

Searching for auxiliary variables

Linda Mohay, Hungarian Central Statistical Office

International Workshop on Household Survey Nonresponse 2015 Leuven, Belgium

Introduction

The Hungarian Labour Force Survey (LFS) is one of the most important household surveys in Hungary and therefore LFS response rate is always a subject of concern. As in Hungary there is no administrative data available to use as auxiliary variables for nonresponse analysis censuses (in every tenth year) offer a single great possibility to learn something about nonrespondents.

Census based link study was performed in the UK as well (ONS, 2004, 2012, 2013) and they already have some results and applied techniques providing good basis for our analysis. In the course of this research we planned to link census dataset with three household surveys (LFS, Household budget survey and Travel habits survey), but so far only LFS was linked. LFS has the largest sample out of the three surveys (about 36 thousand households in a quarter of a year). LFS is a rotating panel and one household spend six years in the sample. There are household-level dispositional codes and only a negligible amount of person-level refusals occurs.

The LFS data was linked with census data, where the survey fieldwork was carried out close to the Census day (1st October). From the sample of LFS (Wave 1 cases) a window of three months either side (i.e. July to December inclusive) was chosen for analysis (11439 cases). The response and the contact rates are in table 1:

Table 1:

Completed	Not contacted	Refusal	Other nonresponse
69,6%	19,7%	10,3%	0,4%

We investigated the following questions in our linked-database:

- characteristics of refusal and non-contacted households
- characteristics of households which took part in LFS after they refused the cooperation(or non-contacted) in the first wave

You could see the examined variables in Table 2.

Table 2:

Variables from LFS (about all sample units – interviewer observations):	Variables from census:
nature of neighborhood	type of the flat
type of the building	degree of conveniences
state of the building	type of the household
population density	age structure of household
type of settlement	proportion of elderly household members

proportion of employed household members
proportion of unemployed household members
proportion of dependant household members
number of household members
gender of head of household
labor market situation of head of household
distance from the center

Logistic Regression analysis was used to identify the characteristics most strongly associated with non-contacts and refusals.

LFS households that were difficult to contact tended to be:

- households living in a good shaped building
- households living in densely populated area
- households living in central region
- single adult households
- households with greater proportion of employed adults
- where households reference person is younger

Households who refused to take part in the LFS were most likely to be:

- living in greater cities
- living in urban area
- living in central region
- and the households reference person is younger

We found the above associations but we have to notice that in general, the models do not fit well.

The values for Nagelkerke's pseudo R squared are low (0,085 in non-contact model and 0,073 in refusal model).

It was interesting for us that the most strongly associated variables were not from the census database but from LFS paradata and other interviewer observations. We are interested in if we should collect other background variables during fieldwork for a grounded regular nonresponse analysis. On the other hand we would like to use the census dataset for further analysis and we interested in other variables we can examine related to nonresponse. (In the 2011th census database we have geocodes assigned to the households. Are there any suggestion to use them in nonresponse analysis?)

We were interested in the households we could not recruit in the first wave but in the second one they took part in LFS. We examined LFS data from January of 2012 until the end of 2014. You can see the distribution of them in Table 3.

Table 3:

households non-contacted at least in one wave	19059
cooperative household after non-contacted in any wave	8132 42,67%
cooperative household after non-contacted in first wave	4838 25,38%

households refused to cooperate at least in one wave	15621
cooperative household after refusal in any wave	1634 10,46%
cooperative household after refusal in first wave	1122 7,18%

We have a policy in LFS that if a household gets a non-contacted disposition code in one wave it remains in the sample until the sixth wave but if a household refuses to cooperate it should be attempted to recruit it once more and if it fails then this household falls out of the sample.

In the current link-study we examined the later cooperate households as well but the number of cases is low. (See Table 4.)

Table 4:

households non-contacted in first wave	2239
cooperative household after non-contacted in first wave	636 28,4%
households refused to cooperate in first wave	1159
cooperative household after refusal in first wave	109 9,4%

Logistic Regression analysis was used to identify the characteristics most strongly associated with later cooperate households, but we couldn't find any characteristic so far. It seems that these households are more similar to the cooperative households than the non contacted or refusal households.

We interested in the question how we should treat sample elements with different dispositional codes: Should we try to contact or persuade them in every wave? How can we decide the appropriate treatment - considering the cost of an attempt and on the other hand the potential bias. How can we measure the effect of our applied method? Are there any differences in best practices between household panel surveys regarding of the treatment of nonresponse through waves? How much is this similar to following up on nonrespondents in one time survey.

References:

ONS (2004) "The Scottish Household Survey Report of the 2001 Census-linked study of survey non-response" available at <http://www.gov.scot/Resource/Doc/933/0009409.doc>

ONS (2012) "A comparison of the 2011 Census and the Labour Force Survey (LFS) labour market indicators", available at http://www.ons.gov.uk/ons/dcp171776_290711.pdf

ONS (2013) "Non-response Weights for the UK Labour Force Survey? Results from the Census Non-response Link Study", available at <http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/articles-and-reports/non-response-weights-for-the-uk-labour-force-survey.pdf>