to be non-respondents, whereas in strata 314 all non-respondents are in 3-rd income quintile. Conditional on the work status, persons who do not work have the highest non-response rates, and the self-employed have the lowest non-response rates, for all stratas. Among Jewish population, the Ultra-Orthodox Jews have the highest non-response rates.

Item non-response study shows that the null hypothesis $I_r(y \mid X, x^* = x_i^*) = I_r(y \mid X, x^* = x_j^*)$ for variables "Gross salary from all jobs" and "General Optimism" is rejected for all stratas of Jewish population, conditioning on each of three administrative covariates. For Arab population, conditioning on income from work and on work status, no significant dependence of values of $I_r(y \mid X, x^*)$ was detected for the variable "Gross salary from all jobs". For the variable "General

Table 2: Estimated means of administrative covariates, by strata

Strata	N resp	N non-resp	Income from work			Work Status			Degree of Religiosity		
			Respondents	Non-respondents	p-value	Respondents	Non-respondents	p-value	Respondents	Non-respondents	p-value
112	477	136	3.2470	3.2111	0.0055	1.5625	1.6876	<0.0001	1.4173	1.5846	<0.0001
225	235	36	3.3188	3.0877	<0.0001	2.1324	2.2832	<0.0001	1.2525	1.9048	<0.0001
314	95	27	3.2249	3.0000	0.2318	1.3337	2.8503	<0.0001	1.1802	2.0146	<0.0001
413	115	13	3.2677	2.8763	0.0008	1.5631	1.7353	0.0001	---	---	---
422	133	22	2.5654	2.2700	0.0007	2.2244	2.1927	0.3195	---	---	---

Table 3: Testing MAR assumption in ISS 2006 data

Strata	Adm. Covariate	Value	Interview non-response				Item non-response							
			N_f	Working last week			Gross salary from all places of work				Optimism - general			
				I_f	Std(I_f)	Min p-value	N_r	I_r	Std(I_r)	Min p-value	N_r	I_r	Std(I_r)	Min p-value
112	Income from work	1	62	0.6290	0.1572	0.0005 2 vs 5	37	2.3736	0.1978	<0.0001 1 vs 5	49	0.2449	0.0959	<0.0001 1 vs 5
		2	78	0.5769	0.1338		56	0.1607	0.1493		63	0.1270	0.0611	
		3	103	0.5825	0.1188		79	0.3418	0.1576		83	0.2169	0.0704	
		4	108	0.5833	0.1137		84	0.2143	0.2293		87	0.1609	0.0559	
		5	97	0.6495	0.1343		70	0.3857	0.2216		76	0.1316	0.0543	
	Work Status	1	417	0.6259	0.0586	<0.0001 1 vs 2	299	0.2408	0.0957	<0.0001 1 vs 3	330	0.1697	0.0299	<0.0001 1 vs 3
		2	31	0.2903	0.1601		27	0.6667	0.3057		28	0.2143	0.1189	
		3	165	0.8364	0.1035		78	0.4615	0.1719		119	0.2185	0.0546	
	Degree of Religiosity	1	415	0.6217	0.0586	<0.0001 2 vs 3	291	0.2784	0.0957	<0.0001 1 vs 3	329	0.1398	0.0276	<0.0001 1 vs 2
		2	113	0.5044	0.1031		82	0.2195	0.1732		94	0.2553	0.0703	
		3	85	1.0941	0.1551		31	0.8710	0.2481		54	0.3333	0.1011	
225	Income from work	1	22	0.5455	0.2515	<0.0001 3 vs 4	11	0.0000	0.0000	0.0291 4 vs 5	18	0.3333	0.1830	<0.0001 3 vs 4
		2	22	0.0000	0.0000		17	0.0000	0.0000		22	0.1818	0.1258	
		3	24	0.1250	0.1290		20	0.0000	0.0000		23	0.0870	0.0809	
		4	25	0.7200	0.2695		18	0.5000	0.5315		19	0.4211	0.1888	
		5	27	0.3333	0.1790		19	0.9474	0.6530		24	0.3333	0.1553	
	Work Status	1	111	0.3514	0.0871	<0.0001 1 vs 3	79	0.3418	0.2056	<0.0001 1 vs 3	98	0.2449	0.0653	<0.0001 1 vs 3
		2	9	---	---		6	---	---		8	---	---	
		3	154	0.4675	0.0873		30	0.9000	0.5054		130	0.4154	0.0685	
	Degree of Religiosity	1	204	0.2353	0.0560	<0.0001 1 vs 3	95	0.5684	0.2321	0.0162 1 vs 2,3	188	0.3617	0.0541	<0.0001 1 vs 2
		2	38	0.3158	0.1522		14	0.0000	0.0000		34	0.2941	0.1166	
		3	29	1.6552	0.2750		5	---	---		13	0.3077	0.2025	
314	Income from work	1	8	---	---	0.1434 3 vs 1,2,4,5	5	---	---	0.0027 3 vs 4	8	---	---	0.0009 3 vs 4
		2	14	0.0000	0.0000		10	0.9000	0.8625		14	0.1429	0.1603	
		3	26	0.2308	0.1527		24	0.3750	0.3584		24	0.3333	0.1519	
		4	22	0.0000	0.0000		22	0.8182	0.5577		22	0.1818	0.1335	
		5	13	0.0000	0.0000		13	0.0000	0.0000		13	0.0000	0.0000	
	Work Status	1	79	0.0759	0.0515	<0.0001 1 vs 3	70	0.5143	0.2398	0.0176 1 vs 3	77	0.1818	0.0673	0.0033 1 vs 3
		2	4	---	---		4	---	---		4	---	---	
		3	39	1.9231	0.2294		6	---	---		14	0.2857	0.2042	
	Degree of Religiosity	1	70	0.1714	0.0832	<0.0001 1 vs 3	56	0.4821	0.2679	0.0078 1 vs 3	66	0.2121	0.0797	<0.0001 1 vs 3
		2	9	---	---		5	---	---		5	---	---	
		3	26	1.6154	0.2769		10	1.0000	0.9769		12	0.5000	0.2457	
413	Income from work	1	11	0.0000	0.0000	<0.0001 2 vs 5	8	---	---	0.3405 3 vs 2,4	11	0.3636	0.2431	0.0009 1 vs 3
		2	22	0.6818	0.2728		15	0.0000	0.0000		17	0.2353	0.1626	
		3	23	0.2609	0.1764		21	0.4286	0.4389		21	0.0952	0.0917	
		4	19	0.0000	0.0000		16	0.0000	0.0000		19	0.0000	0.0000	
		5	24	0.2500	0.1684		21	0.4286	0.4592		22	0.0000	0.0000	
	Work Status	1	82	0.2561	0.0915	<0.0001 1 vs 3	66	0.0000	0.0000	0.2043 2 vs 1,3	75	0.0800	0.0447	<0.0001 1 vs 2
		2	17	0.3529	0.2523		15	1.2000	0.9144		15	0.2667	0.1689	
		3	29	0.4138	0.1840		10	0.0000	0.0000		25	0.0800	0.0784	
422	Income from work	1	15	0.4000	0.2686	0.0480 2 vs 3	7	---	---	1.0000	13	0.1538	0.1375	0.2852 1 vs 3
		2	19	0.3158	0.2086		15	0.0000	0.0000		17	0.1176	0.1214	
		3	19	0.4737	0.2627		14	0.0000	0.0000		16	0.0000	0.0000	
		4	7	---	---		4	---	---		5	---	---	
		5	0	---	---		0	---	---		0	---	---	
	Work Status	1	59	0.4576	0.1392	0.0177 1 vs 3	39	0.0000	0.0000	1.0000	50	0.1200	0.0660	0.0219 1 vs 3
		2	1	---	---		1	---	---		1	---	---	
		3	95	0.4105	0.1023		4	---	---		85	0.0976	0.0452	

Optimism", the values of $I_r(y \mid X ,x^*)$ depend significantly on work status, and, for strata 413, on income from work.

As in the interview non-response case, significance levels for Arab population are lower than for Jewish population. We conclude that there is a significant departure from the MAR assumption in the ISS 2006 data:

1. There are significant differences between respondents' and non-respondents' distributions of three administrative covariates: income from work, work status and degree of religiosity. These covariates are strongly correlated with key ISS variables and may be linked to the whole survey sample.
2. Conditionally on the ISS design variables—population group, religion, sex and age—non-response distribution depends on the orthogonal to the above set covariates (such as degree of religiosity of Jews), and on the covariates which are strongly correlated with the ISS variables. For Arab population, departures from MAR are less significant than for Jews.

Findings (1) and (2) contradict MAR definitions (2a) and (2b), respectively, therefore MAR assumption in the ISS is rejected in "local" sense: the ISS survey design variables do not control for bias in the researched variables, and the probability to be non-respondent depends on the researched ISS variables.

6. Conclusions and discussion

In the current study, non-parametric statistical tests for checking MCAR and MAR assumptions are proposed. The main advantages of using the width between sharp bounds for conditional mean as a test statistic are as follows:

1. No assumptions are made regarding the true missing data generating mechanism in the specific survey;
2. The width of the interval between estimated sharp bounds, which is used in the current study as a test statistic, provides additional information about identification uncertainty due to non-response. This information, together with estimated sharp bounds, is of great interest in applied policy-oriented research.
3. Reliable estimators for sufficiently small cells can be obtained due to independence between the interval width and sample size;

Linking survey data with administrative databases of census type provides valuable information about the non-respondents. , This information may be utilized for non-response studies and survey design improvement. In particular, using administrative covariates allows checking MAR assumption, which is untestable using the survey data alone. The proposed methodology, applied for the Social Survey, can be generalized to the household surveys, with dwellings as sampled units. In this case, the neighborhood can be characterized, and variables like average income, rate of working persons, dominant degree of religiosity can by calculated.
We find that in the 2006 survey the non-response propensity varies significantly between population groups assumed to be homogenous according to the survey design. However, MAR assumption can be restored by incorporating these administrative covariates in the survey design.

Acknowledgement We are grateful to Charles F. Manski and John V. Pepper for their helpful comments.

References

Abraham, K.G., Maitland, A., and Bianchi, S., (2006) *Nonresponse in the American time use survey*, Public Opinion Quarterly, Vol. 70, No. 5, pp. 676-703
Chiu, P.L., Riddick, H., and Hardy, A.M., (2001) *A comparison of characteristics between late/difficult interviews in the National Health Interview survey*, Proceedings of Annul Meeting of the American Statistical Association

Deville, J.C., Sarndal, C.E., and Sautory, O., (1993) *Generalized raking procedure in survey sampling*, Journal of American Statistical Association, Vol. 88, No 423, pp. 1013-1020
Dixon, J., (2002) *The effects of item and unit non-response on estimates of labor force participation*, Proceedings of Annul Meeting of the American Statistical Association
Dixon, J., (2004) *Using census match data to evaluate models of survey nonresponse*, Proceedings of Annul Meeting of the American Statistical Association
Dorsett, R., (2004) Using matched substitutes to adjust for nonignorable non-response: an empirical investigation using labor market data, *Policy Studies Institute*, London, UK
Central Bureau of Statistics, (2006) Social Survey 2006, *Questionnaire and Methodology*, CBS internet publications, Israel
Furman, O., (2005) Comparative analysis of wages and work indicators from the income tax records of wage earners, ICBS Technical paper No. 14, (in Hebrew)
Groves, R.M., and Couper, M.R., (1998) Nonresponse in household interview surveys, John *Willey and Sons Inc*., New York, USA
Hawala, S., (2008) *Partially synthetic data via expert knowledge, modeling and matching*, Proceedings of Annul Meeting of the American Statistical Association
Horowitz, J.L., and Manski, C.F., (1998) *Censoring of outcomes and regressors due to survey nonresponse: identification and estimation using weights and imputations*, Journal of Econometrics, vol. 84, pp. 37-58
Horowitz, J.L., and Manski, C.F., (2000) *Nonparametric analysis of randomized experiments with missing covariates and outcome data*, Journal of the American Statistical Association, vol. 95, No. 449
Imbens, G.W., and Manski, C.F., (2004) *Confidence intervals for partially identified parameters*, Econometrica, Vol. 72, No. 6, pp. 1845-1857
Jenkins, P.S., Lynn, P., Jackle, A., and Sala, E., (2005) *Linking household survey and administrative record data: what should the matching variables be?* University of Essex, Discussion Paper No. 489
Little, R.J., (1982) *Models for nonresponse in sample surveys*, Journal of the American Statistical Association, vol. 95, No. 449
Pfefferman, D., and Sikov, A., (2008) *Estimation and imputation under nonignorable nonresponse with missing covariate information*, Falk Institute of Economic Research Publications, Discussion Paper No. 08.06
Portnoi, H., *Characterization of the Jewish population by level of religiosity based on linkage to educational institutions*, ICBS Technical paper No. 19 (in Hebrew)
Riphahn, R., and Serfling, O., (2005) *Item non-response on income and wealth questions*, Empirical Economics, Vol. 3, No. 2
Rubin, D.B., (1976) *Inference and missing data*, Biometrica, Vol. 63, pp. 581-592
Schechtman, E., Yitzhaki, S., and Artzev, Y., (2005) *Who does not respond in the household expenditure survey: an exercise in extended Gini regressions*, CBS Working paper No. 12, Israel
Sherman, R.P., (2000) *Tests of certain types of ignorable non-response in surveys subject to item nonresponse or attrition*, American Journal of Political Science
Snedecor, G.W., and Cochran, W.G., (1989) Statistic methods, *Eighths Edition*, *Iowa State University Press*, USA
Vartivarian, S., and Little, R.J., (2003) *On the formation of weighting adjustment cells for unit non-response*, The University of Michigan Department of Biostatistics Working Paper Series, No. 10
Wang, J.H., (2004) *Non-response in the Norwegian business tendency survey*, Statistics Norway, Publication 2004/3