



*Øyvin Kleven, Bengt Oscar Lagerstrøm
and Ib Thomsen*

**A Simple Model for Studying the
Effects of Proxy Interviewing:
The Norwegian Labour Force
Survey as a Case Study**

Preface

This report is a paper by Øyvind Kleven¹, Bengt Oscar Lagerstrøm² and Ib Thomsen³ presented at the 17th annual International Workshop on Household Survey Nonresponse, Omaha, Nebraska, US, August 20-30 2006.

The Norwegian Labour Force Survey (LFS) is a longitudinal survey of households and individuals. Over the years, a great deal of effort has been devoted to fight non-response and to keep the cost low. Indirect or proxy interview is a common way to fight both nonresponse and cost. About 10-15 per cent of the respondents in the Norwegian LFS are proxies. It is therefore important to have an idea of the impact of proxy interviewing on response errors and response probabilities. Using information from the employment register on each respondent we study these two effects upon employment. Previous studies show that proxy interviews introduce an underreporting of employment. We study the question: Does the joint impact of reducing nonresponse, and the use of proxy interviews lead to higher or lower quality?

Key words: proxy interviewing; nonresponse.

¹ Division for Data Collection Methods, Statistics Norway, P.B. 8131 Dep., N-0033 Oslo Norway. E-mail: kle@ssb.no.

² Division for Sample Surveys, Statistics Norway, P.B. 8131 Dep., N-0033 Oslo Norway. E-mail: bol@ssb.no.

³ Division for Statistical Methods and Standards, Statistics Norway, P.B. 8131 Dep., N-0033 Oslo Norway. E-mail: ith@ssb.no.

Contents

- 1. Introduction 3
- 2. Data..... 3
- 3. Why do survey organizations use proxy? 3
- 4. Proposing a simple model to study the effect of proxy 4
- 5. Conclusions 6
 - 5.1 Further work..... 7
- 6. References 8

Tables

- Table 1. Outcome by introducing proxy on nonresponse error and incorrect answers. 4
- Table 2. LFS sample by employment status in the employment register and the LFS. 5
- Table 3. The direct observed sample by employment given by employment status in the employment register and the LFS. 5
- Table 4. The proxy sample by employment given by employment status in the employment register and the LFS. 5
- Table 5. Employment rates by register employment status and proxy/direct interviewing. 6

1. Introduction

Proxy interviewing is sometimes used to reduce costs and nonresponse bias for survey data. By proxy we mean that another person gives an answer on behalf of the person selected or the respondent. In Statistics Norway proxy has a long history in the Labour Force Survey, and has for many years constituted about ten to fifteen percent of the total net sample. The LFS has a relatively short period of fieldwork; the sample units must be interviewed within ten days. Since it is not possible to reach all the respondents in ten days, we allow the interviewers to make proxy interviews. Since the Norwegian LFS is a sample of families, it is easy and cost efficient to accomplish this rule. The effect of proxy interviewing on the accuracy of the employment estimate have to some degree been studied earlier. Thomsen and Zhang (2001) investigated the effects of using administrative registers in economic short-term statistics. Solheim et al. (2001) studied the impact among young adults and Foss (2004) touch the topic in a comparability of quality issues in The LFS and the Employment register. Self-reported data are usually assumed to be more accurate, complete, and reliable. Some research suggests that the distribution of response from proxies often differs from those of self-respondents without allowing us to conclude which is better (Moore, 1988). Since there seldom are external sources available that can be used to validate the self- and proxy reports, it is normally hard to conclude. It is, however, evident that what we gain in reducing the nonresponse errors by allowing proxy interview can be lost in terms of the “total survey error” by the fact that proxy gives more incorrect answers. In this paper we suggest a simple model to distinguish between the two effects.

2. Data

In this paper we exploit the connection between the employment/unemployment status from the Norwegian Labour Force Survey (LFS) and the employment/unemployment status in the Norwegian Employment Register (NER). LFS has a sample of about 24,000 respondents (for more documentation on LFS see: http://www.ssb.no/english/subjects/06/01/aku_en/). The selection unit is the formal family. The respondent's spouse/cohabitant or parents are allowed to give the information by proxy report. NER is a combination of several administrative registers. The core is the Central Population Register (CPR). The main sources for labour force variables are the Register of Employees, the Register of Wage Sums, the Tax Return Register, the Register unemployment and the Business Register. Statistics Norway has developed a system of joint utilization of these registers (for more documentation see: http://www.ssb.no/english/subjects/06/01/regsys_en/). Hence we are able to match each respondent's status in the two sources on micro level.

In this study we use data from the 1st quarter of 2005. There were a total of 2,801 proxy reports to compare with the 18,764 self-reports. The data is further described in appendix table 1. The nonresponse level was 11.5%.

3. Why do survey organizations use proxy?

There are two major advantages from the survey organization point of view. The use of proxy gives information on respondents that otherwise would have been nonrespondents. Allowing proxy interviewing can give a lower variance in the estimates and more importantly less bias since the persons interviewed by proxy is believed to be systematically different from those interviewed directly. The other advantage is due to the cost efficient argument. Asking someone in the family about the respondent is more efficient than tracing respondents that are hard to get in contact with. If proxy were banned, Statistics Norway would probably have to spend considerably more money to reach about the same response rate that we do to day.

Whether to use proxy or not can be expressed as a trade off between at least two sources of error. One motivation for using proxy is to reduce nonresponse error. A disadvantage in using proxy can be that the answers given on behalf of other persons can be more incorrect than given from the respondent. Table 1 shows an analytical approach to the problem.

Table 1. Outcome by introducing proxy on nonresponse error and incorrect answers.

		Nonresponse error		
		Increasing	About the same	Decreasing
Incorrect answers	Increasing	1	2	3
	About the same	4	5	6
	Decreasing	7	8	9

4. Proposing a simple model to study the effect of proxy

Both our sources have different types of “measurement errors” or “errors of observation” (for a review of different types of measurement errors see Groves et al. 2004). The Employment register suffers from time lag and misclassification. The LFS suffer from incorrect answers due to misinterpretation, tendency to report more social desirable answers etc. There are also slightly different ways to classify employment status in the two sources (see Foss (2004) for details).

We believe it is inappropriate to treat the Employment register as the gold standard or the true value hence defined the “measurement error” as the deviance between the survey response and the register. In direct interviews, the LFS is considered to be the truth, and therefore we propose that the relationship between the two sources obtained from the direct interviews should be used as the gold standard to measure the effect of proxy interviewing. We will from now on use the term proxy effect rather than “measurement error”.

The difference in employment status between the direct interview and the proxy sample is rather big. The estimate from the direct interview shows an employment rate of 71.7, the corresponding estimate from the proxy sample is 55.2 (see Appendix table 1. for more details). Assuming that the proxy sample is a random sample from the whole sample this difference may be called the proxy effect. The assumption may be evaluated by means of post stratification when appropriate data are available. In what follows we will present an example in which we split the observed difference between the mean of the proxy sample and the mean of the direct interviewed sample into two components:

- A proxy effect which is due to the fact that proxy is used, and
- A selection effect which is due to the fact that the proxy sample is not a random sample

In table 2, 3 and 4 the LFS sample is cross-classified with employment status given by the employment register. A unique personal identification number link the two sources at micro level.

Table 2. LFS sample by employment status in the employment register and the LFS.

		<i>Register</i>		
		<i>Employed</i>	<i>Not employed</i>	
<i>LFS</i>	<i>Employed</i>	13,205	1,802	15,007
	<i>Not employed</i>	723	5,835	6,558
		13,928	7,637	21,565

Table 3. The direct observed sample by employment given by employment status in the employment register and the LFS.

		<i>Register</i>		
		<i>Employed</i>	<i>Not employed</i>	
<i>LFS</i>	<i>Employed</i>	11,918	1,544	13,462
	<i>Not employed</i>	572	4,730	5,302
		12,490	6,274	18,764

Table 4. The proxy sample by employment given by employment status in the employment register and the LFS.

		<i>Register</i>		
		<i>Employed</i>	<i>Not employed</i>	
<i>LFS</i>	<i>Employed</i>	1,287	258	1,545
	<i>Not employed</i>	151	1,105	1,256
		1,438	1,363	2,801

We calculate the employment rate for direct interview to $13,462/18,764 = 0.717$, Table 2, and the corresponding employment rate for proxy interview to $1,545/2,801 = 0.552$ from table 3.

If the proxy sample is random sample from the whole sample this difference may be noted the proxy effect. Under this assumption the proxy effect is -0.165

From table 3 it is clear that this assumption is not feasible. The employed according to the register is substantially underrepresented in the proxy sample.

A different approach is to divide the sample into two, employed in the register and not employed in the register, and calculates employment rates for proxy/direct interview. This approach gives the rates for proxy/direct interview as showed in table 5.

Table 5. Employment rates by register employment status and proxy/direct interviewing.

	<i>Register</i>	
	Employed	Not employed
Direct int.	11,918/12,490=0.954	1,544/6,274=0.246
Proxy	1,287/1,438=0.895	2,58/1,363=0.189

Proxy effect.

Under the assumption that the proxy sample is random given the register status, we can estimate the proxy effect within each register status as

Employed in register: Proxy effect = -0.059

Not employed in register: Proxy effect = -0.057.

From this we find that both proxy effects are around -6% and clearly different from the one estimated above.

Selection effect.

Under the same assumption as for the proxy effect, we can estimate the employment rate if direct interviewing was used on the whole sample as

$$(13,928/21,565) * 0.954 + (7,637/21,565) * 0.246 = 0.703 \tag{1}$$

If only the direct sample was used, the employment rate would be

$$13,462/18,764 = 0.717 \tag{2}$$

The difference between these two estimates, we denote the selection effect.

5. Conclusions

If only the direct sample was used, we seem to overestimate the population employment rate because the persons with register status employed are clearly overrepresented in the direct sample.

If a combination of the direct sample and the proxy interviews is used, the employment rate is

$$15,007/21,565 = 0.696 \tag{3}$$

This underestimates the population employment rate because of the of the proxy effect. Comparing the three estimates (1), (2), and (3), it seems as if the inclusion of proxy interviews underestimates the employment rate with approximately 0.007, while using only the direct sample overestimates the same rate with about 0.014.

If poststratification is used in the estimation, it seems as if proxy interviewing should be avoided.

5.1 Further work

The approach used in the example must be parameterized and generalized allowing several auxiliary variables. A crucial assumption is to assume that the proxy sample is a random sample within each post-stratum. This assumption is hardly correct, and to evaluate its importance we shall use an approach similar to the one presented in Zhang (1999). Cost considerations should also be elaborated.

6. References

Groves, Robert M., Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau (2004) *Survey Methodology*, New York: Wiley

Foss, Aslaug Hurlen (2004) *Quality measurement of the Norwegian Labour Force Register*. Paper presented at the European Conference on Quality and Methodology in Official Statistics (Q2004). Mainz, Germany, 24-26 May 2004.

Moore, J.C (1988) Self/proxy Response Status and Survey Response Quality. *Journal of Official Statistics*, 4, 155-172

Solheim, Leiv, Inger Håland and Bengt Oscar Lagerstrøm (2001) *Proxy interview and measurement error in the The Norwegian Labour Force Survey*. Paper presented at 12th International Workshop on Household Survey Nonresponse, Oslo Norway 12-14 September 2001.

Thomsen, I and Zhang, L.-C. (2001). The Effect of Using Administrative Registers in Economic Short Term Statistics. *Journal of Official Statistics*. Vol 17, no 2, pp 285-294.

Zhang, L.-C. (1999). A Note on Poststratification when Analysing Binary Survey Data subject to Nonresponse. *Journal of Official Statistics*, Vol 15, no 2, pp 329-334

Appendix table 1. Employment status among direct interviewed and proxy interviewed. Labour Force Survey 1st quarter 2005

	Direct interview				proxy			
		Employed	Not employed	N		Employed	Not employed	N
All	100.0	71.7	28.3	18,764	100.0	55.2	44.8	2,801
Gender								
Male	100.0	74.8	25.2	9,321	100.0	57.9	42.1	1,563
Female	100.0	68.7	31.3	9,443	100.0	51.7	48.3	1,238
Age group								
16-19	100.0	34.5	65.5	1,129	100.0	30.1	69.9	518
20-24	100.0	66.3	33.7	1,482	100.0	54.9	45.1	370
25-39	100.0	83.9	16.1	5,320	100.0	70.7	29.3	546
40-54	100.0	87.4	12.6	5,708	100.0	80.0	20.0	641
55-66	100.0	66.7	33.3	3,702	100.0	56.5	43.5	481
67-74	100.0	11.8	88.2	1,423	100.0	6.1	93.9	245
Education								
Low	100.0	44.7	55.3	3,001	100.0	37.3	62.7	740
Middle	100.0	74.1	25.9	10,283	100.0	61.4	38.6	1,332
High	100.0	86.0	14.0	5,170	100.0	78.5	21.5	544
Male/ Age group								
16-19	100.0	32.8	67.2	574	100.0	26.4	73.6	288
20-24	100.0	68.7	31.3	723	100.0	61.4	38.6	210
25-39	100.0	87.2	12.8	2,686	100.0	76.9	23.1	277
40-54	100.0	90.4	9.6	2,819	100.0	84.0	16.0	362
55-66	100.0	71.4	28.6	1,819	100.0	62.8	37.2	277
67-74	100.0	14.4	85.6	700	100.0	6.0	94.0	149
Female/ Age group								
16-19	100.0	36.4	63.6	555	100.0	34.8	65.2	230
20-24	100.0	63.9	36.1	759	100.0	46.3	53.8	160
25-39	100.0	80.5	19.5	2,634	100.0	64.3	35.7	269
40-54	100.0	84.5	15.5	2,889	100.0	74.9	25.1	279
55-66	100.0	62.3	37.7	1,883	100.0	48.0	52.0	204
67-74	100.0	9.3	90.7	723	100.0	6.3	93.8	96