1. Introduction

As one of many efforts to cope with non response, Statistics Norway has established several databases of paradata relevant to non response (Thomsen et al …).

In a previous study this data were used to study the feasibility of using proxy interviewing in the Labour Force surveys. Allowing family members to answer some labour market related questions on behalf of another member of the family, is one of the most efficient ways to reduce non response. The conclusion was that this practice should be continued in spite of a minor underreporting of labour market participation (Thomsen & Villund). In a follow-up study this conclusion was confirmed using a more comprehensive methodology (Zhang et al).

In this study, we shall follow the Norwegian Labours Force Study from 1996 to 2015. During this period, the response rate has decreased from 91 to 80 percent. We ask what the consequences of this development have for the nonresponse bias.

In order to study this we chose 7 auxiliary variables, several of which are used in the estimation process as well. All variables (multinomial and continuous) were dichotomized, thus simplifying the task, but also reducing the contained information. However, the covariation between each variable and the response

rate is essentially maintained. It is well known that the nonresponse bias is a function of this covariation (Bethlehem …).

In Section 2 we present the nonresponse biases (absolute and relative) for each of the 7 auxiliary variables.

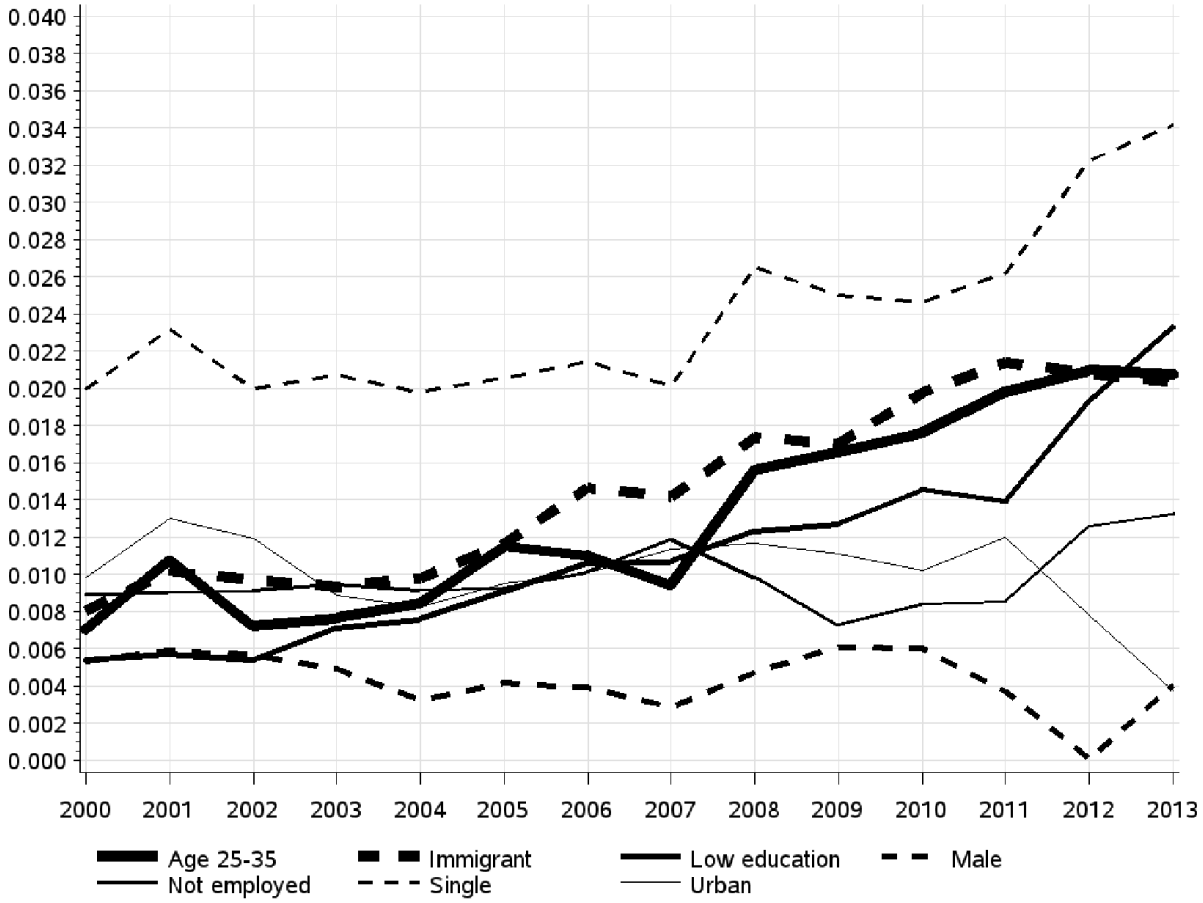
In section 3 we argue that the average of the 7 biases is a methodological sound indicator at survey level for the present study.

2. Bias of each auxiliary variable.

As a first step, we dichotomize seven auxiliary variables into seven binomial ones: $D_1, D_2, \ldots, D_7$. We estimate the nonresponse bias as the difference between the proportions observed in the net sample and the selected sample.

The results are shown in figure 1.

Figure 1: Category level indicators for selected binary variables. LFS 2000-2013



It is seen that for all variables, the bias remains positive during the whole period. We have observed similar results in other surveys, and in section 3 we shall suggest a simple and transparent indicator for the whole survey. We find from figure 1 that some biases are practically constant over time, while others increase. During the same period, the non response rate has increased about 20%. In other words the non response rate is a poor indicator of the bias.

To find a better indicator, we shall use the following expression for the bias:

$$B(\frac{n_k}{n}) = \frac{n_k}{n} - \frac{N_k}{N} = \frac{N_k}{N}\frac{r_k - r}{r} \quad (2.1)$$

Where N_k and n_k are the number of observations for which the variable $D_k = 1$ in the total sample and the response sample respectively.

$$r = \frac{n}{N} \quad , \quad r_k = \frac{n_k}{N_k} \quad ,$$

From (2.1) follows that the bias depends on two factors: $\frac{N_k}{N}$ being the sample proportion of the variable, and $\frac{r_k - r}{r}$ the relative bias. This term is the ratio between a measure of the representativity and the response rate, and is a good indicator of the bias. This is easily generalized to all categorical variables. The relative bias is shown in figure […] Figure 2: […]

Equation (2.1) is the binomial version of a more general result

$$B\left(\frac{n_k}{n}\right) = cov(D_k, R)/r_k$$

Where

$$R = \begin{cases} 1|response \\ 0|nonresponse \end{cases}$$

$$P(R = 1) = \frac{n}{N}$$

$$P(D_k = 1) = \frac{N_k}{N}$$

$$P(R = D_k = 1) = \frac{n_k}{n}\frac{N_k}{N}$$

3. Survey Level Indicator

Let

$$Z_1 = \frac{D_1 + D_2 + \cdots + D_7}{7}$$

$$Z_2 = \frac{D_1/P_1 + D_2/P_2 + \cdots + D_7/P_7}{7}$$

Z_1 and Z_2 are both positively correlated with the auxiliary varibles, and therefore with the response rate. We can therefore use the bias of Z_1 and Z_2 as survey level indicators of absolute and relative bias respectively.

$$B_1 = cov(Z_1, R) = \frac{1}{7}\sum_{k=1}^{7}\frac{N_k}{N}(\frac{r_k - r}{r})$$

$$B_2 = cov(Z_2, R) = \frac{1}{7}\sum_{k=1}^{7}(\frac{r_k - r}{r})$$

Results for the indicators are shown in figure […]

Again it is seen in this survey that the relative bias is a good indicator of the bias [???]. The two survey level indicators are simply averages of the marginal indicators. That does not mean that they are independent of the covariance between all the auxiliary variables. The distribution of Z_1 and Z_2 depends on these covariances.

4. Conclusions

- The nonresponse bias has increased over the period studied. If not at [every?] variable level, it certainly has at survey level.
- For categorical variables the ratio between a measure of representativity and the response rate seems to be a good [useful?] indicator.
- Independent of the choice of indicator, it is important to secure access to [comparable / time series?] data such that a retrospective study can be performed.
- Further work should be done to investigate how this retrospective approach may be used to design follow-up strategies [???]