Technical Note No. 3

Note on Longitudinal Weights

Technical Report on the Longitudinal Sampling Weights for the Yemen National Social Protection and Monitoring Survey

Updated version

Marcel de Toledo Vieira
Departamento de Estatística
Universidade Federal de Juiz de Fora, Brazil

Cristiano Ferraz
Departamento de Estatística
Universidade Federal de Pernambuco, Brazil

Recife, October 2014.

1. Introduction

This report describes the statistical procedures adopted for calculating the longitudinal sampling weights of the Yemen National Social Protection and Monitoring Survey (NSPMS). The NSPMS has the Yemeni resident population (excluding non-household communities such as refugees, nomads and internally displaced persons, hotels, dormitories, prisons and hospitals) as its target population. The NSPMS is a longitudinal household survey that aims to provide parameter's estimates quarterly, and to accommodate the Social Welfare Fund (SWF) program impact assessment. Longitudinal NSPMS data currently available provide the necessary information for the implementation of the techniques described here.

This report is organised in three further sections. Section 2 revises the main aspects of the NSPMS sample design that guided the longitudinal sampling weights composition. Section 3 presents the building steps of the longitudinal sampling weights, and Section 4 includes some concluding remarks.

2. The NSPMS sample design

The NSPMS follows a two phase sampling design. In phase one, a stratified cluster sampling design with unequal selection probabilities is taken, where Enumeration Areas (EAs) are considered as the primary sampling units (clusters) selected within each governorate (stratum). In phase two, a stratified simple random sample of households is selected from each EA selected at the first phase. The second phase stratification is based on screening information raised at the first phase sample, and comprises three groups: one treatment and two control groups. Twelve households were sampled from each EA, as presented in Table 1. Thirty EAs are selected from each of the 21 governorates, providing a total sample size of 7560 households.

1

Table 1. *Second phase strata groups and sample size allocation*

Stratum	Description	Sample allocation
i)	Treatment	5
ii)	Control 1	5
iii)	Control 2	2

Further detailed descriptions on the NSPMS sample design can be found in Vieira and Ferraz (2012).

3. Longitudinal sampling weights

In the longitudinal surveys context two types of sampling weights have to be calculated: (i) cross-sectional weights at round t for use with single wave analyses for each round t; and (ii) longitudinal weights at round t for use with longitudinal analyses considering rounds up to round t.

Cross-sectional weights allows for new entrants and adjusts for non-response at each wave and, when population census data is available, population weighting adjustments that relate to population distributions at time t may be applied. Methodology adopted for calculating cross-sectional weights at round 1 of the NSPMS was described in Ferraz and Vieira (2013) and that was also adopted for calculating NSPMS cross-sectional weights for rounds 2, 3 and 4.

Alternative approaches may be adopted for developing longitudinal weights. In fact, we could calculate as many as 2^t -1 sets of longitudinal weights to allow for analysis of all possible combinations of non-response patterns in a panel of t rounds. However, as for most longitudinal surveys (and as for the British Househould Panel Survey, for example), we have adopted here the simplified approach which deals only with attrition non-response and which result in the necessity of calculating t sets of longitudinal weights. Under this approach, only cases who have responded at each round up to and including the latest round of the survey will have positive longitudinal weights at that round.

If we consider, for example, a survey with 2 waves, then the longitudinal weight at wave 2 could: (i) account for unequal selection probabilities at wave 1; (ii) adjust for unit nonresponse which may occur at waves 1 and 2; and (iii) adjust (via post-stratification, raking or calibration) so that weighted sample estimates for certain auxiliary variable match their respective known population parameters. Longitudinal weights, therefore, allow for different selection probabilities and nonresponse at wave 1 and attrition, and are adjusted, at each wave, to take account of previous wave respondents' absence through refusal at the current wave or through some other way of sample attrition. Longitudinal weights are calculated in order to guarantee the property that weighted sample moments are

consistent for population moments with respect to the joint sampling/nonresponse probability distribution.

Longitudinal sampling weights described in this report reflect the NSPMS sampling design and also the application of adjustment terms for dealing with unit non-response cases found at the first round of data collection process and attrition. The lack of currently available up-to-date population census data prevented us from applying calibration adjustments to the longitudinal weights.

Retaining notation used by Vieira and Ferraz (2014), and Ferraz and Vieira (2014), let the following quantities be defined:

- π_{hi} is the first phase inclusion probability of EA i within governorate h ;
- $\pi_{g\,j|hi}$ is the second phase conditional inclusion probability of household j within group g given the selected EA i within governorate h ;
- $\pi_{hig\,j} = \pi_{hi}\,\pi_{g\,j|hi}$. is the inclusion probability of household j within group g , at EA i of governorate h .

Moreover, let:

- $d_{higj} = 1/\pi_{hig\,j}$ be the basic cross-sectional sampling-design-weight for household j within group g , at EA i of governorate h .

In order to cope with the attrition problem, a longitudinal weighting adjustment procedure based on the simplified approach described above was adopted to correct the basic sampling-design-weight.

Let $\hat{\varphi}_{hig,t} = \dfrac{m_{h(r),t}}{m_h}\dfrac{n_{g/hi(r),t}}{n_{g/hi}}$ be the estimated propensity score for responses within group g at EA i whithin governorate h to be considered for the longitudinal weighting adjustment at round t. In this expression, m_h is defined as the number of EAs selected in governorate h , $m_{h(r),t}$ is the number of EAs that were surveyed at all rounds up to round t, $n_{g/hi}$ is the sample size in the classification stratum g within EA i of governorate h, and $n_{g/hi(r),t}$ is the number of households that responded the questionnaire at all rounds ut to round t in the classification stratum g within EA i of governorate h.

Then:

- $\tilde{w}_{hig\,j,t} = \dfrac{d_{hig\,j}}{\hat{\varphi}_{hig,t}}$ is the sample longitudinal weight at round t adjusted for unit nonresponse and attrition at governorate h, within EA i and group g.

Details on the NSPMS cross-sectional sampling weights can be found in Ferraz and Vieira (2014).

4. Concluding remarks

In this report, information of the NSPMS longitudinal sampling weights building process were presented. These information are useful for the release of official estimates based on longitudinal survey data, including estimates for changes over time and policy evaluation based upon impact econometric analysis, which should allow for the consideration of the longitudinal sampling weights.

Further information on methodology for the analysis of longitudinal complex survey data, considering weights and all the other sampling design features may be found in Vieira and Skinner (2008) and Vieira (2009). Moreover, both a theoretical and an empirical study on the impacts of complex sampling designs in longitudinal analysis of socio-economic and demographic data are presented in Skinner and Vieira (2007).

6. References

Ferraz, C. and Vieira, M. D. T. (2014) *Technical Report on the Cross-Sectional Sampling Weights for the Yemen National Social Protection and Monitoring Survey*. Technical Report. Brasília, International Policy Centre for Inclusive Growth.

Vieira, M. D. T. and Ferraz, C. (2014) *The Technical Report on the Sampling Design for the Yemen National Social Protection and Monitoring Survey*. Technical Report. Brasília, International Policy Centre for Inclusive Growth.

Skinner, C. J. and Vieira, M. D. T. (2007) Variance Estimation in the Analysis of Clustered Longitudinal Survey Data. Survey Methodology, v. 33, p. 3-12.

Vieira, M. D. T. (2009) Analysis of Longitudinal Survey Data. 1. ed. Saarbrücken: VDM Verlag Dr. Müller, v. 1, 246 p.

Vieira, M. D. T. and Skinner, C. J. (2008) Estimating Models for Panel Survey Data under Complex Sampling. Journal of Official Statistics, v. 24, p. 343-364.