Response representativeness and measurement profile risks in the LFS

Barry Schouten and Kasper Leufkens

Discussion paper (201104)

Statistics Netherlands

The Hague/Heerlen, 2011

Explanation of symbols

.	= data not available
*	= provisional figure
**	= revised provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2010–2011	= 2010 to 2011 inclusive
2010/2011	= average of 2010 up to and including 2011
2010/’11	= crop year, financial year, school year etc. beginning in 2010 and ending in 2011
2008/’09–2010/’11	= crop year, financial year, etc. 2008/’09 to 2010/’11 inclusive

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands - Grafimedia

Cover
TelDesign, Rotterdam

Information
Telephone +31 88 570 70 70
Telefax +31 70 337 59 94
Via contact form: www.cbs.nl/information

Where to order
E-mail: verkoop@cbs.nl
Telefax +31 45 570 62 68

Internet
www.cbs.nl

ISSN: 1572-0314

60083201104 X-10

Response representativeness and measurement profile risks in the LFS

Barry Schouten and Kasper Leufkens

Summary: We investigate the relationship between nonresponse error and measurement error as a function of a number of survey design features. Both types of survey error are quantified using indicators. Nonresponse error is analysed in terms of maximal nonresponse bias. Measurement error is decomposed into measurement profile risk and response bias. A measurement profile is a certain response style or behaviour.

In this study measurement profiles are based on observed differences between employment registers and the Labour Force Survey. Response representativeness and measurement error are investigated as a function of the number of interviewer visits, as a function of whether an appointment was made with the household for a visit, and for proxy reports and self reports.

The proposed indicators can be used for the evaluation of survey quality over time, the monitoring of data collection and may be used as objective functions in adaptive survey designs.

Keywords: Nonresponse bias; Response bias; Response propensity

1. Introduction

In many regular survey settings only indirect evidence of measurement errors and nonresponse errors can be found. The reason for this is the very existence of the survey itself; the observations made in the survey are not recorded elsewhere. The

Dutch Labour Force Survey (LFS) is an exception as it can be linked to a number of government registers and administrative data that are informative on the employment status of the respondent and his or her search for a (new) job. As such it allows for an investigation of both survey errors. Similar analyses were made by Isaksson et al. (2008), Kleven et al. (2008), and Villund (2009).

In this paper, we investigate the relationship between nonresponse error and measurement error. Nonresponse error is investigated in terms of the maximal nonresponse bias, i.e. a worst-case assessment of nonresponse bias, see Schouten et al. (2009). Measurement error is decomposed into what we have termed measurement profile risk and the response error. The measurement profile risk is the probability that a respondent will show a certain response style that may lead to response errors. Response errors are the actual differences between the observed and true values. Measurement profiles are in this paper operationalized as differences between registers and the LFS. We view such differences as indicators of response error. We distinguish two measurement profiles, which we have termed social desirability and satisficing.

Just like the LFS itself, the employment registers may suffer from measurement errors. We do not assume that the registers hold true values, but we do assume that these errors do not relate to response behaviour in the LFS and that they are stable in time. These assumptions are not testable, but allow us to use the registers as a benchmark.

Response representativeness and measurement risk profiles are evaluated as a function of the number of interviewer visits, as a function of whether an appointment was made with the household for a visit, and for proxy reports and self reports. In the Dutch LFS proxy reporting is allowed in order to increase the response rate and to reduce travelling costs. Around 40 percent of the responses are obtained by proxy reporting. Interviewers make six contact attempts at maximum per household. If the interviewer comes at an inconvenient time, it is possible to schedule an appointment. The purpose of the analysis is to provide insights into the relationship between both types of error and into the impact of the survey design features on response quality. We aim at developing a framework that allows for generalizations beyond survey specific questionnaire items.

The indicators for nonresponse and response error may serve three purposes. First, the indicators may be used to assess the quality of a survey in time, e.g. over months, quarters or years. Such an assessment may support analyses of survey redesigns. Second, the indicators may be used during data collection as quality monitoring tools. Third, the indicators may form the input to decisions in adaptive survey designs (Wagner 2008, Schouten and Calinescu 2010).

In section 2 we define the concepts of maximal nonresponse bias and measurement risk profiles. In section 3 we provide background information on the LFS and the available registers. In section 4 we apply concepts of section 2 to the LFS. In section 5 we discuss the operationalization of measurement risk profiles. We end with a general discussion in section 6.

2. Maximal nonresponse bias and measurement profile risk

In this section we define the two concepts behind our analysis, maximal nonresponse bias and measurement profile risk.

We assume simple random sampling and compare the survey response to the total sample. The survey design may be included by means of inclusion probabilities or design weights. For ease of notation we leave these quantities out. Let $i = 1,2,3,\ldots,n$ be the sample unit labels, i.e. n is the size of the sample.

Let R_i be the 0-1 indicator for response of unit i with 1 indicating a response and 0 a nonresponse. The response propensity for unit i is denoted by ρ_i. Furthermore, let y_i be the true value of some survey variable Y for sample unit i.

We are interested in the absolute bias of the response mean

$$\bar{y}_R = \frac{1}{R}\sum_{i=1}^{n} R_i y_i \,, \tag{1}$$

where R is the size of the response.

It can be shown that the absolute bias of (1) is approximately equal to the covariance of the survey variable and response propensities divided by the average response propensity or response rate

$$|B(\bar{y}_R)| = |E(\frac{1}{R}\sum_{i=1}^{n} R_i y_i) - \bar{y}| = \frac{|\operatorname{cov}(y,\rho)|}{\bar{\rho}}, \tag{2}$$

with $\bar{y}$ is the sample mean and $\bar{\rho}$ the response rate. The absolute bias can be bounded from above by

$$|B(\bar{y}_R)| = \frac{|\operatorname{cov}(y,\rho)|}{\bar{\rho}} \le \frac{S(\rho)S(y)}{\bar{\rho}}\,, \tag{3}$$

where $S(\rho)$ and $S(y)$ are standard deviations. After dividing by $S(y)$ we get the so-called *maximal absolute nonresponse bias* B_m, see Schouten et al. (2009),

$$\frac{|B(\bar{y}_R)|}{S(y)} \le \frac{S(\rho)}{\bar{\rho}} = B_m(\rho)\,. \tag{4}$$

The maximal nonresponse bias represents a worst case bias following from strongest possible correlation between response and survey variable.

Instead of defining measurement error directly, we define an intermediate concept, the *measurement profile*. A measurement profile is comparable to a response style, e.g. social desirability, acquiescence, extreme answers, midpoint responding (see Baumgartner and Steenkamp (2001)). Tourangeau and Rasinski (1988) introduced four phases, Interpretation and comprehension, Information retrieval, Judgment, and Report, that are followed by respondents in answering survey questions. Linked to these four phases, several authors have come up with response styles, the most well-known being satisficing (Krosnick 1991). Response styles are insufficiencies in one or more of the four phases. Satisficing means for instance that respondents shortcut the first three phases and move to the fourth phase directly. Acquiescence is a special form of satisficing where the respondent always provides positive answers.

When the respondent shows a certain measurement profile then his or her answers are prone to measurement errors. The motivation for defining such profiles is that we want to abstract from measurement error on survey specific variables. There are many survey themes, many survey data collection strategies and many survey questionnaires. As a consequence, there is little scientific value in investigating the measurement error of one survey item in a specific survey. The measurement risk profile is not linked to individual survey variables but rather is an overall mood or tendency of the respondent.

Let M_i be the 0-1 indicator for a certain measurement profile. The probability that sample unit i shows the profile is θ_i, which we will call the *measurement profile risk*. When $M_i = 0$, then there is no measurement error caused by the specific measurement profile. When $M_i = 1$, then a measurement error ε_i^Y for variable Y may occur as consequence of the measurement profile. Clearly, the measurement error itself is survey variable specific.

In the survey $\tilde{y}_i = y_i + M_i \varepsilon_i^Y$ is observed instead of y_i. The bias of the response mean now changes to

$$|B(\bar{\tilde{y}}_R)| = |B(\bar{y}_R) + E(\frac{1}{R}\sum_{i=1}^{n} R_i M_i \varepsilon_i^Y)|, \qquad (5)$$

the sum of the bias of the true response mean plus the expected average measurement error. It is easy to show that in fact (5) is approximately equal to

$$|B(\bar{\tilde{y}}_R)| = |B(\bar{y}_R) + E(\frac{1}{R}\sum_{i=1}^{n} R_i M_i \varepsilon_i^Y)| = |B(\bar{y}_R) + \frac{\mathrm{cov}(\rho\theta, \varepsilon^Y)}{\bar{\rho}} + \bar{\theta}\bar{\varepsilon}^Y|, \qquad (6)$$

where $\mathrm{cov}(\rho\theta, \varepsilon^Y)$ is the covariance between the product of response propensities and measurement profile risks, and the measurement errors. The absolute bias in (6) can again be standardized and bounded from above by assuming optimal correlations

$$\frac{|B(\bar{\tilde{y}}_R)|}{S(y)} = \frac{|B(\bar{y}_R) + \frac{\mathrm{cov}(\rho\theta, \varepsilon^Y)}{\bar{\rho}} + \bar{\theta}\bar{\varepsilon}^Y|}{S(y)} \leq \frac{|B(\bar{y}_R)| + \frac{|\mathrm{cov}(\rho\theta, \varepsilon^Y)|}{\bar{\rho}} + \bar{\theta}|\bar{\varepsilon}^Y|}{S(y)}$$
$$\leq B_m(\rho) + \frac{S(\rho\theta)}{\bar{\rho}}\frac{S(\varepsilon^Y)}{S(y)} + \bar{\theta}|\bar{\varepsilon}^Y|. \qquad (7)$$

In the upper bound to the absolute bias we find five terms of which two terms are survey item specific and three are not. The general terms are the maximal absolute bias, $B_m(\rho)$, the ratio of the standard deviation of the product of response and measurement risk propensities and the response rate, $S(\rho\theta)/\bar{\rho}$, and the average measurement profile risk, $\bar{\theta}$. We will refer to the second term as the maximum interaction bias $I_m(\rho,\theta)$, i.e.

$$I_m(\rho,\theta) = \frac{S(\rho\theta)}{\bar{\rho}}. \tag{8}$$

The maximum interaction bias represents the combined impact of nonresponse and measurement risk. It will be zero if there is no variation in the product of the propensities. Hence, if nonresponse and measurement error are unrelated the propensities will have the tendency to attenuate each other and the product will have less variation. However, when respondents have a higher risk of measurement error, then the variation will become sharper and the maximal interaction bias bigger. Surveys without measurement error do not have an interaction bias. The profile risks θ_i are equal to zero and, as a consequence, the standard deviation in (8) equals zero. However, in surveys with full response the interaction bias may still exist as the profile probabilities may differ for population units.

The survey item specific terms in (7) are the ratio of standard deviations of the measurement error and survey item itself, $S(\varepsilon^Y)/S(y)$, and the absolute mean measurement error, $|\bar{\varepsilon}^Y|$. Since we are not interested in the survey item specific terms we will mostly ignore them in the following sections.

It is straightforward to distinguish more than one measurement profile. For instance, respondents may be prone to give socially desirable answers but may also show acquiescence behaviour. The two profiles may appear simultaneously but also separately.

Suppose, we assume K measurement profiles, labelled as $k = 1,2,3,\ldots,K$. Corresponding indicators are denoted by M_i^k. θ_i^k is the risk that unit i will have measurement profile k. Now, (6) and (7) change to

$$|B(\bar{\bar{y}}_R)| = |B(\bar{y}_R) + \sum_{k=1}^{K} \frac{\operatorname{cov}(\rho\theta^k, \varepsilon^{k,Y})}{\bar{\rho}} + \sum_{k=1}^{K} \bar{\theta}^k \bar{\varepsilon}^{Y,k}| \tag{9}$$

and

$$\frac{|B(\bar{\bar{y}}_R)|}{S(y)} \le B_m(\rho) + \sum_{k=1}^{K} I_m(\rho,\theta^k)\frac{S(\varepsilon^{Y,k})}{S(y)} + \sum_{k=1}^{K} \bar{\theta}^k |\bar{\varepsilon}^{Y,k}|. \tag{10}$$

In section 4, for illustration, we will distinguish two profiles, one profile we link to social desirable answers and one profile we link to trying to reduce the interview time.

3. The LFS and employment registers

In this section we give an exploratory analysis of differences between the Labour Force Survey (LFS) and employment registers. This analysis is the starting point for the definition of measurement profiles in sections 4.2 and 4.3.

The objective of the LFS is to provide reliable information about the labour market. For the Dutch LFS a sample of addresses is drawn each month. The sampling frame is a list of all occupied addresses in the Netherlands, which is derived from the municipal registration of population data. The LFS is based on a stratified two-stage cluster design of addresses. Strata are formed by geographical regions. Municipalities are considered as primary sampling units and addresses as secondary sampling units. All households residing at an address, up to a maximum of three, are included in the sample (in the Netherlands, there is generally one household per address). Since most target parameters of the LFS concern people aged 15 through 64 years, addresses with only persons aged 65 years and over are undersampled. Furthermore, households with persons aged between 14 and 26 or from non-western countries are oversampled. For 2008 this lead to a total survey sample of 76,401 addresses. The persons living at these addresses form the first wave of the LFS.

In the first wave, data are collected by means of computer assisted personal interviewing (CAPI). All sampled addresses receive a letter in which it is stated that an interviewer will come by in the upcoming weeks in order to interview the household members. Interviewers make six contact attempts per household at maximum. If the interviewer comes at an inconvenient time, it is possible to schedule an appointment for the interview. In the Netherlands participation in the LFS is not mandatory. Households can thus refuse to participate in the survey.

For all members of participating households, demographic variables are observed. For the target variables only persons aged 15 years and over are interviewed. When a household member cannot be contacted, proxy interviewing is allowed by members of the same household. Households, in which one or more of the selected persons do not respond for themselves or in a proxy interview, are treated as nonresponding households. The respondents aged 15 through 64 years are re-interviewed four times at quarterly intervals. In these four subsequent waves, data are collected by means of computer assisted telephone interviewing (CATI). During these re-interviews a condensed questionnaire is applied to establish changes in the labour market position of the respondents. Proxy interviewing is also allowed during these re-interviews.

In this paper we only consider the first wave data of 2008. Furthermore we only consider persons aged 15 through 64. Although the LFS sample is based on addresses, it is possible to convert this into persons using the municipal registration of population data. At some of the sampled addresses nobody is living and some of the addresses do not exist. In 2008 the total sample of persons for which response was sought consisted of 135,332 persons. Interviewers obtained response for 78,321 persons.

Table 3.1: Fractions of contact attempts, appointments made and proxy interviewing

Contact attempt	*Sample*	*Response*
0	2.4	0.6
1	38.9	43.1
2	26.8	29.0
3	14.4	14.6
4	7.4	7.0
5	4.1	3.1
6	6.0	2.5
Appointment		
no	70.0	56.5
yes	30.0	43.5
Proxy response		
no	-	62.7
yes	-	37.3

Table 3.1 shows the fraction of contact attempts made by interviewers for both the LFS sample and the response. Sometimes persons contact Statistics Netherlands after receiving the announcement letter. Most of them do not want to participate and a few want to make an appointment. For these persons the number of contact attempts by the interviewers is zero. In the remainder of this paper, we do not show the results for zero contact attempts as this group is very small.

The average number of contact attempts is lower for respondents than for the complete sample. In Table 3.1 we also show the fraction of appointments that were made. Clearly, appointments were mainly made by respondents. Table 3.1 also shows that 37.3 percent of the response was obtained by proxy interviewing.

In order to investigate survey errors, we augment the LFS with data from the POLIS register. For all persons living in the Netherlands this register contains the amount of money received from working in employment and from social benefits. The register does not contain income from self-employment. Using this register we can determine whether a person was working in employment and the number of jobs this person had.

In Table 3.2 the fraction of people working in employment based on the POLIS register is given for the Dutch population, the LFS sample, and the LFS response. For LFS respondents we can also determine whether they are working as an employee based on their answers in the survey. This fraction is shown in Table 3.2. The same is done for employees with more than one job. Table 3.2 contains their fraction in the Dutch population, the LFS sample, and the LFS response.

Table 3.2: Fraction of people working as an employee and with multiple jobs

	Employed		*Multiple jobs*	
	POLIS	*LFS*	*POLIS*	*LFS*
Population	66.3	-	5.0	-
Sample	67.2	-	5.1	-
Response	69.2	67.3	5.3	5.3

From table 3.2 we conclude that there is a difference of almost 2% in the proportion of persons that are employed, while the proportion of persons having multiple jobs is similar in size. The fractions are averaged over the response. Hence, these results do not preclude that on the individual level there may be differences that cancel out over respondents.

Besides people working in employment, we also focus on people that are looking for a job. The Centre for Work and Income[1] (CWI) assists (unemployed) people in finding a job. In order to receive an unemployment benefit, one has to register here. Also other people without a job that want to work can register at CWI.

Table 3.3 contains the fraction of people registered at CWI for the Dutch population, the LFS sample, and the LFS response. In the LFS people are asked whether they are registered at the Centre for Work and Income. Table 3.3 also contains the fraction of respondents that answer this question affirmatively. There is a considerable difference between the fraction of respondents that say there are registered at CWI (5.0%) and the fraction of them that can be found in the CWI register (3.7%). We consider this difference to be caused by socially desirable answers, i.e. some respondents indicate that they have subscribed themselves to CWI because they do not want to make a bad impression.

Table 3.3: Fraction of people registered at the Centre for Work and Income

	Registered at CWI	
	CWI	*LFS*
Population	4.1	-
Sample	4.1	-
Response	3.7	5.0

4. Application

We, first, investigate how the nonresponse bias of register variables evolves as a function of three design features, the number of visits, the possibility to make appointments and proxy reporting. Second, we analyse the occurrence of differences between the LFS and registers as a function of the same design features as an exploratory analysis of measurement profiles. Third, we combine the two types of error and investigate their relation by plotting maximal nonresponse bias, maximal response risk and measurement profile risk as a function of design features. Fourth and final, we elaborate on the impact of single variables on nonresponse and measurement profiles using so-called partial R-indicators.

4.1 Nonresponse bias in register variables

In this section we show how the nonresponse bias of the register variables evolves as a function of the three selected survey design features: the number of visits, the

[1] The centre for work and income is nowadays called Public Employment Service (UWV).

possibility to make appointments and proxy reporting. For this purpose we have replaced the LFS answers by the POLIS and CWI data. In other words we treat the POLIS and CWI variables as survey questions. This way we can explore the impact of nonresponse separate from measurement differences.

Table 4.1.1 shows how the fraction of respondents working as a POLIS employee evolves when the maximum number of contact attempts grows. The columns represent four different strategies: 1) no proxy and appointments allowed, 2) no proxy allowed but appointments are accepted, 3) no appointments allowed but proxy is accepted, and 4) both proxy and appointments allowed. The sample fraction of POLIS employees, i.e. including nonresponse, equals 67.2%.

Table 4.1.1: The fraction of employees as a function of the maximum number of visits, proxy interviewing and appointments.

Number of	*Proxy, appointment*			
attempts	*No, No*	*No, Yes*	*Yes, No*	*Yes, yes*
1	56.5	59.1	62.5	64.4
2	61.1	63.6	65.3	67.3
3	63.0	65.3	66.6	68.4
4	64.2	66.1	67.3	68.9
5	64.7	66.4	67.5	69.1
6	65.0	66.7	67.7	69.2
Sample	67.2			

In general the fraction of persons working as an employee grows with the number of contact attempts. For all contact attempts, the fraction of employees increases when we take into account response obtained after making an appointment or from proxy interviewing is taken into account. Especially proxy interviewing appears to boost the fraction. These results can be explained by the fact that employed people are less often at home than persons who do not work.

Table 4.1.2 shows how the fraction of CWI registered persons evolves when the maximum number of contact attempts is increased. Again the fraction is shown as a function of proxy interviewing and scheduled appointments. The sample fraction equals 4.1%

Table 4.1.2: The fraction of CWI registered persons as a function of the maximum number of visits, proxy interviewing and appointment scheduled

max. contact attempts	no proxy no appoint	no proxy	no appoint	
1	6.0	5.6	4.5	4.1
2	5.3	5.0	4.1	3.8
3	5.1	4.8	4.0	3.7
4	5.1	4.8	4.0	3.7
5	5.1	4.8	4.0	3.7
6	5.1	4.8	4.0	3.7
sample	4.1			

In general the fraction of respondents registered at the CWI decreases with the number of contact attempts. For all contact attempts the fraction decreases when response was obtained after an appointment was made or from proxy interviewing. People registered at the CWI do not work and are therefore more likely to be at home when an interviewer visits the household.

We now turn to the impact of nonresponse in terms of maximal nonresponse bias as defined in (4). We modelled response propensities in a logistic regression using six auxiliary variables: type of household, age, ethnicity, urbanization degree of residence, average house value in zip code and an indicator for paid job. The indicator for paid job is derived from the POLIS register. Age and average house value are included in the model as categorical variables.

The selected six variables are used in all assessments of representativeness at Statistics Netherlands. They are selected because literature and experience at Statistics Netherlands has shown that they generally relate to nonresponse. Their selection is thus not LFS specific. To aid comparison the variables are all entered to the model; no variable selection was made.

Figure 4.1.1: The maximal nonresponse bias after 1, 2, 3, 4, 5 and 6 or more contact attempts. Diamonds represent response without appointment and proxy. Triangles is response without proxy. Filled circles is response without appointment. Solid circles present bias for all response.

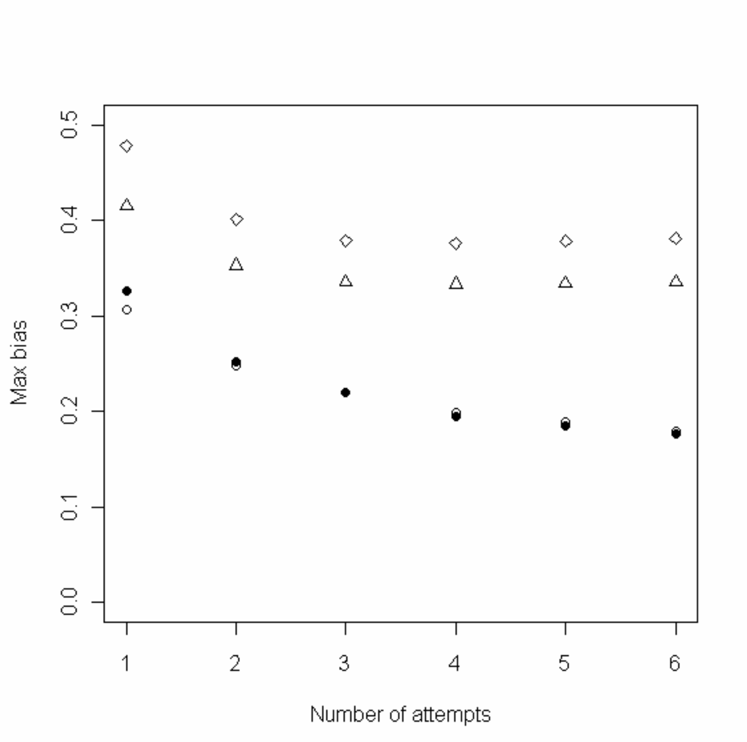


Figure 4.1.1 plots the maximal nonresponse bias against the number of contact attempts for

1. Sample units that respond themselves without an appointment
2. Sample units that respond themselves possibly after an appointment
3. Sample units that respond without an appointment possibly by proxy
4. Sample units that responds possibly after an appointment and possibly after by proxy

Respondents under 2 include the respondents under 1 plus the respondents that required an appointment. Similarly, respondents under 3 include the respondents under 1 plus the proxy respondents. Respondents under 4 consist of all respondents. Hence, response is both cumulative in the number of contact attempts as in appointment and proxy.

From figure 4.1.1 we can conclude that all design features have a positive effect on nonresponse bias, that is; the maximal nonresponse bias decreases. The impact of proxy reporting is strongest and also the number of contact attempts shows a clear decrease from one to three attempts. Appointments only lead to a moderate decrease of maximal bias. This decrease vanishes almost completely when proxy is allowed. A larger number of contact attempts correspond in all cases to a smaller bias. However, after three attempts the decrease in bias is small; implying that for nonresponse bias doing more than four contact attempts is not necessary.

Figure 4.1.2 presents a so-called response-representativity plot, see Schouten and Bethlehem (2009). It plots R-indicators against response rates for the various design choices. The R-indicator is a simple transformation of the standard deviation of estimated response propensities $R(\rho) = 1 - 2S(\rho)$. It takes values between zero and one, where a value of one corresponds to optimal representativeness and a value of zero to maximal deviation from representative response. Representative response is defined as equal response propensities. The lines in figure 4.1.2 correspond to different maximal bias levels. The top line is 10% and the bottom line is 40% maximal bias.

From 4.1.2 we conclude that (not surprisingly) all design features give a higher response rate. However, the representativeness indicator goes down between the first and third contact attempt. When appointments are allowed then there is a further decrease in representativeness. Proxy reporting has a positive effect on the variation in response propensities. When proxy reporting is allowed, then the variation in response propensities is reduced.

Summarizing, all design features lead to a higher precision of survey estimates but only the number of visits and proxy reporting lead to an improvement of response representativeness. Hence, these two design features are interesting to consider as treatments in adaptive survey designs when we would restrict focus to the quality of survey response. In section 4.3 we will also add survey data quality.

4.2 Differences between LFS and registers

In this section we take the analysis of section 3 a step further in order to derive measurement profiles.

For LFS respondents we can determine whether they are working as an employee on the basis of their answers in the survey. We compare this with their status according to the POLIS register. For 2008 we find that 2,374 respondents work as an employee according to the LFS but are not registered as such in the POLIS. Vice versa, there are 3,818 respondents in the POLIS who are not employees according to the LFS. Out of 78,321 respondents we find differences between the LFS and POLIS register from 6,192 persons. For these respondents we say there is an employee error. In section 4.3 we view these errors as a form of satisficing behaviour.

Figure 4.1.2: Response-representativity plot for number of contact attempts. Diamonds is response without appointment and proxy. Triangles is response without proxy. Filled circles is response without appointment. Solid circles is all response. Straight lines correspond to maximal bias of 10, 15, 20, 25, 30, 35 and 40%.

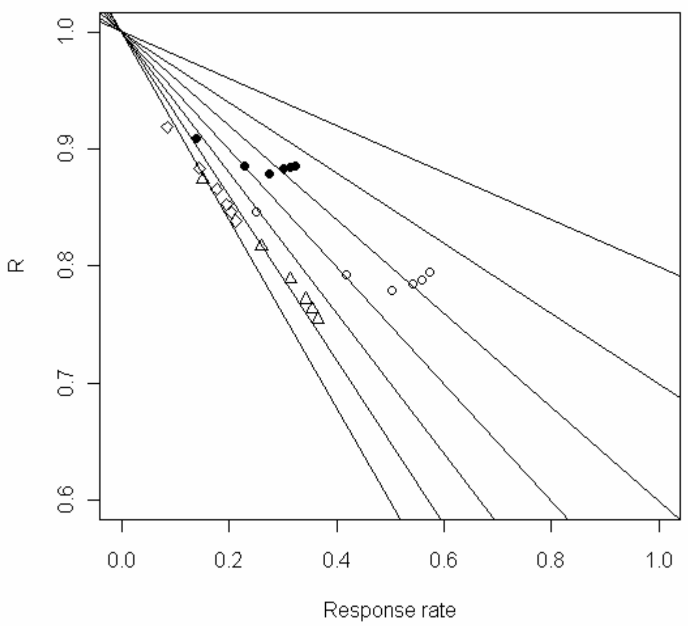


Table 4.2.1 shows the fraction of respondents for which there is an employee error as a function of maximum contact attempts, proxy interviewing and appointment scheduling. In general, the fraction of employee error remains constant over the maximum number of contact attempts. Only from one to two maximum contact attempts, the fraction decreases in all columns. Including data from appointments marginally lowers the fraction. Data from proxy interviews leads to an increase in the fraction of employee errors.

Table 4.2.1: The fraction of employee error as a function of the maximum number of visits, proxy interviewing and appointment scheduled

max. contact attempts	no proxy no appoint	no proxy	no appoint	
1	6.5	6.5	7.6	7.7
2	6.3	6.2	7.5	7.6
3	6.3	6.2	7.5	7.6
4	6.3	6.2	7.4	7.5
5	6.3	6.2	7.5	7.6
6	6.3	6.2	7.5	7.6

In the LFS respondents are asked whether they are registered at the CWI. We compare the answer to this question with the status according to the CWI register. We find that 1,593 respondents answered that they were registered at CWI but cannot be found in the CWI register. Vice versa, there were 573 respondents in the CWI register, who answered that they were not registered at CWI. Out of 77,646 respondents, we find a difference between the LFS and CWI register for 2,166 persons. For these respondents we say there is a CWI error. In section 4.3 we view these errors as caused by socially desirable answers.

Table 4.2.2 shows the fraction of respondents with a CWI error as a function of maximum contact attempts, proxy interviewing and appointment scheduling. In general, the fraction CWI error remains constant over the maximum number of contact attempts. Including data from appointments marginally lowers the fraction of CWI error. Data from proxy interviews leads to a decrease of the CWI error fraction.

Table 4.2.2: The fraction of CWI error as a function of the maximum number of visits, proxy interviewing and appointment scheduled

max. contact attempts	no proxy no appoint	no proxy	no appoint	
1	3.4	3.2	3.0	2.8
2	3.2	3.1	3.0	2.8
3	3.2	3.0	2.9	2.7
4	3.2	3.0	2.9	2.8
5	3.2	3.1	3.0	2.8
6	3.2	3.1	3.0	2.8

Table 4.2.3: The fraction of employee/CWI error as a function of the maximum number of visits, proxy interviewing and appointment scheduled

max. contact attempts	no proxy no appoint	no proxy	no appoint	
1	9.6	9.4	10.4	10.3
2	9.3	9.0	10.1	10.0
3	9.2	8.9	10.1	10.0
4	9.1	8.9	10.1	10.0
5	9.2	9.0	10.2	10.1
6	9.2	8.9	10.1	10.1

For each respondent we combine the employee error and CWI error into employee/CWI error. For each respondent, this variable is 1 in case at least one of the previously mentioned errors is present and 0 otherwise. Table 4.2.3 shows the fraction of respondents for which there is an employee/CWI error as a function of maximum contact attempts, proxy interviewing and appointment scheduling.

4.3 Nonresponse bias and measurement risk profile

Although the LFS contains mostly behavioural questions, some response styles may be expected. Respondents may feel that they should have a job or, when they do not, should look for a job. In society it is considered undesirable to not have a job and live on an allowance from the government. Respondents may also lack a strong motivation to concentrate on answering questions about their employment status or feel that the survey does not apply to them because they are outside the labour force population. Such feelings may lead to acquiescence or disacquiescence or more generally to satisficing, i.e. respondents shortcut the answering process and answer promptly without consideration of the questions asked. Satisficing may also be the result of repeated questions about different jobs. When a person has multiple jobs, then questions are posed for each job separately.

Literature shows there is a clear trade off between nonresponse and measurement error when considering proxy reports. See Lemaitre (1988), Moore (1988), Dawe and Knight (2007) and Thomsen and Villund (2008) for investigations into the impact of proxy reporting on LFS key statistics. Literature concludes that the gain in response rate outweighs the increase in measurement errors due to proxy interviewing.

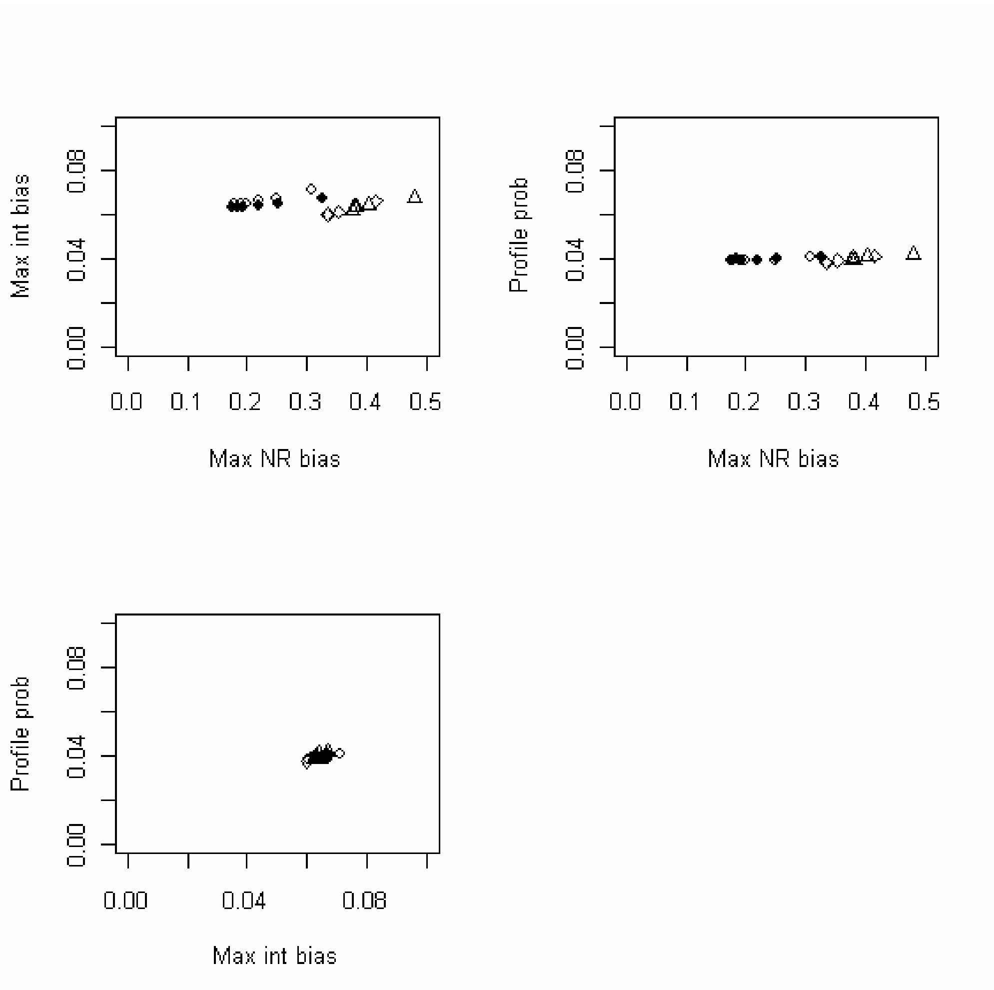
The three survey design features that we consider in this paper are expected to have different impacts on nonresponse and measurement error. The number of visits and proxy reporting mostly reduce nonresponse due to noncontact, while appointments decrease the number of refusals. We conjectured beforehand that each design feature should positively impact the representativeness of the response. There is no direct link between the number of visits and measurement error as the respondent will often not be aware of the amount of effort that was put into obtaining contact. Of course respondents that are harder to reach may be prone to certain response styles. Proxy reporting and appointments clearly may affect the occurrence of measurement errors directly. Proxy reporting involves answers from a less knowledgeable person. An appointment is made at a point in time that is convenient to the respondent. One may conjecture that appointments positively affect measurement errors while proxy reports have a negative impact.

We distinguish two measurement profiles. The first profile consists of two groups of persons: 1) persons that are not working as an employee according to the POLIS registration but respond in the LFS that they are an employee, and 2) persons that are not employed and are not subscribed to an employment office but respond that they are subscribed in the LFS. We label this profile as the tendency to provide socially desirable answers. The second profile consists of persons that are employed according to the POLIS registration but respond that they are not employed. This profile gets the label satisficing. In short:

- Profile 1 (socially desirable): Not employed in POLIS but employed in LFS, or not employed, no employment office registration but subscription to employment office in LFS
- Profile 2 (satisficing): Employed in POLIS but not employed in LFS.

The second profile is labelled as satisficing because this type of difference occurs mostly when a respondent has multiple jobs and does not mention one of the jobs that is registered as an employment.

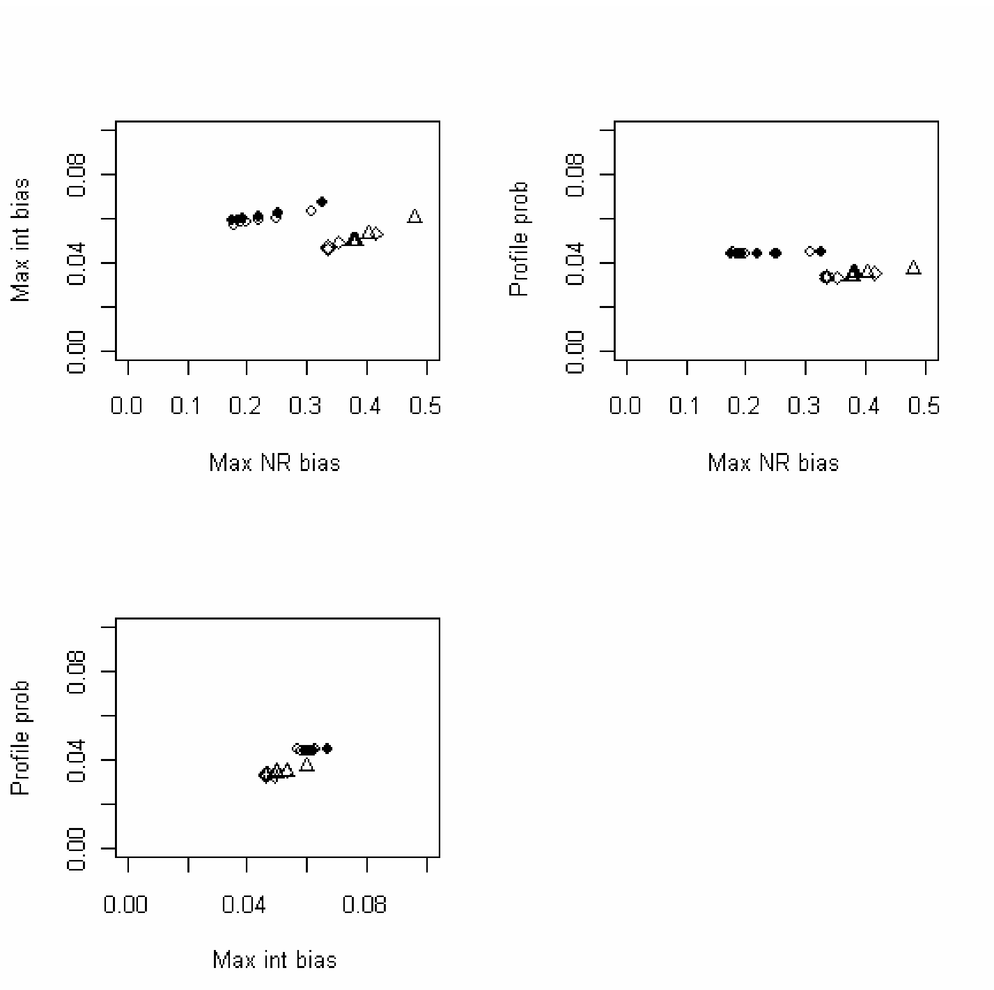
Figure 4.3.1: Maximal nonresponse bias, maximal interaction bias and average profile risk for profile Socially desirable for number of contact attempts. Diamonds is response without appointment and proxy. Triangles is response without proxy. Filled circles is response without appointment. Solid circles is all response.



The profile risks θ_i^k are estimated using the same model as for the response propensities, i.e. a model with type of household, age, ethnicity, urbanization degree of residence, average house value in zip code and an indicator for paid job.

Figures 4.3.1 and 4.3.2 provide the maximal nonresponse bias $B_m(\rho)$, the maximal interaction bias $I_m^k(\rho, \theta^k)$ and the average profile risk $\bar{\theta}^k$ for the two profiles. The plots are 2D projections of a 3D plot. In the figures the three projections to 2D are shown. The 3D plot is not shown as some points are not visible and as a consequence the plot leads to a confusing, less informative picture

Figure 4.3.2: Maximal nonresponse bias, maximal interaction bias and average profile risk for profile Satisficing for number of contact attempts. Diamonds is response without appointment and proxy. Triangles is response without proxy. Filled circles is response without appointment. Solid circles is all response.



Based on the two figures we can draw a number of conclusions. Compared to the nonresponse bias, the interaction bias and profile risks are much more stable when different design choices are made. For the first profile, socially desirable, both the interaction bias and the average profile risk are almost the same. There is a slight increase in interaction bias when proxy reporting is allowed. The average profile risk is around 4.0%. For the second profile, Satisficing, both the interaction bias and the average profile risk go up when proxy reporting is allowed. The average profile risk is around 3.5% for self-reports and 4.5% when proxy reporters are added.

4.4 The impact of single variables on nonresponse and measurement profiles

In the previous sections we restricted ourselves to overall measures of response representativeness and measurement profile risk. In this section we refine these measures and investigate the impact of single variables. We consider maximal nonresponse bias, maximal interaction bias and profile rates for the six auxiliary variables that we used in the previous sections: type of household, age, ethnicity, urbanization degree of residence, average house value in zip code area and an indicator for paid job.

The detailed analysis of the impact of single variables is needed when the indicators are used for adaptive survey designs. Such designs need targeting of population subgroups that have a relative large impact on survey error, i.e. that affect response representativeness and/or measurement profile risk. Adaptive survey designs tailor strategies to these subgroups in such a way that their contribution so survey error is minimized given constraints on costs.

We start with the profiles risks as these are most straightforward. Tables 4.4.1 to 4.4.6 show the profile 1 (socially desirable) and 2 (satisficing) rates for each of the six variables as a function of proxy and appointments (N,N = no proxy and no appointments, N,Y = no proxy but appointments allowed, Y,N = proxy allowed but no appointments, Y,Y = proxy and appointments allowed). In section 4.3 we concluded that the number of visits does not impact the profile rates. For this reason we leave this design feature out of the analysis to make tables more transparent.

Table 4.4.1: Profile rates for type of household given design features

	Profile 1				*Profile 2*			
	N,N	*N,Y*	*Y,N*	*Y,Y*	*N,N*	*N,Y*	*Y,N*	*Y,Y*
Single	3.9	3.5	3.9	3.5	3.0	2.9	3.0	2.9
Single parent	7.7	7.2	7.2	6.9	3.5	3.3	4.8	4.5
Unmarried couple	2.0	1.7	2.1	1.7	2.4	2.3	2.5	2.6
Unmarried couple + children	3.8	3.1	3.5	3.0	2.8	3.1	3.6	3.9
Married couple	2.8	2.7	2.8	2.7	4.1	4.0	4.5	4.5
Married couple + children	4.3	4.3	4.3	4.4	3.7	3.5	5.3	5.2
Other	3.9	3.7	3.7	3.6	3.2	3.3	3.6	4.0

Table 4.4.2: Profile rates for age given design features

	Profile 1				*Profile 2*			
	N,N	*N,Y*	*Y,N*	*Y,Y*	*N,N*	*N,Y*	*Y,N*	*Y,Y*
15-20	12.3	12.0	9.8	10.1	7.9	6.9	9.0	8.5
20-25	7.5	6.8	6.0	5.6	4.9	5.5	6.8	6.9
25-30	3.7	3.3	3.5	3.1	2.3	2.2	2.4	2.4
30-35	3.2	3.5	3.0	3.1	2.5	2.5	2.7	2.9
35-40	3.1	3.1	2.8	2.8	3.1	2.9	3.4	3.5
40-45	3.4	3.5	3.0	3.2	3.1	3.2	4.0	4.2
45-50	3.5	3.6	3.1	3.1	3.0	2.8	3.8	3.7
50-55	4.5	4.0	3.8	3.4	3.0	3.0	3.6	3.9
55-60	3.8	3.4	3.4	3.0	3.8	3.4	4.3	4.1
60-65	2.7	2.8	2.7	2.8	4.5	4.8	5.1	5.4

Table 4.4.3: Profile rates for ethnicity given design features.

	Profile 1				*Profile 2*			
	N,N	*N,Y*	*Y,N*	*Y,Y*	*N,N*	*N,Y*	*Y,N*	*Y,Y*
Native	3.6	3.5	3.7	3.6	3.4	3.3	4.4	4.5
1st gen western non-native	7.0	6.9	7.1	7.1	4.0	3.9	3.7	3.4
2nd gen western non-native	3.5	3.5	3.6	3.7	4.0	3.9	4.8	4.7
1st gen non-western non-native	7.1	7.0	6.7	6.4	4.0	3.8	4.4	4.3
2nd gen non-western non-native	8.4	7.8	6.4	6.2	5.0	5.4	7.5	8.0

Table 4.4.4: Profile rates for urbanization degree given design features.

	Profile 1				*Profile 2*			
	N,N	*N,Y*	*Y,N*	*Y,Y*	*N,N*	*N,Y*	*Y,N*	*Y,Y*
Very strong	4.8	4.4	4.7	4.4	3.4	3.5	3.9	4.1
Strong	4.0	3.9	3.9	3.8	3.2	3.2	4.0	4.2
Moderate	3.5	3.6	3.5	3.8	3.5	3.5	4.7	4.6
Little	4.0	3.8	4.0	4.0	4.0	3.7	5.2	4.9
Not	3.9	3.9	4.1	4.1	3.4	3.2	4.7	4.7

Table 4.4.5: Profile rates for average house value (×1000) given design features.

	Profile 1				*Profile 2*			
	N,N	*N,Y*	*Y,N*	*Y,Y*	*N,N*	*N,Y*	*Y,N*	*Y,Y*
No values available	4.0	3.5	3.8	3.7	3.3	4.1	4.6	5.3
0-75	4.8	4.5	4.5	3.9	3.6	3.9	4.5	4.9
75-100	6.0	5.9	5.7	5.7	3.0	3.0	3.2	3.4
100-150	6.1	5.8	5.9	5.7	2.9	2.8	3.3	3.2
150-200	4.3	4.2	4.3	4.1	2.2	2.3	2.9	3.0
200-250	3.3	3.3	3.4	3.6	2.9	2.6	3.5	3.5
250-300	3.6	3.4	3.5	3.4	3.3	3.1	4.3	4.2
300-400	2.9	2.8	3.3	3.4	4.5	4.3	5.8	5.7
400-500	3.3	3.7	3.6	3.9	7.9	7.1	9.8	9.2
>500	2.6	2.7	3.5	3.3	10.2	10.8	11.5	12.5

Table 4.4.6: Profile rates for employment status given design features.

	Profile 1				*Profile 2*			
	N,N	*N,Y*	*Y,N*	*Y,Y*	*N,N*	*N,Y*	*Y,N*	*Y,Y*
Not employed	11.2	11.3	12.0	12.5	0	0	0	0
Employed	0	0	0	0	5.3	5.1	6.6	6.5

From tables 4.4.1 to 4.4.6 we can conclude that profile 1 rates are especially high for single parents, young persons, non-native persons with the exception of 2nd generation western, and the persons living zip code areas with a low average house value. The profile 2 rates are high for married couples, younger persons and older persons, second generation non-western non-natives, and the persons living in zip code areas with a high average house value. By definition the profile 1 rate is zero for employed persons and the profile 2 rate is zero for unemployed persons.

The picture with respect to the impact of design features is mixed. Generally, for the profile 1 rates there is no strong impact of the design features. Exceptions are for younger persons and second generation non-western non-natives where the profile 1 rates drop when proxy is allowed. When proxy is allowed the profile 2 rates go up for single parents and married couples, younger persons, second generation non-natives, persons living in the less urbanized areas of the country and various house value groups.

Detailing nonresponse bias and interaction bias is less straightforward. For this purpose we employ so-called partial R-indicators, see Shlomo et al. (2009). Partial R-indicators decompose the variance of estimated propensities in between and within variance terms in an Analysis of Variance fashion. The numerator in the maximal nonresponse bias $B_m(\rho)$ and the maximal interaction bias $I_m(\rho,\theta)$ both are variance terms, the variance of response propensities and the variance of response times profile propensities, respectively. We use the partial R-indicators to investigate the contribution of single variables to these numerators.

There are two types of partial R-indicators: unconditional and conditional. The unconditional partial R-indicator does not correct for collinearity between variables and corresponds to the univariate impact of a variable on representativeness of response. It takes values between zero and a half, where zero means that the variable is representative and does not contribute to a deviation from representative response. Conditional partial R-indicators give the impact of single variables corrected for the impact of the other variables in the model. Again the indicator takes values between 0 and a half, and again a value of zero indicates absence of impact on representativeness of response. The conditional partial R-indicator is strictly smaller than the unconditional partial R-indicator as it is corrected for other variables. In the tables some conditional values are larger than their unconditional counterparts due to sampling variation.

The unconditional partial R-indicator for a single categorical variable is equal to the between variance of propensities when the population is stratified by the categories of that variable. The unconditional partial R-indicator is denoted by $P_u(Z)$ where Z is the variable of interest. $P_u(Z)$ represents the amount of variance attributable to Z. The conditional partial R-indicator is denoted by $P_c(Z)$ and is equal to the within variance when the population is stratified by the categories of all variables with Z excluded.

In this paper we restrict ourselves to partial R-indicators at the variable level. Shlomo et al. (2009) also present partial R-indicators at the category level. These indicators measure the contribution of categories of variables to the lack of representative response.

Partial R-indicators are designed for measuring the contribution of single variables to the lack of representative response. Here we also use them to decompose the interaction bias. Tables 4.4.7 to 4.4.10 contain partial R-indicators for nonresponse for the six variables as a function of the three design features (with maximum number of contact attempts on the rows). Tables 4.4.11 and 4.4.12 contain the partial R-indicators for the interaction between nonresponse and risk profiles.

Various conclusions can be drawn from the nonresponse partial R-indicators. As expected, the conditional partial R-indicators are smaller than the unconditional partial R-indicators. However, the decrease is small, indicating that the six variables have almost an orthogonal impact on response representativeness. The partial R-indicators are lowest when proxy is allowed but appointments are excluded. Allowing for appointments generally increases the indicators while the number of visits has a small impact. The contribution of variables changes but in most settings age is the strongest variable, i.e. collects most of the variation in response propensities.

Table 4.4.7: Nonresponse unconditional and conditional partial R-indicators for urbanization, age, ethnicity, type of household, housevalue and job status when appointments and proxy are not allowed.

		Urban	*Age*	*Ethnicity*	*Household*	*Housevalue*	*Job*
P_u	1	0.01	0.03	0.01	0.02	0.01	0.02
	2	0.01	0.05	0.02	0.04	0.01	0.02
	3	0.01	0.06	0.02	0.04	0.01	0.02
	4	0.01	0.06	0.02	0.05	0.01	0.01
	5	0.01	0.06	0.02	0.06	0.02	0.01
	6	0.01	0.06	0.02	0.06	0.02	0.01
P_c	1	0.01	0.02	0.01	0.01	0.01	0.01
	2	0.01	0.04	0.01	0.02	0.01	0.01
	3	0.01	0.04	0.01	0.02	0.01	0.01
	4	0.01	0.04	0.01	0.03	0.02	0.01
	5	0.01	0.05	0.01	0.04	0.02	0.01
	6	0.01	0.05	0.01	0.04	0.02	0.00

Table 4.4.8: Nonresponse unconditional and conditional partial R-indicators for urbanization, age, ethnicity, type of household, housevalue and job status when proxy is allowed.

		Urban	*Age*	*Ethnicity*	*Household*	*Housevalue*	*Job*
P_u	1	0.02	0.03	0.02	0.03	0.01	0.01
	2	0.03	0.04	0.03	0.04	0.02	0.01
	3	0.04	0.04	0.03	0.04	0.03	0.00
	4	0.03	0.04	0.03	0.04	0.03	0.00
	5	0.03	0.04	0.03	0.03	0.03	0.00
	6	0.03	0.04	0.03	0.03	0.03	0.00
P_c	1	0.01	0.02	0.01	0.02	0.01	0.01
	2	0.02	0.02	0.01	0.02	0.02	0.00
	3	0.02	0.03	0.02	0.02	0.02	0.00
	4	0.02	0.03	0.02	0.02	0.02	0.01
	5	0.02	0.03	0.02	0.02	0.02	0.01
	6	0.02	0.02	0.02	0.02	0.02	0.01

Table 4.4.9: Nonresponse unconditional and conditional partial R-indicators for urbanization, age, ethnicity, type of household, housevalue and job status when appointments are allowed.

		Urban	*Age*	*Ethnicity*	*Household*	*Housevalue*	*Job*
P_u	1	0.02	0.05	0.02	0.03	0.01	0.03
	2	0.03	0.08	0.03	0.04	0.02	0.02
	3	0.03	0.09	0.04	0.05	0.02	0.01
	4	0.03	0.10	0.04	0.06	0.02	0.01
	5	0.03	0.10	0.04	0.07	0.02	0.01
	6	0.02	0.10	0.04	0.07	0.03	0.00
P_c	1	0.01	0.04	0.02	0.01	0.01	0.02
	2	0.02	0.07	0.02	0.01	0.02	0.02
	3	0.02	0.08	0.03	0.03	0.02	0.01
	4	0.02	0.08	0.03	0.04	0.03	0.01
	5	0.02	0.09	0.03	0.04	0.03	0.00
	6	0.02	0.09	0.03	0.05	0.03	0.00

Table 4.4.10: Nonresponse unconditional and conditional partial R-indicators for urbanization, age, ethnicity, type of household, housevalue and job status when both appointments and proxy are allowed.

		Urban	*Age*	*Ethnicity*	*Household*	*Housevalue*	*Job*
P_u	1	0.04	0.04	0.03	0.06	0.03	0.02
	2	0.06	0.05	0.05	0.07	0.05	0.01
	3	0.07	0.05	0.06	0.08	0.06	0.01
	4	0.07	0.05	0.06	0.07	0.06	0.02
	5	0.07	0.05	0.06	0.06	0.06	0.02
	6	0.07	0.05	0.06	0.06	0.06	0.02
P_c	1	0.02	0.03	0.02	0.04	0.02	0.01
	2	0.03	0.03	0.03	0.05	0.03	0.01
	3	0.04	0.04	0.03	0.05	0.03	0.02
	4	0.04	0.04	0.03	0.04	0.04	0.02
	5	0.04	0.04	0.03	0.04	0.04	0.02
	6	0.04	0.04	0.03	0.04	0.04	0.03

Tables 4.4.11 and 4.4.12 show that generally the partial R-indicators for the interaction between profile risk and nonresponse are smaller than for nonresponse alone. This implies that the response and profile propensities attenuate each other. This picture is consistent with the overall impact of nonresponse and interaction bias. The conditional indicators are in some cases slightly bigger than the unconditional indicators, probably due to sample variation. The parrtial R-indicators increase when proxy is allowed and sometimes also when appointments are included. The strongest variable is employment status, which is not surprising as it

closely related to the definition of the two profiles. Age and household type also have an impact on risk profile 1, while for risk profile 2 all variables influence the variation in propensities except for ethnicity.

Table 4.4.11: Profile 1 interaction unconditional and conditional partial R-indicators for urbanization, age, ethnicity, type of household, housevalue and job status when NN = no appointments, no proxy, NY = no proxy, appointments, YN = proxy, no appointments and YY = proxy and appointments.

		Urban	*Age*	*Ethnicity*	*Household*	*Housevalue*	*Job*
P_u	NN	0.00	0.00	0.00	0.00	0.00	0.01
	NY	0.00	0.01	0.00	0.00	0.00	0.02
	YN	0.00	0.01	0.00	0.00	0.00	0.02
	YY	0.00	0.01	0.00	0.01	0.00	0.03
P_c	NN	0.00	0.00	0.00	0.00	0.00	0.01
	NY	0.00	0.01	0.00	0.00	0.00	0.02
	YN	0.00	0.01	0.00	0.01	0.00	0.02
	YY	0.01	0.02	0.01	0.01	0.01	0.03

Table 4.4.12: Profile 2 interaction unconditional and conditional partial R-indicators for urbanization, age, ethnicity, type of household, housevalue and job status when NN = no appointments, no proxy, NY = no proxy, appointments, YN = proxy, no appointments and YY = proxy and appointments.

		Urban	*Age*	*Ethnicity*	*Household*	*Housevalue*	*Job*
P_u	NN	0.00	0.00	0.00	0.00	0.00	0.01
	NY	0.00	0.01	0.00	0.00	0.01	0.01
	YN	0.00	0.00	0.00	0.00	0.01	0.01
	YY	0.01	0.01	0.00	0.01	0.01	0.02
P_c	NN	0.00	0.01	0.00	0.00	0.01	0.01
	NY	0.00	0.01	0.00	0.00	0.01	0.02
	YN	0.00	0.01	0.00	0.00	0.01	0.01
	YY	0.00	0.02	0.00	0.00	0.02	0.03

5. Operationalizing measurement profiles

In this paper we implemented measurement profiles as observed differences with respect to external registrations. One may question whether these differences really imply measurement profiles, i.e. whether it is justified to label this as response styles. Another important question is whether alternative operationalizations of profiles can be constructed. In this section we briefly discuss both questions. However, this paper merely launches the idea of measurement risk profiles; more research and empirical evidence are clearly needed.

In section 4.3 we labelled the observed deviances between the LFS survey and the two registers as socially desirable answering and satisficing tendencies. Do the observed differences indeed always deduce to such behaviours? The answer is negative for two reasons. First, it supposes that the registers are without error and that the definitions and reference dates in the LFS and registrations are equal. This may not be true. The registrations themselves may be subject to errors as in fact they depend on data collection too. Furthermore, there may be small differences in the definitions of employment status between LFS and registrations. The registrations are updated on a monthly basis while the LFS may be referring to any date during a month. Nonetheless, it is believed that differences in definition between LFS and registrations are small and that the registrations are accurate because of their importance to policy making. The second reason that jeopardises the definition of measurement profiles, as they are used here, is more important and of a conceptual nature. Although a respondent may fail to give a correct answer to one or two survey questions, and for these questions show the specified response style, for other questions the respondent may show different behaviour. In other words does a response style correspond to consistent behaviour that is latent throughout the entire questionnaire or may styles change. In the case of the LFS it is clear that when the employment status or the subscription to an employment office are subject to socially desirable answering or satisficing, then the respondent is forced to stick to these styles in order to be consistent. Furthermore, the computer-assisted questionnaire will select questions based on earlier answers. If the respondent would change styles then it would immediately follow that earlier answers are false. In general, however, one may assume that response styles may change when the topics of the survey change, i.e. when the questionnaire shifts to a new block of questions about a completely new topic. For instance, the survey may contain question blocks about political interest and health. Respondents may only be motivated to answer health questions and satisfice to the questions about political interest. Hence, it is recommendable that any definition of a specific response style accounts for the full variety of survey topics. Nevertheless, it seems reasonable to assume that certain response styles are indicative of the respondent’s mood or mind set and will arise through the entire questionnaire.

Measurement profiles as for representativeness of response can only be identified with additional information that is external to the survey. Without such information, undesirable response styles cannot be distinguished from true answers. Registrations and administrative data are strong examples of external data given that the quality level of the data is high. Alternatively, one may rely on paradata, i.e. data about the data collection process collected by interviewers, data collection staff or data collection systems. Examples are:

- The average time length per question, or the average length of blocks of questions
- The frequency of failures to edit rules, inconsistencies or lack of coherence in answers

- The proportion of missing items
- Coded recordings of sections of the questionnaire, e.g. via Computer Assisted Recording of Interviews (CARI)
- Audit trails of web surveys
- An interviewer assessment of the pace of the interview indicating satisficing

It is paramount, however, that paradata do not suffer from measurement errors themselves and can be linked to response styles in a natural and logical way. For example, very short periods between questions and answers indicate satisficing. Time spans can be recorded without error in computer-assisted surveys. However, if an interviewer is present, it must be accounted for that interviewers help or motivate respondents and may thus change the length of the answering process. It is thus less straightforward how to link time span to satisficing in interviewer-assisted surveys as opposed to web surveys. Interviewer assessments may be focussed on specific response styles, assuring a direct link with these styles, but the assessments may be subject to interviewer variance and errors.

Hence, although the possibilities are almost unlimited, for quality assessment the implementation of measurement profiles must be done with a lot of care and based on experience and empirical evidence.

6. Discussion

In this paper we propose quality indicators for nonresponse and measurement errors. The indicator for nonresponse error is the maximal nonresponse bias; a worst case bias assuming that nonresponse and survey outcome correlate optimally. Measurement error is split into a measurement profile risk and the actual response error. Measurement profiles are comparable to response styles. When a respondent fits the profile then there is a risk of errors for any survey question. The profile is viewed as a survey specific phenomenon but a survey item non-specific phenomenon. For measurement error two indicators are used: the maximal interaction bias and the average profile probability. The maximal interaction bias conforms to an optimal correlation between response errors and profile risks. It is again a worst case assessment. The average profile risk is the rate of respondents that shows the specified profile. The indicators are based on response propensities and profile propensities. Both are estimated using a set of auxiliary variables.

The indicators are operationalized and applied to the LFS. Two profiles are distinguished: persons that are prone to social desirable answers and persons that may show satisficing behaviour. The profiles are determined from deviations between LFS answers and information available from two registrations; the tax board registration on employment and the employment office registrations. Clearly, the labels we attached to the two profiles are subjective interpretations of the response style. They may be labelled differently by other researchers. Also they are only indirect assessments of such profiles. It remains to be discussed, therefore, if such profiles are useful concepts.

In the application we evaluated three design features: the number of contact attempts, appointments and proxy reporting. We found that all features affect the nonresponse bias positively. The strongest positive impact is of proxy reporting. Contrary, we found that only proxy reporting has a mild impact on the Satisficing measurement profile. When proxy is allowed the interaction bias and average profile probability go up. The other profile, Socially desirable, is not affected by any of the design features.

The concepts of measurement profiles can be used in survey design choices, for instance in adaptive survey designs. Adaptive survey designs assign different design features to different population units in order to optimize quality given constraints on costs or vice versa to minimize costs given constraints on quality. The quality indicators in this paper may be used as quality functions in such optimizations. However, both practically and conceptually more research is needed to evaluate the usefulness and relevance of these measures.

References

Baumgartner, H., Steenkamp, J.E.M. (2001), Respons styles in marketing research: a cross—national investigation, Journal of Marketing Research, 28, 143 – 156.

Dawe, F., Knight, I. (2007), A study of proxy response in the Labour Force Survey, *Labour force survey user guide* Vol. 1, Ch. 11, 50-59.

Isaksson, A., Lundquist, P., Thorburn, D. (2008), Use of process data to determine the number of call attempts in a telephone survey, Paper presented at Q2008, July 9 – 11, 2008, Rome, Italy.

Kleven, O., Lagerstrom, B., Thomsen, I., (2008), A simple model for studying the effect of proxy interviewing: The Norwegian labour force survey as a case study, *Statistics Norway.*

Krosnick, J.A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. Applied Cognitive Psychology, 5, 213-236.

Lemaitre, G. (1988), A look at response errors in the Labour Force Survey, *The Canadian Journal of Statistics,* Vol. 16, 127-141.

Moore, J. (1988), Self/Proxy response status and survey response quality. A review of liturature, *Journal of official statistics*, Vol 4, No. 2, 155-172.

Schouten, B., Bethlehem, J. (2009), Representativeness indicators for measuring and enhancing the composition of survey response, RISQ deliverable 9, available at www.risq-project.eu .

Schouten, B., Calinescu, M. (2010), Optimizing quality of response through adaptive survey designs, Discussion paper, CBS, Den Haag.

Schouten, B., Cobben, F., Bethlehem, J. (2009), Indicators for the representativeness of survey response, Survey Methodology, 35 (1), 101 – 113.

Shlomo, N., Skinner, N., Schouten, B., Carolina, N., Morren, M. (2009), Partial indicators for representative response, RISQ deliverable, available at www.risq-project.eu.

Thomsen, I., Villund, O. (2008) Using administrative registers to evaluate the effects of proxy interviews in the Norwegian labour force survey, working paper.

Tourangeau, R. and Rasinski, K.A. (1988). Cognitive processes underlying context effects in attitude measurement. Psychological Bulletin, 103, 299-314.

Villund, O. (2009), Measuring working hours in the Norwegian Labour Force Survey. A pilot study of data quality using administrative registers, Statistics Norway Reports 2009/3.

Wagner, J. (2008), Adaptive survey design to reduce nonresponse bias, PhD thesis, University of Michigan, USA.