Explicit Non-Response Weighting on a Cross-National Survey

Peter Lynn, University of Essex, UK

Short paper for Nonresponse Workshop 2020

ABSTRACT

The European Social Survey (ESS)[1] is a biennial survey that aims to provide survey data for comparative analysis. Between-country comparisons constitute a major use of the data. Non-response adjustment on the ESS currently consists of a population calibration adjustment to the design weights. However, a majority of the participating countries are now able to provide informative individual-level auxiliary variables from the sampling frame, typically a population register. The question therefore arises as to whether the ESS weights could be improved by adding a step of explicit non-response adjustment prior to the calibration, based upon the frame variables. However, issues of between-country consistency and comparability arise. Not all countries can supply frame variables. And amongst those that can, the variables differ in both substance and detail. This presentation poses the question of how we should judge, in this context, whether adding an adjustment step is likely to prove beneficial. The nature of the available frame variables will be described and some illustrative analysis of their relationship with non-response, both unconditionally and conditional on the existing weights, will be presented.

CURRENT ESS WEIGHTING PRACTICE

Currently, ESS produces design weights for each country and then uses a single calibration step to deal with statistical error due to random sampling variation and non-response. [ESS documentation refers to this as post-stratification, though it is not strictly post-stratification.] The default approach is to adjust design-weighted distributions in each country to match European Labour Force Survey (ELFS) weighted distributions in terms of:

- Gender x age (3 groups) x education (3 groups); and
- NUTS2 regions

In practice, the approach often differs from the default for various reasons: at Round 8 the default procedures were strictly followed in only 9 of the 23 countries. For examples, some ESS countries do not participate in the ELFS so an alternative source of weighting targets must be found. Also, the number of NUTS2 regions in a country ranges from 1 (CY, EE, LU, LV, MT) to 39 (DE), so NUTS3 regions are dropped from the adjustment in some of the smaller countries, while NUTS1 regions are used in the largest countries.

Possible limitations of this approach include:

- Auxiliary variables cannot be incorporated unless they are available for the population (and not only the gross sample);
- Population targets come from a survey, which is also subject to same two error sources as ESS (sampling variance and non-response);

[1] www.europeansocialsurvey.org

- The use of consistent target variables and categories (apart from region) in each country implicitly assumes similar error effects in each country.

RESEARCH QUESTION

Could the ESS weights be improved by adding a step of explicit non-response adjustment prior to the calibration, based upon sampling frame variables?

FEASIBILITY

The number of ESS countries providing micro-level auxilliary variables from the sampling frame or linked sources has increased in recent rounds. At Round 8 (2016-17), 18 out of 23 countries (78%) provided auxiliary microdata. Of these, 11 (43%) included one or more variables other than gender, age and region (which are already used in post-stratification). At R9, 24 out of 28 (86%) have included auxiliary microdata.

Aside from age, gender and region, the most commonly provided auxiliary variable is citizenship. Other variables provided by minorities of countries include language, country of birth, level of education, marital status and number of children.

Several countries also provide small-area indicators such as settlement type, population density, or economic activity profile. At a minimum, sampling strata could be used as adjustment variables in all countries.

It would therefore be feasible to introduce a step of sample-based non-response weighting to provide an adjusted design weight prior to the post-stratification stage. However, the variables and categories used would by necessity differ between countries. The association of auxilliary variables with non-response and with substantive variables of interest is of course also likely to vary between countries. It has been estimated that this additional step might require an additional amount of person time in the range of 2 weeks to 3 months, depending on the methods adopted (see questions, below).

NONRESPONSE MODELS

With minimum effort, a simple model (e.g. logit) could be fitted in each country, without variable selection, using all variables exactly as supplied. However there are several possible enhancements that are likely to improve the quality of the weighting but also to add significantly to the required researcher time:

- A prior step of combining categories and transforming variables, based on univariate empirical considerations;
- Stepwise fitting;
- Choosing between alternative categorisations/transformations, based on predictive empirical considerations.

Furthermore, it is unclear whether we should seek to influence the models in ways which may improve consistency between countries, e.g. by prioritising variables which are available in at least similar form across many countries, or by excluding variables that are only available in one or two countries.

Two simple case studies, for Finland and Germany, using ESS8 data, are presented in the appendix. These appear to suggest that there is some scope for the additional step of non-response adjustment to improve sample composition.

QUESTIONS FOR DISCUSSION

To what extent should we be concerned about inconsistencies between countries in a) the variables defining the nonresponse adjustment, and b) the impact of the adjustment on nonresponse error?

What criteria could be used for deciding the best approach to non-response weight adjustment?

How best should we assess whether weights are "improved" by any particular approach? (assuming that the subsequent 'post-stratification' step will be retained, and bearing in mind that between-country comparisons are a central use of the ESS data)

APPENDIX: TWO CASE STUDIES

Case Study 1: Finland ESS8

Of nine individual-level auxiliary variables supplied on the SDDF, only two (region [six regional state administrative agencies] and number of children) have no significant effect in a logistic regression model. The model, retaining significant predictors after forward stepwise fitting, is shown in table 1. Four of the seven predictors in the model are quite distinct from variables currently included in post-stratification.

Table 1: Logistic regression model predicting response, ESS8, Finland

Predictor	Odds Ratio	Std. Err.	P>z
Age group (ref: 20-29)			
15 – 19	1.723	.314	0.003
30 – 39	.9041	.128	0.478
40 – 49	.7369	.105	0.032
50 – 59	1.004	.142	0.978
60 – 69	1.221	.181	0.176
70 or over	1.174	.169	0.264
Marital status (ref: single)			
Married	1.454	.130	0.000
Divorced or legally separated	1.364	.172	0.014
Education (ref: Upper secondary)			
Short-term tertiary	1.391	.182	0.012
Bachelor degree	2.120	.286	0.000
Masters/Doctorate	2.141	.314	0.000
Other	0.680	.068	0.000
Gender (ref: male)			
Female	0.824	.061	0.009
Citizenship (ref: other)			
Finnish, Estonian, or Russian	2.300	.700	0.006
Language (ref: Finnish)			
Swedish or Norwegian	1.247	.205	0.178
Other	.6383	.125	0.022
Region (ref: urban)			
Semi-urban	1.186	.123	0.099
Rural	1.716	.193	0.000
Baseline odds	0.467	.151	0.018

Characteristics associated with a reduced propensity to respond are:

- Age 30-49
- Single (never married)
- Secondary or 'other' education only
- Female

- Not a citizen of Finland, Estonia or Russia
- First language not Finnish, Swedish or Norwegian
- Living in an urban area

It appears that weighting adjustment based on this model would correct some imbalances in the responding sample that would otherwise remain. Whether this improves substantive estimates notably remains an open question, however. We would welcome suggestions for exemplar analyses/estimates that could be used as criteria to test the impact of weighting and/or discussion of whether empirical testing of the effects of weights is appropriate.

Case Study 2: Germany ESS8

All five of the auxiliary variables contribute to the model (table 2). Characteristics associated with a reduced propensity to respond are:

- Age Group 7 (presumably, the oldest group?)
- Female
- Not a German citizen
- Living in East Germany
- Living in a larger municipality

Only two of these five variables (citizenship and municipality size) are distinct from the variables used for post-stratification.

Predictor	Odds Ratio	Std. Err.	P>z
Age group (ref: group 1)			
Group 2	0.671	.060	0.000
Groups 3-4	0.797	.060	0.003
Group 5-6	0.816	.063	0.008
Group 7	0.465	.045	0.000
Missing	0.566	.189	0.088
Gender (ref: male)			
Female	0.828	.038	0.000
Missing	2.029	.548	0.009
Citizenship (ref: German)			
Other	0.529	.046	0.000
Region (ref: West Germany)			
East Germany	0.867	.043	0.004
Municipality size (ref: smaller)			
Larger	0.772	.050	0.000
Baseline odds	0.901	.081	0.244

Notes: Municipality size was provided as a 10-category ordinal variable (without labels). There was a significant difference between the first 3 categories and the other 7 categories, but not otherwise between categories. Age groups are assumed to be ordinal, but are not labelled.