The use of R-indicators in the Danish LFS

Introduction

Like many other surveys, the Danish Labour Force Survey (LFS) has experienced a decline in its response rate. A lower response rate increases the risk of non-response bias in the case where the non-response is not random. Hence, when it comes to accuracy, the goal in itself should not be increasing the response rate, but rather to have a missing-at-random pattern to decrease the bias and give better accuracy of the estimates. However, a high response rate will increase the precision of the estimates and increase the level of detail to which the results can be analyzed.

In an attempt to increase the LFS response rate, we have introduced a number of measures. The two measures that will be analyzed in this paper are 1) a lottery for cash prizes and 2) an extended data collection period. Both measures have had a positive effect on the response rate.

After seeing a positive effect on the response rate, we started wondering if the extra respondents made the data more representative or if we were introducing more bias. Hence we wanted to find a way to measure representativeness. For this, R-indicators have previously been used for the LFS, which use the response propensities to present how far the composition of the response data deviates from the original sample. We will now shortly introduce the concept of R-indicators and apply the measure on the Danish LFS.

Introduction to representativeness and R-indicators

We define representativeness based on individual response propensities, which denote the likelihood or tendency of individuals to participate in a survey or questionnaire.

> *Strong definition*; A response subset is representative with respect to the sample if the response propensities are the same for all units in the population and if the response of a unit is independent of the response of all other units. (Schouten, et al., 2008)

In practice, it is not possible to test the strong definition, and therefore we need a weaker definition:

> *Weak definition*; A response subset is representative of a categorical variable if the average response propensity over the categories is constant. (Schouten, et al., 2008)

We can estimate the response propensity for each category, allowing the weak definition to be tested in practice. We use logit models to estimate the response propensities.

Denoting the response propensities by ρ_i, the standard deviation for the response propensities for the population is given by

$$S(\rho) = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(\rho_k - \bar{\rho})^2}$$

which is a distance function that takes the value 0 if all response propensities are equal, and increases when the variation between the response propensities increases. Since the maximum value of $S(\rho)$ is 0.5, Schouten, et al. proposes the following R-indicator which take values on the interval [0,1].

$$R(\rho) = 1 - 2S(\rho)$$

Here $R(\rho) = 1$ corresponds to the ideal case where all sampled individuals have the same response propensity.

Since the R-indicator measures the representativeness with respect to the original sample, it is important for the analysis, that the original sample is representative. If design weights are used, one should be sure to include them in the calculations of the R-indicator.

Stratified sampling

If the sample is not created using a random sampling method, individuals in the population will not have equal probabilities of being selected for inclusion in the sample. This affects the estimated response propensity for the groups since we likely sample a larger proportion of individuals known to have a lower response propensity. To address this issue effectively, we integrate design weights into the computation of the R-indicator. The mean response propensity is estimated by $\hat{\bar{\rho}} = \frac{1}{N}\sum_{k=1}^{n} \hat{\rho}\,\omega_k$ where ω_k denotes the design weight for individual k, and the sum of the weight is the population size N. The standard deviation for the response propensities in the population is subsequently calculated as

$$\hat{S}(\hat{\rho}) = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}\left(\hat{\rho}_k - \hat{\bar{\rho}}\right)^2 \omega_k}$$

Resulting in the R-indicator

$$\hat{R}(\hat{\rho}) = 1 - 2\sqrt{\frac{1}{N-1}\sum_{k=1}^{N}\left(\hat{\rho}_k - \hat{\bar{\rho}}\right)^2 \omega_k}$$

Estimating the response propensity

Since the individual response propensities are unknown, we estimate the response propensities using a logit model. In this analysis, we have used the same variables and interaction terms as we use when calibrating the weights.

Interaction terms used to estimate response propensities

*gender * age*	2 genders, 12 age groups; 5 year intervals
*region * age*	5 regions, two age groups; <75 and >75
labour market status	register based, 10 levels
*education level * age*	3 educational levels, 7 age groups; 5 year intervals, 45-74, 75+

Various variables and interactions will give different results, since it will measure variations of the weak definition of representativeness.

Maximal absolute bias and maximal root mean square error

We want to be able to attach bounds to an R-indicator, such that changes in the R-indicator can be interpreted. Schouten et al. shows, that for any survey item Y, the R-indicator can be used to set upper bounds to the non-response bias. The bounds are given by

$$B_m(\rho, y) = \frac{\left(1 - R(\rho)\right)S(y)}{2\hat{\rho}}$$

These bounds are different for each survey item y, thus we define a hypothetical survey item for the purpose of comparison. We let this survey item have $\hat{S}(y) = 0.5$. The bound of the bias are then estimated by

$$\hat{B}_m(\hat{\rho}) = \frac{(1 - \hat{R}(\hat{\rho}))}{4\hat{\bar{\rho}}}$$

Framework of the Danish LFS

The Danish LFS operates on a mixed-mode data collection framework. Initially, the entire sample is invited to participate online (CAWI), followed by phone interviews (CATI) for non-respondents. Respondents are surveyed four times: during the first two quarters consecutively, followed by a break of two quarters, and then interviewed again for two consecutive quarters. Consequently, half of the sample is refreshed each quarter. The sample size for each quarter is approximately 35,000.

Lottery for cash prizes

In the 4th quarter of 2022, an experiment was initiated in the Labor Force Survey (LFS), where a lottery for cash prizes was introduced among the completed responses. All responses were included in the lottery, but 20 percent of the respondents were not informed about the lottery, thus acting as a control group in assessing the effect. In total, a lottery of 20,000 DKK (2,700 EUR) in prizes was drawn, distributed among 11 individual prizes. With just over 14,000 interviews in the 4th quarter, this corresponds to an expenditure of 0.70 DKK (0.01 EUR) per interview.

Effect on response rates

The response rate for the prize group was overall 2.1 percentage points higher than for the control group. The effect was most pronounced among individuals under 30 years old, who generally have the lowest response rate. The prize had a lesser effect among individuals aged 30-49 compared to the age group 50-64. This may be related to the concern where a cash prize could evoke thoughts of identity theft attempts and other dishonest solicitations. However, this does not overshadow the positive effect - there are still 1.3 percentage points more responses in the age group. There is also a significant positive effect among 50-64 year olds, where the response rate is 2.4 percentage points higher. Among individuals aged 65 and above, the effect is minimal, but this is expected and insignificant, as this group has by far the highest response rate and the potential for increasing the response rate is subsequently lower.

Effect on refusal rates

Incomplete interviews are divided into respondents who simply did not respond and those who actively refused to participate. Everyone receives an invitation letter, but not everyone has a registered telephone number – therefore some would-be refusals are also included in the numbers for the non-contact.

The proportion of individuals who refused to participate is 0.5 percentage points higher in the prize group (7.9 percent) than in the control group (7.4 percent). The proportion of refusals remains roughly unchanged for the youngest and oldest age groups, while it increases in the middle age groups. The effect on the refusal rate is greatest in the age group 30-49 years old (+1.1 percent) and second greatest in the age group 50-65 years old. The reason for this is unknown, but possibly these age groups are most vigilant against possible fraud and are therefore more skeptical about the payout of cash prizes.

Effect on representativeness

We estimated the response propensities using the logit model with the same variables and interaction terms as we use when calibrating the weights. We did this for the prize group, the control group and the entire group. The R-indicator is slightly lower for the prize group, indicating that the representativeness is slightly better in the control group. The R-indicators for the two groups are quite close with respect to the chosen variables. When time allows, we will estimate confidence intervals for the R-indicator.

	Sample size	*Response rate*	*R-indicator*	*Max bias*
Prize group	24,358	0.441	0,764	0,049
Control group	6,099	0.419	0,776	0,024
Total	30,457	0.437	0.766	0,055

Overall effect

The introduction of prizes leads to a 0.5 percentage point increase in explicit refusals to respond, highlighting the importance of ensuring the credibility of our communication channels. However, this negative effect is more than offset, as there is an overall positive effect on the response rate at 2.1 percentage points, thus significantly increasing the response rate.

The R-indicator is slightly higher for the control group, which indicates that the respondents in the prize group is slightly less representative with respect to the variables used in the estimation of the response propensities than the control group. We still need to calculate confidence intervals to determine if the difference is significant.

Extended data collection period

In the Danish LFS everyone receives an invitation letter with a link to answer the survey online. After a couple of days, and if the survey has not been answered through the link, we will send another letter and then start trying to get in touch with those with a registered telephone number. The share of the sample, who has a registered telephone number, and their willingness to pick up the phone when we call, has decreased over the years. Furthermore, a share of respondents complained, that we did not give them enough time to answer online before we started calling them. Therefore, we decided to extend the data collection period from two to four weeks, giving the respondents more time to answer, and having a more spread-out and therefore less intense reminder process.

The extension of the data collection period was implemented for the entire sample in the fourth quarter of 2023, and there was no control group for this measure for budgetary reasons. However, the response rate in the implementation quarter was 4.3 percentage points higher than in the previous quarter. By extending the data collection period, we also saw an increase in the share of respondents who fill out the survey online (CAWI) from approximately half to now almost two-thirds. Most of the interviews are still collected during the first two weeks after the reference week.

We estimate the response propensities using the logit model with the same variables and interaction terms as we use when calibrating the weights. The resulting R-indicators are shown in the table below. We see no clear correlation between the response rate and the R-indicator.

R-indicator from 2021-2023

Year and quarter	*Response rate*	*R-indicator*	*Max bias*
2021, 1	0.536	0.829	0.040

2021, 2	0.497	0.822	0.044
2021, 3	0.454	0.818	0.047
2021, 4	0.442	0.749	0.061
2022, 1	0.462	0.778	0.054
2022, 2	0.411	0.782	0.057
2022, 3	0.395	0.785	0.058
2022, 4	0.402	0.783	0.058
2023, 1	0.413	0.793	0.055
2023, 2	0.404	0.796	0.054
2023, 3	0.402	0.803	0.053
2023, 4	0.444	0.805	0.049

It should be noted, that we in this period also have improved our letters, making them more relatable for the respondents. The letters have been continuously adjusted as we operate within an adaptive survey design framework. Furthermore, we gradually implemented a policy of not re-interviewing seniors starting in the first quarter of 2023.

An alternative approach to assessing the optimal length of the data collection period is to compute the R-indicator and maximum bias daily throughout the collection period. It is anticipated that, eventually, despite potential slight increases in the response rate, both the R-indicator and maximum bias will reach their minimum levels.

Concluding remarks

By introducing a lottery for cash prizes and extending the data collection period from two to four weeks, we have observed a positive effect on the response rate. However, we still need to calculate confidence intervals for the R-indicator to determine if the changes in the R-indicator, and consequently the representativeness, are significant.

Lastly, a few final thoughts on how we might proceed from here.

Test of other variables in the logit model

In this analysis, we chose to estimate the response propensities based on the same variables that we use to calibrate the weights after data collection. Perhaps it would be more interesting to consider the representativeness on other variables, which we do *not* correct for later. This could show us whether we have representativeness problems not accounted for. Other variables of interest could be income, citizenship or social status.

Analysis on mode effects

Upon extending the data collection period, we observed an uptick in the proportion of respondents completing the survey online. Exploring the effects of this change on survey results would be insightful. During such analysis, it would be crucial to account for both selection bias and measurement bias.

Referencer

Schouten, B., Cobben, F. & Bethlehem, J., 2008. Indicators for the Representativeness of Survey Response. *Survey Methodology 35(1)*, November.