Regular Local Sample Survey-based Monitoring for Food Security and Child Nutrition

Stephen J. Haslett
Massey University, Palmerston North, New Zealand s.j.haslett@massey.ac.nz

Abstract

Rapid, frequent, local monitoring of food security and child nutrition is a requirement for good food assistance allocation in a changing environment. Resources are however an issue, so that efficient survey designs that will also provide sound subpopulation estimates can have considerable benefit. The UN World Food Programme (WFP) in Nepal has been monitoring food security since 2002. WFP field surveillance capacity consists of a database management system (e-WIN) and integrated electronic data collection using tablets for field surveillance staff. The system allows rural, field-based household food security monitoring and analysis in near real-time. The annual sample is approximately 4,000 households, all of which are essentially rural. Data collected includes food security, market situation, water and sanitation, migration patterns, and child nutrition. This household survey is one of the core components of the Nepal Food Security Monitoring System (Nepal Khadhya Surakshya Anugaman Pranali: NeKSAP), currently being institutionalized into the Nepal government monitoring system. The data collected has also been used for non-food security purposes such as nutrition (Helen Keller International - HKI, Nepal Ministry of Health and Population), education (Research Input and Development Action - RIDA, UNICEF, Nepal Ministry of Education), and child protection (UNICEF). The survey design has evolved in line with changing information requirements. In 2010, probability sampling was introduced to give better representation of seasonality and geographical area, subject to the continuing to limit the survey to essentially rural areas, but in 2010 and 2011 no estimates of accuracy were calculated. Further major revision of the sample design in 2011 involved a complete redesign of the sample to improve estimates of quarterly and annual change, and to provide measures of accuracy (i.e. standard errors). Increased accuracy was achieved by use of rotation sampling, which divided the sample into four nationally-based subgroups in each quarter, resampling after initial selection in the following quarter, the following year and the following year plus one quarter. In each quarter, one new rotation group is introduced and one dropped, so that the four rotation groups sampled in each quarter have been in the sample 1, 2, 3 and 4 times respectively. This type of household survey design, which WFP intends extending to and implementing in other countries in which it provides food assistance, will be used to illustrate the methodology and various practical aspects of using rotation sampling for repeated regular monitoring of food security and child nutrition.

Key Words: composite estimation, correlation patterns, repeated sample surveys, rotation sampling, survey design.

1. Introduction

Rotation sampling was first proposed in the 1940s by Jessen (1942) and developed for more than two periods by Patterson (1950). It has been in use internationally since the 1960's. See, for example, Bellhouse (1989), Blight & Scott (1973), Eltinge (1994), Fuller, Adam & Yansaneh (1992), Gbur (1988), Haslett (1986), Kalton & Citro (1993), Park, Choi, & Kim (2007), Towhidi & Namazi-Rad (2010), Williams & Mallows (1970), and Wolter (1979).

Basically, rotation sampling gains its extra accuracy for estimating quarterly and annual change in comparison with a new sample each period by dividing the sample into groups each of which reflects the sample and the population as a whole. At each time point, one or more rotation groups are introduced and stay in the sample for a fixed number of periods (or rotate in and out of the sample over a fixed period), while other(s) are dropped. The patterning of inclusion and exclusion is carefully controlled, so that if annual as well as quarterly change estimates are of interest for a quarterly survey, then rotation groups will not be finally dropped until six quarters after their introduction to the sample.

Because there is generally a relatively high positive correlation over time for variables collected from households within each rotation group, time series correlation is the source of the method's efficiency.

The method is particularly good when estimating changes between time periods (usually quarterly or annually, but also monthly for monthly surveys), so has considerable potential for monitoring.

Rotation schemes also generally involve somewhat less fieldwork than choosing a completely new sample each period, since fieldwork staff return to a number of households that have already been visited. The method can also to some extent mitigate the long term response fatigue associated with cross-sectional / panel surveys, since in rotation designs respondents stay in the survey only for a limited number of periods.

2. Background

WFP Nepal has been implementing the food security monitoring since 2002. The WFP field surveillance capacity is composed of an advanced database management system (e-WIN) and field surveillance staff (34 field monitors) initially equipped with Personal Digital Assistance (PDA) and satellite phone with modem connection, and now using tablet and internet connection, which enables provision to decision makers of field-based monitoring information and analysis in near real-time. On an annual basis, the system until end-2010 collected some 4,000 household observations and the data include a variety of thematic areas such as food security, market situation, water and sanitation, and migration patterns. In 2010 a child nutrition module, which collects data on Infant and Young Child Feeding (IYCF) indicators, was introduced in collaboration with Helen Keller International (HKI). See also Tiwari (2010) and Government of Nepal, WFP & FAO (2010).

The data collected through the system has been used by different thematic stakeholders for purposes other than food security, such as nutrition (HKI, Ministry of Health and Population), education (RIDA, UNICEF, Ministry of Education) and child protection (UNICEF). The food security monitoring and analysis system (NeKSAP) is currently being institutionalised into the Nepal government system.

Household monitoring sampling design has evolved in line with changing information requirements. See also Ojha (2004). In 2010, probability sampling was introduced to NeKSAP to achieve better representation of seasonality and geographical area, subject to the continuing limitation of the survey to what are essentially rural areas.

The further revision of the sample design at the end of 2011 involved two main parts:

1. Modularising, as well as amending and supplementing the questions in the questionnaire.
2. Redesigning the sample to improve estimates of quarterly and annual change, and provide measures of accuracy (i.e. standard errors) for survey estimates.

The 2011 redesign also considered alignment with the 2010 Nepal Living Standards Survey (NLSS-III), with an aim of comparing findings from NeKSAP and NLSS-III for the rural parts of Nepal, and where possible track changes over time using NLSS-III as a baseline.

The 2011 NeKSAP redesign's rotation sampling scheme divided the sample into four nationally-based rotation groups in each quarter, resampling after initial selection in the following quarter, and the same quarters in the following year. In each quarter, one new rotation group is introduced and one dropped, so that the four rotation groups sampled in each quarter have been in the sample 1, 2, 3 and 4 times respectively, with the exception of the initial period when some rotation groups are dropped sooner.

3. NeKSAP sample re-design

The intent was to retain as many features of the 2010 design as possible, both for ease of implementation on a tight time line, and because the number of field monitors and their location place limitations on feasible designs for the NeKSAP survey. The re-design also needed to be flexible so it can be expanded either in terms of sample size or questionnaire length should additional field resources become available. Statistics of interest from NeKSAP include measures of both level (i.e. the current value) and change (e.g. differences between successive quarters or years) for a range of food security and other variables, and measures derived from them.

The 2011 design and the 2012 re-design for NeKSAP consist of 12 strata, as detailed in Table 1:

Table 1: Sample frame: 2011 Nepal Census of Housing and Population

Stratum	Stratum number	Total wards	Total households
Mountain East, Central and West	1	26	2229
Mountain Far-West	2	10	854
Mountain Mid-West	3	12	706
Rural Hills Central	4	41	6237
Rural Hills East	5	35	3300
Rural Hills Far-West	6	18	1526
Rural Hills Mid-West	7	29	3234
Rural Hills West	8	55	5614
Rural Terai Central	9	52	7533
Rural Terai East	10	34	6715
Rural Terai Mid-West and Far-West	11	15	4208
Rural Terai West	12	19	3410

For the 2012 re-design, wards (i.e psu) within strata were selected with probability proportional to the number of households they contain, based on the provisional results of the 2011 Nepal Census of Housing and Population. Equal numbers of households were selected in each sampled ward.

Four rotation groups are sampled in each quarter. Each rotation group contains two wards or primary sampling units (psu) per stratum, making a total sample size of 12 x 4 x 2 = 96 psu per quarter. With 10 households sampled with equal probability within wards, this gives a sample of 96 x 10=960 households a quarter, and 3840 households per year.

- There are four rotation groups in the sample in every quarter.
- Each rotation groups contains the same primary sampling units (psu) / clusters / wards and the same household in every quarter in which it is included.
- For each variable of interest (e.g. Food Consumption Score – FCS) the estimate from each rotation group in each quarter is calculated separately, along with its estimate of accuracy (i.e. standard error) in a way that allows for the stratification and clustering.
- The four estimates in each time period are independent of one another, but each is positively correlated with the estimate from earlier periods for the same rotation group if it has been sampled in any previous quarters.
- It is this correlation that improves estimates of change, since the rotation pattern retains each rotation group for two successive quarters, drops it, and then re-introduces it for exactly the same quarters in the following year.

Each rotation group of the twenty rotation groups in total for a five year period needs two clusters / wards / psu sampled per stratum, making 40 psu to be sampled per stratum for all 20 rotation groups combined. Note that the design for each rotation group is consequently a two psu per stratum design, and at each time point (since there are four rotation groups being used) there are 12 x 2 x 4=96 psu sampled in total. With 10 units / households selected per sampled psu, this gives a total sample of 960 households per quarter and 960 x 4=3840 households per year. Note however that because the rotation pattern resamples the same households, this does not correspond to 3840 different households per year, but to 24 x 9 x 10=2160 different households, since there are nine different rotation groups used each year. Note that the effective sample size is increased relative to a new sample of 960 household every quarter, through the efficiency gains derived from the correlation of some rotation groups with data collected in earlier quarters.

4. Estimation

- Combining the estimates uses a linear model, i.e. least squares, in a way that accounts for the correlation between the results from the same rotation group in different quarters.
- Estimates of level and change between quarters or years all come from the fitting of this model. The model provides one overall estimate of level or change by combining the estimates from all sampled rotation groups up to the current time.

The simplest and earliest example of a rotation sampling scheme is in Jessen (1942), where there are two time periods with equal sample sizes in each, and the sample selected in the first period contains some but not all of units selected in the first period. Using the same schemata as before the rotation scheme for this design is simply as given in Table 2:

Table 2: Rotation Groups for simple repeated survey

	1	2
1	x	
2	x	x
3		x

Note: Rows denote rotation groups, 1, 2 & 3; columns denote times 1 &2. Rotation group 1 is the unmatched units at time 1, rotation group 2 is the matched units at times 1 & 2, rotation group 3 is the unmatched units at time 2.

Consider one variable collected in the survey (or derived from it) at both times one and two, and let us find the best estimate of the means β_1 and β_2 at times one or two, and of the difference between the two time periods, β_1 - β_2 .

Let $\overline{Y}_{ui}$ = mean of unmatched units in time period *i*, with sampling error e_{ui} , for *i=1, 2*

$\overline{Y}_{mi}$ = mean of matched units in time period *i*, with sampling error e_{mi} , for *i=1, 2*

ρ = correlation between matched units at times 1 and 2

β_i = parameter of interest at time *i*, for *i=1,2.*

u= number of unmatched units at times 1 and 2
m = number of matched units at times 1 and 2
n. = sample size at each of times 1 and 2, so that *n=u+m*
S^2 = the common population variance.

Note that the sample does not need to be a simple random sample, but if it is not then means must be computed as weighted means based on inverse of selection probability, so must correlations, and u and m must be replaced by their effective sample sizes u^* and m^* from the complex design used.

Then in matrix notation

$$\begin{pmatrix} \bar{Y}_{u1} \\ \bar{Y}_{m1} \\ \bar{Y}_{u2} \\ \bar{Y}_{m2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} e_{u1} \\ e_{m1} \\ e_{u2} \\ e_{m2} \end{pmatrix}$$

or more compactly

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \qquad (1)$$

Now the variance of $\boldsymbol{Y}$ is given by

$$var(\boldsymbol{Y}) = \boldsymbol{\Sigma} = S^2 \begin{pmatrix} 1/u & 0 & 0 & 0 \\ 0 & 1/m & 0 & \rho/m \\ 0 & 0 & 1/u & 0 \\ 0 & \rho/m & 0 & 1/m \end{pmatrix}$$

The best linear unbiased estimate (BLUE) of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \qquad (2)$$

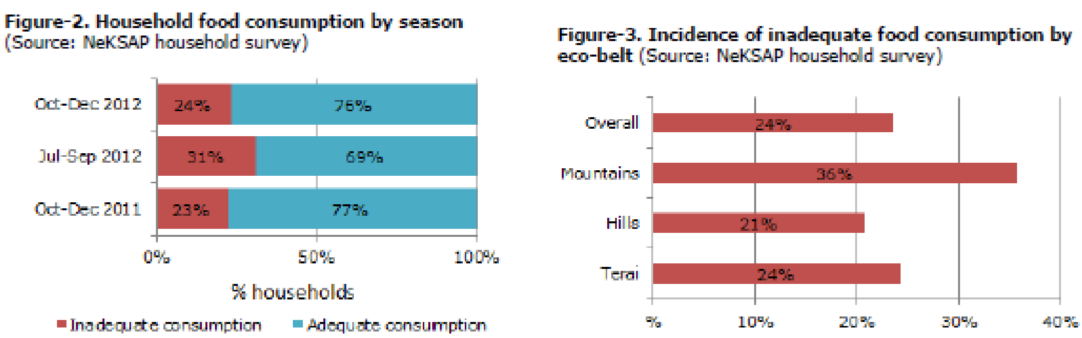
with

$$var(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1} \qquad (3)$$

The situation for the NeKSAP rotation scheme is a little more complicated, but the same principles still apply - as do the numbered equations (1) to (3) once the relevant matrices have been suitably redefined. Further detail is given in Haslett (2011).

5. Results

Figure 1: Household food consumption by season and area.



Source: Figures 2 and 3 from Nepal Food Security Bulletin, Issue 37, January 2013, UN World Food Programme.

Survey data from the NeKSAP redesign are used to estimate change from quarter-to-quarter and year-to-year. Although local estimates are calculated, standard errors recommend caution. An example of the results is given in Figure 1. Quarterly reports are published by WFP in its *Nepal Food Security Bulletin*.

References

Bellhouse, D. R. (1989) Optimal estimation of linear functions of finite population means in rotation sampling, *Journal of Statistical Planning and Inference*, 21, 69-74.

Blight, B. J. N. and Scott, A. J. (1973) A stochastic model for repeated surveys, *Journal of the Royal Statistical Society, B*, 35, 61-66.

Eltinge, J. L. (1994) Comparison of time and cross-sectional aggregation under a time series random component model, *Journal of Time Series Analysis*, 15, 167-181.

Fuller, W. A., Adam, A. and Yansaneh, I. S. (1992) Estimators for longitudinal surveys with application to the U.S. Current Population Survey, *Statistics Canada Symposium*, 309-324, Statistics Canada (Ottawa).

Government of Nepal, WFP & FAO (2010) *Nepal Khadhya Surakshya Anugaman Pranali (NeKSAP) Nepal Food Security Monitoring System: Framework Document*, Government of Nepal, UN World Food Programme, UN Food and Agriculture Organization.

Gbur, E. E., Jr (1988) Rotation sampling, *Encyclopedia of Statistical Sciences*, Volume 8, 193-195 Banks, David L. (ed.), Read, Campbell B. (ed.) and Kotz, Samuel (ed.), Wiley, New York.

Haslett, S. J. (1986) *Time series methods and repeated sample surreys*, Unpublished PhD Thesis, Victoria University of Wellington, http://researcharchive.vuw.ac.nz/handle/10063/971

Haslett, S. J. (2011) NeKSAP *Household Food Security and Child Nutrition Monitoring Survey: 2012 Questionnaire Re-design, Survey Re-design, Estimation, and Calibration with 2010 NLSS-III*, Report prepared for UN World Food Programme.

Jessen, R. J. (1942) Statistical investigation of a sample survey for obtaining farm facts. *Iowa Agric. Exp. Sta. Res. Bull*. 304.

Kalton, G. and Citro, C. F. (1993) Panel surveys: Adding the fourth dimension, *Survey Methodology*, 19, 205-215.

Ojha, D. P. (2004) Inventory of Poverty and Impact Monitoring (PIM) Approaches in Nepal, Poverty Monitoring Unit, GTZ, Nepal http://siteresources.worldbank.org/INTPAME/Resources/Country-studies/nepal_pim.pdf

Park, Y. S., Choi, J. W. and Kim, K. W. (2007) A balanced multi-level rotation sampling design and its efficient composite estimators, *Journal of Statistical Planning and Inference*, 137, 594-610.

Patterson, H. D. (1950) Sampling on successive occasions with partial replacement of units, *Journal of the Royal Statistical Society B* 12, 241-255.

Tiwari, B. (2010) *District Poverty Monitoring and Analysis System in Nepal*, http://tribhuvan.academia.edu/BishwaTiwari/Books/1462913/District_Poverty_Monitoring_and_Analysis_System_in_Nepal

Towhidi, M. and Namazi-Rad, M-R. (2010) An optimal method of estimation in rotation sampling, *Advances and Applications in Statis*tics, 15, 115-136.

Williams, W. H. and Mallows, C. L. (1970) Systematic biases in panel surveys due to differential nonresponse, *Journal of the American Statistical Association*, 65, 1338-1349.

Wolter, K. M. (1979), Composite estimation in finite populations, *Journal of the American Statistical Association*, 74, 604-613.