

## Using Register Data to Evaluate the Effects of Proxy Interviews in the Norwegian Labour Force Survey

*Ib Thomsen<sup>1</sup> and Ole Villund<sup>2</sup>*

We combine data from the Norwegian Labour Force Survey with register data in order to evaluate the impact of proxy interviews on the survey-based employment rate estimates. The method compares estimates under different models for proxy response and nonresponse models, over a relatively long time series from 1997 to 2008. Using register-based employment as an auxiliary variable, we try to differentiate between the effect of the measurement and the effect of the fact that proxy-interviewed people are not selected at random. We label these effects “proxy effect” and “selection effect” respectively, and suggest methods for estimating them. Our conclusion, after also including the impact of nonresponse, is that proxy interviews probably result in a better employment rate estimate, even though they introduce some underreporting. The reason is that proxy interviews provide data on some hard-to-reach people who have a labour-market situation more similar to that of those not reached at all. We find that including the proxy responses has approximately the same effect as post-stratification of the direct responses, using register-employment status as the auxiliary variable.

*Key words:* Nonresponse; proxy response; nonsampling error; selection bias.

### 1. Introduction

The purpose of this study is to evaluate the impact of proxy interviews on the employment rate in the Norwegian Labour Force Survey (LFS), and is based on studies by Thomsen et al. (2007) and Kleven et al. (2006). In a proxy interview, also called “indirect interview,” one person (the proxy) answers questions on behalf of another person (the sampled subject). Self-responses, also called “direct interviews,” are responses provided by the individual. Currently proxy interviews constitute about 15 percent of the LFS response sample, while the unit nonresponse rate is about 15 percent of the total sample.

The motivation for using proxy interviews in the LFS is primarily to save working time, and thus keep costs down. To maintain the current response level in the LFS using only direct interviews, the data collection would take longer. Abandoning proxy interviewing without increasing the resources would probably increase the nonresponse rate considerably.

<sup>1</sup> Statistics Norway, Division for Statistical Methods, P.O. Box 8131, 0033 Oslo, Norway. Email: ith@ssb.no

<sup>2</sup> Statistics Norway, Division for Labour Market Statistics, P.O. Box 8131, 0033 Oslo, Norway. Email: vil@ssb.no

**Acknowledgments:** We would like to thank Dr. Jan Fredrik Bjørnstad, Head of Research, and Dr. Li-Chun Zhang, Senior Researcher, both at Statistics Norway, for valuable comments and suggestions. A number of referees and an Associated Editor have also contributed to improving this article.

It can be argued that proxy interviewing maintains the precision within the available resources, by extending the response sample size. The assumption is that a considerable number of those interviewed by proxy are less reachable or completely unreachable within reasonable time. Nonresponse as well as the proxy response is associated with several common factors such as people being young, urban and of foreign origin. Proxy response could be viewed as an intermediate between self-response and no response at all, along a dimension of “reachability.” A typical explanation for the negative association between foreign origin and self-response is language issues. Language problems could affect both the motivation for participating and the communication. Possible explanations for less reachability among young and urban people include: higher mobile-phone saturation, more frequent change of telephone number, activities outside home, smaller household size. The impact of technological developments is discussed for instance in Steeh (2008). Technological advances have affected the contact possibility in more ways than just increased it. For instance, every mobile phone displays the caller’s number, and this makes it easy to reject a call without even talking to the caller. This function together with answering machines and other technologies could be contributing to declining response rates in countries with high telephone saturation.

Another line of arguing in favour of proxy interviews is that it may reduce some nonresponse bias. We know that both nonrespondents and proxy-respondents have lower register-based employment than self-respondents. Proxy interviews may therefore include more data on nonemployed persons who might otherwise not have responded at all. One possible explanation for lower response rate among the nonemployed is that they feel less obligated to participate in a survey labelled “Labour Force.” This could arise from a feeling that their participation is not as relevant, since they are not employed; or it could be a more general negative attitude to governmental agencies. Some such fundamental factors are plausible, since the nonresponse bias for employment remains significant after controlling for obvious demographic variables such as age, gender and region.

The goal of this study is to assess the advantages and disadvantages of proxy interviews for the employment rate estimates. In Section 2 we review some previous results on proxy interviews. Section 3 describes the data sources and the linking of data. Section 4 presents a method that uses register data to evaluate the effects of including proxy interviews in the response sample. In Section 5 we include also the effect of nonresponse.

## **2. Previous Studies**

In survey methodology, proxy response is recommended as a cost-saving alternative, especially when conducting face-to-face interviews. Several studies conclude that, if proxy interviews were replaced by self-response, the result would be substantially increased cost (Moore 1988). However, proxy response is often thought to differ systematically from self-response. According to survey theory, proxy responses are less accurate and rely more on generic than episodic information in recalling facts (Groves et al. 2004). Several authors have proposed that the effect of proxy interviewing varies with the type of questions, and there is no general consensus on how seriously proxy interviewing affects the data quality.

Kalsbeek et al. (2007) and others mention a possible cognitive basis for the superior quality of self-response over proxy response. There are plausible arguments for self-response being more accurate in cases where more vivid memory or more detailed information is crucial. However, thus far we cannot conclude that the better information value of self-response gives much more accurate employment estimates.

O'Muircheartaigh (1991) studied a reinterviewed sample from the U.S. Current Population Survey, focusing on response variance. This study initially reveals evidence of better quality of proxy response than self-response. However, it also points out the self-selection of proxy respondents, as well as possible additional effect of selection made by the interviewer. The latter process may be more or less formalized. According to the Norwegian LFS protocol, parents can answer on behalf of their children, but not vice versa. Other selection processes are left to the judgement of the interviewer. One hypothesis O'Muircheartaigh proposes is that the particular survey instrument does not provide an opportunity to utilize the advantage of better information that self-respondents may have. More demanding survey questions, either more complicated or more personal, could result in a more pronounced quality difference.

In an experimental study conducted under laboratory conditions, Boehm (1989) focused on the difference in reliability between proxy responses and self-responses, as well as the agreement between the two response types. Another topic was the relationship between self-rated confidence and the actual performance. This study had the advantage of being able to control both the interview situation and selection of interview type, but had a relatively small sample size. The results show a considerable disagreement between proxy responses and self-responses, and more so among those classified as not employed (from self-response). Furthermore the proxy interviews resulted in lower reliability. There was a poor correlation on the one hand between the proxies' self-rated confidence and knowledge, on the other the accuracy of the information supplied. The study therefore concluded that screening of proxies based on self-rating would be less useful. No formal screening process is applied in the LFS, other than barring offspring from responding about their parents. It is possible that some individual screening is at work where the interviewer detects severe problems in communicating, for instance with nonnatives.

In a large-scale study, Martin and Butcher (1982) found a very high level of agreement on employment between self-response and proxy response. The nonresponse rate was 28 per cent, which could introduce a substantial response bias. Pairs of adults closely related were each interviewed both about themselves and the other person, and this should have ruled out selection bias in the interview type. Because the LFS has the same restriction that only the spouse or a parent is a possible proxy, the results are all the more relevant for our case. There was great variation in the agreement rate between different types of questions. Questions about income scored considerably lower than employment classification and attitude questions even lower.

Lemaitre (1988) reports on response errors in the Canadian LFS by examining a re-interviewed subsample. Both the initial interview and the reinterview allowed proxy response as well as self-response. That means the combined data contains some individual units with both direct and proxy responses about the same reference week. Of interest here are the topics important in order to classify a person as employed: "Had a job, did not

work” and “Worked during reference week.” There was more inconsistency when the interview and reinterview were of two different types. If both were self-response or both proxy, the inconsistency was about half that of the two different interview types. Although proxy interviews introduce some measurement errors, the study reveals that two direct interviews also produced up to 5 percent inconsistent answers. Given a correct and reliable self-response, inconsistency between different-type interviews can be attributed to proxy error. The study finds more inconsistency in proxy-proxy pairs than when both are direct interviews. However, this does not tell us about the validity of proxy interviews, but indicates lower reliability in proxy interviews than in direct ones. The inconsistency is lowest when both interviews are direct, and this indicates higher reliability.

Dawe et al. (2007) report on a proxy response study based on questions in the British LFS. First a sample of households with at least two adults was contacted and only proxy interviews were conducted. After some time, each person for whom proxy information had been given was interviewed directly. The data for the two interviews referred to the same period, and the time between interview and reinterview was kept short. The design assumes a high internal reliability for the self-response, and does not measure this explicitly. By comparing the proxy and direct responses to the same questions, three quality indicators were constructed: proportion of consistent answers; rate of missing data due to proxy interviews; estimated overall effect on the whole sample (gross error rate). The study focuses on the relative quality differences in terms of the nature of the questions and the relation between the proxy and the subject. Economic activity status (employment etc.) was one of the questions with highest quality in respect of high consistency, few missing data and low gross error rate. Questions that require more detailed or numerical answers decrease the proxy-response quality considerably. The study made a comparison of the proxy’s relationship with the subject, and found that spouses gave more consistent responses about each other than parents did about their children. This was not the case for all types of questions, but, for economic activity, a spouse as proxy was clearly better than a parent answering on behalf of an offspring.

Solheim et al. (2001) studied the effect of proxy interviews on employment for the age group 16 – 29 years, in LFS data collected in the 1st and 4th quarters of 2000. Logistic regression was used to control for age and register-employment, with separate models for students and other young people. The overall result was an underestimation of employment rate of about 1.5 percentage points. The effect was larger for young age groups and students not living at home. This result and those of Dawe et al. (2007) should warrant some attention in respect of the LFS, since disproportionately many young people are interviewed by proxy in this survey.

### **3. Data**

#### *3.1. The Survey Data*

The Norwegian Labour Force Survey (LFS) is a continuous sample survey, with a sample size of about 24,000 persons per quarter. The sampling frame is all registered resident families having one or more members between 15 and 74 years old. The sample design

is a one-stage cluster sampling where all family members between 15 and 74 years old in the sampled families are included. The mode of data collection is exclusively computer-assisted telephone interviews. The statistical unit is person, and each person is interviewed once every quarter, for eight consecutive quarters. The reference period is one week, and the interview is performed within ten days after the reference week.

Since families are sampled, proxy interviewing by family members is possible. The survey protocol allows proxy response by parent or spouse for most items, including questions on employment status. The variable of interest in this study is employment rate, defined as the number of employed persons relative to the population size of 15 – 74-year-olds. A person is defined as being employed if they worked at least one hour during the reference week. This definition is in line with recommendations from the ILO (International Labour Organization).

### 3.2. *The Register Data*

The register data used in this study are collected mainly from the Norwegian Labour and Welfare Service's employee register. Jobs are reported by the employers to the employee register; mostly directly from the employers' own IT systems. Thus, the employment status based on this register data constitutes a source of employment information fully independent of the survey responses. The definition of "register employed" is based on records with employer identification, employee identification, starting date and stopping date for the job spell. A person is considered register employed during a reference period if the job starts before the end of the reference period, and stops after the beginning of the reference period. This definition means that you have to work "some time" during the reference period to be considered employed. As the administrative employee register data are somewhat revised after Statistics Norway collects them, we use the term "register data" or "register" in the following.

Since we have register data with employment information, it is tempting to assume that the registered value is the correct value, and estimate the measurement error in LFS employment by the difference between survey value and register value. However, both Kleven et al. (2006) and others argue that an observed divergence between survey and register-based data cannot be considered as an error in the survey. Register properties such as different definitions and time lag, as well as random errors in the register, can also cause divergence between the two data sources.

### 3.3. *The Linked Data*

For each reference quarter, we link the survey sample data to the register data at the micro level, using the personal identification number found in both sources. It is important to note that we are linking people and not jobs. Our aim is not to compare the data quality in the two sources, since we know that both have measurement errors. What we want is to study the effects of including proxy interviews. For this purpose we use the register status as an auxiliary variable, since it is known for the whole population and highly correlated with the target variable.

## 4. Evaluating the Effects of Proxy Interviews

### 4.1. Defining Proxy Effects

We define:

$$Y_i = \begin{cases} 1 & \text{LFS - employed} \\ 0 & \text{LFS - not employed} \end{cases}$$

LFS employment status for unit  $i$ .

$$X_i = \begin{cases} 1 & \text{register - employed} \\ 0 & \text{register - not employed} \end{cases}$$

Register employment status for unit  $i$ .

$$Z_i = \begin{cases} 1 & \text{direct response} \\ 0 & \text{proxy response} \end{cases}$$

Interview mode for unit  $i$ .

$$n_{xz}$$

Response subsample  $s_{xz}$  defined by  $x$ ,  $z$ .

$$n = \sum_{x,z} n_{xz}$$

Response sample

$$p_{xz} = \frac{\sum_{i \in s_{xz}} Y_i}{n_{xz}}$$

LFS employment rate in response subsample defined by  $X$  and  $Z$ .

Following Thomsen et al. (2007) we consider two models for proxy response that are analogous to nonresponse models. The two proxy models result in two ways of calculating proxy effect:

1) PCAR (proxy completely at random): the proxy interviews are distributed completely by chance among all interviews. This means that the direct interviews are a random sample of the total responses. In this case, the proxy effect is simply the difference between the employment probabilities:

$$E = P(Y = 1|Z = 0) - P(Y = 1|Z = 1)$$

with an empirical estimate as the difference between the observed employment rates:

$$e = p_{\bullet 0} - p_{\bullet 1}$$

where

$$p_{\bullet 0} = \frac{\sum_{i=1}^{n_{\bullet 0}} y_i}{n_{\bullet 0}} \text{ and } p_{\bullet 1} = \frac{\sum_{i=1}^{n_{\bullet 1}} y_i}{n_{\bullet 1}}$$

In other words, the PCAR model assumes that all the subjects have equal probability of being interviewed by proxy. Under this simple (and as we shall see unrealistic) model, the

proxy effect is the difference in employment rate between proxy responses and direct responses. Clearly, if the direct response sample is biased then some of the observed difference in employment rate is due to this bias. In order to study this, we introduce an auxiliary variable, register employment.

2) PAR (proxy at random): the proxy interviews are randomly distributed given the auxiliary variable. That means for a given value of register employment, the direct interviews constitute a random sample of all interviews.

In this case, we define two proxy effects, one for  $X = 1$  (register-employed) and one for  $X = 0$  (not register employed)

$$E_0 = P(Y = 1|X = 0, Z = 0) - P(Y = 1|X = 0, Z = 1)$$

$$E_1 = P(Y = 1|X = 1, Z = 0) - P(Y = 1|X = 1, Z = 1)$$

with the empirical estimates:

$$e_0 = p_{00} - p_{01}$$

$$e_1 = p_{10} - p_{11}$$

Figure 1 shows the proxy effects estimated under PCAR and PAR models, quarterly from 1997 to 2008.

We observe that the proxy effect is negative under both models, indicating underreporting among proxy respondents. However, it is also clear that the proxy effect is smaller for the PAR model than for the PCAR model. This is because register employed people are overrepresented among those interviewed directly. It is interesting to ask whether introducing more auxiliary variables would further reduce the proxy effect.

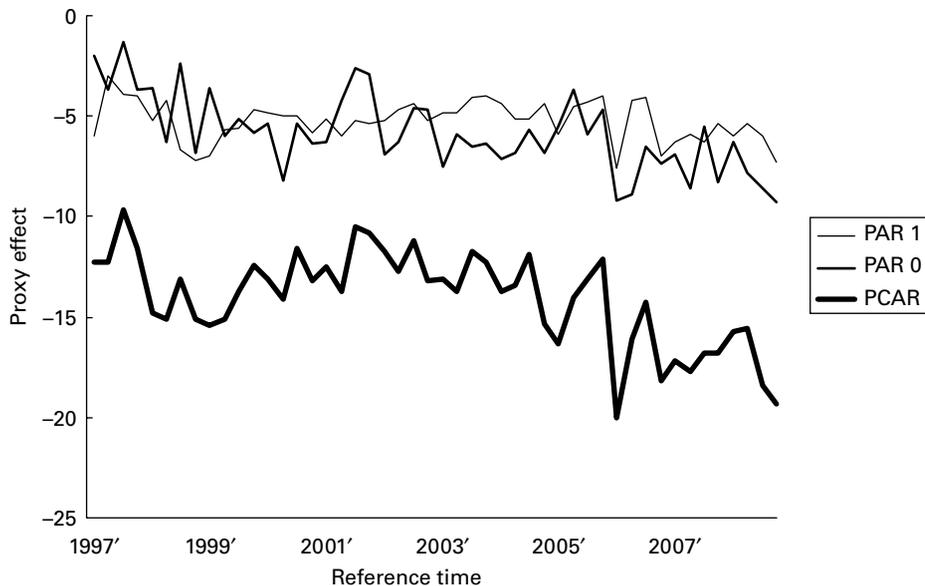


Fig. 1. Proxy effect under different proxy models. Quarterly LFS 1997 – 2008

Both the findings of Kleven et al. (2006) and those of Solheim et al. (2001) make it clear that including auxiliary variables such as education and age would have little effect on the results in Figure 1.

Both the proxy effects are approximately  $-6$  percentage points under the PAR model, on average for the whole period. Under the PCAR model, the proxy effect is about  $-14$  percentage points. Although there are fluctuations, both irregular and seasonal, the PAR model gives a consistently lower proxy effect than the PCAR model. Observing this relationship over a long time series helps to establish this as a significant result. In comparison, one standard error of the employment rate estimate is approximately 1 percentage point.

#### 4.2. Comparing Three Different Employment Rate Estimates Under the PAR Model

The figures for proxy effect shown in Figure 1 are differences in percentage points between employment rates. In order to evaluate the impact on the employment statistics, we look at three different estimates of employment rate. This will provide a clearer picture of whether proxy interviews should be included or not.

One estimate uses only the direct interviews ( $y_{direct}$ ). Another one uses the combined sample of both self-response and proxy interviews ( $y_{combined}$ ). These two estimates are compared to a “benchmark estimate” ( $y_{direct,PST}$ ):

$$\bar{y}_{direct} = p_{01} \frac{n_{01}}{n_{\bullet 1}} + p_{11} \frac{n_{11}}{n_{\bullet 1}}$$

$$\bar{y}_{combined} = p_{00} \frac{n_{00}}{n} + p_{01} \frac{n_{01}}{n} + p_{10} \frac{n_{10}}{n} + p_{11} \frac{n_{11}}{n}$$

$$\bar{y}_{direct,PST} = p_{01} \frac{n_{0\bullet}}{n} + p_{11} \frac{n_{1\bullet}}{n}$$

The benchmark estimate ( $y_{direct,PST}$ ) uses only the direct responses, but is adjusted by post-stratification using register-based employment to create two post-strata. We call this a “benchmark” because it is an unbiased estimate of the directly measured LFS employment rate under the PAR model, given the distribution of X among direct and proxy interviews. The difference between the benchmark and the direct responses constitutes what we term a *selection effect* or an indicator of “representativity” of the self-responses.

We also compare the combined sample estimate to the benchmark, and the difference can be interpreted as the proxy interviews’ overall influence on the employment estimate. Figure 2 shows the three employment rate estimates, quarterly from 1997 to 2008.

We observe that using only the direct interviews results in overestimation – due to the selection effect. Using all interviews results in underestimation – due to the proxy effect. But comparing the magnitude of the two effects, we observe that the combined estimate is closer to the unbiased benchmark. An average over the time series gives about  $-0.8$  percentage points difference for the combined sample (underestimation), whereas the direct sample gives about  $+1.3$  percentage points difference (overestimation). The preliminary conclusion is that the proxy effect is the smaller problem, and that it is better to include the proxy interviews.

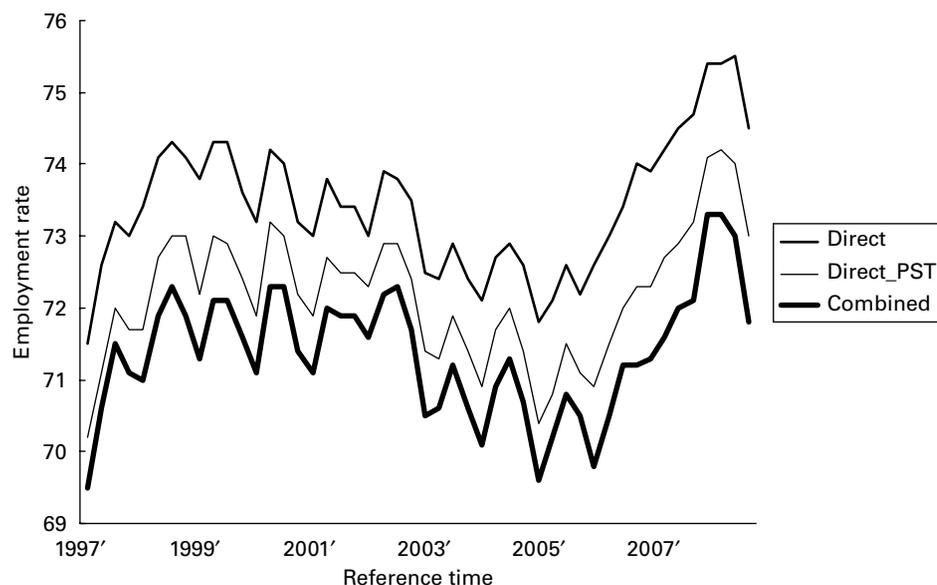


Fig. 2. Three different employment rate estimates. Quarterly LFS 1997 – 2008

The selection effect is mostly due to parents' answering on behalf of their offspring who are less likely to be reached at home. To explain the causes of the proxy effect, more analyses will have to be made.

The findings shown in Figure 2 depend on the method and model chosen. We have included age and education in order to get a more nuanced picture of the proxy effect, and it does vary between subpopulations. However, the proxy effect is negative in all subpopulations, and the overall effect on the LFS employment rate remains approximately the same as in Figure 2.

The PAR model assumes that the response propensity is independent of the LFS employment status for a given register employment status. This is a strong assumption, and in another study we found clear evidence that LFS employed people are over-represented in the sample, even for a given register employment status (Thomsen et al. 2007). This means that the benchmark estimate probably slightly overestimates the LFS employment rate. Consequently, the proxy effect shown in Figure 2 will be slightly overestimated, which further supports our conclusion that the proxy effect is smaller than the selection effect.

Finally, as we know that both nonresponse and proxy response are biased with respect to employment, we wish to investigate the connection between nonresponse and the proxy effect. In the next section, we study data that include the nonresponse units in survey sample data where both response and nonresponse units are linked to register-employment data.

## 5. Including the Effect of Nonresponse

The effect of nonresponse on the Norwegian LFS employment estimates has been studied in several papers, for instance Thomsen and Zhang (2001) and Thomsen et al. (2006).

We take it as an established fact that employed people are overrepresented in the combined response sample. And we have just shown that employed people are overrepresented in the direct response sample. In our study of the effects of proxy interviews on the employment rate, we wish to incorporate both of these characteristics. To do so, we use the fact that the auxiliary variable is available for the whole sample, including nonresponse.

We define:

$$N_x$$

Selected sample size in a group with the same register employment status

$$N = N_0 + N_1$$

Total selected sample size, sum of register employed and register nonemployed.

Following Thomsen et al. (2007), we consider two nonresponse models:

**MCAR:** Missing completely at random, meaning that the response probability is independent of both X and Y.

**MAR:** Missing at random, meaning that the response probability depends only on X.

Each of these nonresponse models can now be considered in combination with the PAR model for proxy response proposed in Section 4.

Under the PAR model for proxy response and MCAR model for nonresponse, the benchmark employment estimate ( $y_{direct,PST}$ ) is unbiased and the conclusions from Section 4.3 hold. It is clear from earlier studies and our own data that the MCAR model for nonresponse is not plausible. Consequently the benchmark estimate ( $y_{direct,PST}$ ) is not an unbiased estimator of the employment rate measured directly. We therefore employ a modified benchmark that is an unbiased estimate under the MAR model:

$$\bar{y}_{direct,PST}^* = p_{01} \frac{N_0}{N} + p_{11} \frac{N_1}{N}$$

In Figure 3 we compare the modified benchmark estimates for each quarter with the series for the direct and combined samples. The benchmark estimate adjusted with the nonresponse (dotted line) is slightly lower than the post-stratified estimate using only the response sample (thin line).

We observe that when we include the effect of nonresponse, the direct-only estimate has an even larger bias. It seems that the inclusion of the proxy interviews has about the same effect as post-stratification of the direct responses by using register employment as a post-stratification variable.

## 6. Conclusion

From the methods and data at hand, we conclude that the employment rate estimates probably are better when the combined sample is used. Extending the response sample size by including proxy interviews introduces some underreporting, but gives a more representative response sample.

The method outlined here can be applied to other variables where relevant auxiliary data are available. By this we mean independent data substantially correlated with the variable of interest. For instance, in the Nordic countries unemployment data at the

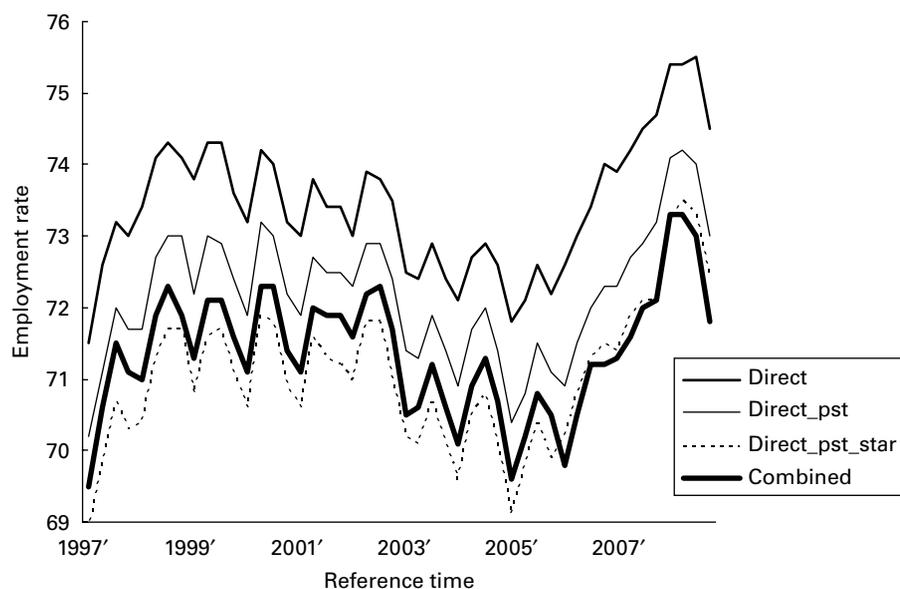


Fig. 3. Comparing employment rate estimates including nonresponse. Quarterly LFS 1997 – 2008

individual level is available both from survey and register sources. In a simple approach, the same method outlined in Sections 4 and 5 can be applied to unemployment. Two limitations should be noted about this approach. Firstly, some strong assumptions have been made in order to estimate the proxy effect. Secondly, we have limited the study to only binominal variables. For an extended approach, more register-based categories can be constructed, for instance: register employment, register unemployment, and “not in workforce.” Proxy effect for employment and unemployment can be measured separately in each of the three register-based categories. The consequence is several more figures to compare at the same time, but we suggest that this exploration-oriented approach has merits for this subject at this stage.

In a planned study, using data on working hours, we intend to follow up on the limitations of the present study.

## 7. References

- Boehm, L.E. (1989). Reliability of Proxy Response in the Current Population Survey. In Proceedings of the American Statistical Association, Survey Research Methods Section, 486–489.
- Dawe, F. and Knight, I. (2007). A Study of Proxy Response in the Labour Force Survey. Labour Force Survey User Guide Vol. 1, Ch. 11, The Office for National Statistics.
- Groves, R.M., Fowler, Jr., F.J., Couper, M.P., Lepkowski, J.M., Singer, E., and Tourangeau, R. (2004). Survey Methodology. Hoboken NJ: Wiley.
- Kleven, Ø., Lagerstrøm, B.O., and Thomsen, I. (2006). A Simple Model for Studying the Effects of Proxy Interviewing. Paper presented at the 17th Annual International

- Workshop on Household Survey Nonresponse, Omaha, NE, U.S.A. Statistics Norway Documents.
- Lemaitre, G. (1988). A Look at Response Errors in the Labour Force Survey. *The Canadian Journal of Statistics*, 16, 127–141.
- Kalsbeek, W.D. and Agans, R.P. (2007). Sampling and Weighting in Household Telephone Surveys. *Advances in Telephone Survey Methodology*, Ch. 2 in Lepkowski, J.M. et al. Hoboken, NJ: John Wiley & Sons.
- Martin, J. and Butcher, B. (1982). The Quality of Proxy Information – Some Results from A Large-Scale Study, *The Statistician*, 31, 3. London: Wiley-Blackwell.
- Moore, J.C. (1988). Self/proxy Response Status and Survey Response Quality. *Journal of Official Statistics*, 4, 155–172.
- O’Muircheartaigh, C. (1991). Simple Response Variance: Estimation and Determinants. *Measurement Errors in Surveys*, Ch. 27 in Biemer et al. New York, NY: Wiley.
- Solheim, L., Håland, I., and Lagerstrøm, B.O. (2001). Proxy Interview and Measurement Error in the Norwegian Labour Force Survey. Paper presented at the 12th International Workshop of Household Survey Nonresponse, Oslo.
- Steeh, C. (2008). Telephone Surveys. In *International Handbook of Survey Methodology*, E.D. de Leeuw, J.J. Hox, and D.A. Dillman (eds), Ch. 12. New York, NY: Lawrence Erlbaum.
- Thomsen, I., Kleven, Ø., and Zhang, L.-C. (2007). Dealing with Nonsampling Errors using Administrative Data. Paper presented at the IAOS Conference, Lisboa.
- Thomsen, I., Kleven, Ø., Wang, J.H., and Zhang, L.-C. (2006). Coping with Decreasing Response Rates in Statistics Norway. *Statistics Norway Reports*.
- Thomsen, I. and Zhang, L.-C. (2001). The Effects of Using Administrative Registers in Economic Short Term Statistics: The Norwegian Labour Force Survey as a Case Study. *Journal of Official Statistics*, 17, 285–294.

Received April 2009

Revised August 2010