Estimating Quarter and Annual Labor Statistics in Pooled Monthly Surveys[1]

Nikkin L. Beronilla, and Quindale Caraos

Abstract

The Labor Force Survey (LFS) was originally conducted once every quarter to estimate labor statistics in the Philippines. Beginning 2021, however, the LFS is conducted monthly in response to the volatile nature of employment due to CoVID-19. To simplify the survey operations, it uses dependent replicates between regular (i.e., quarterly) and additional monthly surveys. However, the trade-off is that the computation of variance in pooled surveys is complicated as the regular-to-additional-month covariance is not equal to zero. As a result, additional steps are required to account for non-zero covariances. Hence, this paper aims to demonstrate a method to compute the variance of labor statistics that captures replicate dependence.

Introduction

The Labor Force Survey (LFS) is a quarterly survey conducted in the first month of the quarter (i.e., January, April, July, and October). A resolution in December 2020, however, was passed by the Philippine Statistics Authority (PSA) Board mandating a monthly conduct of LFS in response to volatile nature of employment due to CoVID-19. Beginning 2021, additional monthly surveys are conducted in-between the regular surveys. This produces eight (8) additional monthly surveys for a total of twelve (12) surveys. The sample size of the regular surveys[2] consist of four (4) replicates (roughly 44 thousand households), except for July which consists of 16 replicates[3]. The sample size of the monthly surveys consists of one (1) replicate which is selected randomly from one of the replicates of the preceding regular survey.

The use of dependent replicates greatly simplifies the survey implementation between regular and additional monthly surveys given the limited time. However, the trade-off comes later when the surveys are pooled for quarter and annual estimates. The computation of variance in pooled surveys is complicated as the regular-to-additional-month covariance is not equal to zero. In 2021 LFS, the variance computation requires additional steps to account for non-zero covariances. Ignoring the non-zero covariance will underestimate the variance of labor statistics. Hence, this paper aims to demonstrate a method to compute the variance of labor statistics that captures replicate dependence.

Variance Estimation

The LFS is a two-stage cluster design, with the barangay or portion of a large barangay, or combination of small barangays as the primary sampling units (PSUs) and the households as

[1] The authors are grateful to the staff of the Philippine Statistics Authority for their helpful comments and suggestions. The usual disclaimer applies.

[2] For clarity, original quarterly surveys are also referred as regular survey while the additional monthly surveys are referred as monthly surveys.

[3] July LFS coincided with the Family Income and Expenditure Survey.

the secondary sampling unit (SSUs).[4] Since it is done monthly beginning 2021, computing for quarter and annual labor statistics are done by pooling these surveys. Computing for the associated variances requires computing the associated regular-to-additional-month covariance. This is done by matching the individuals that were interviewed twice, in the regular survey and in the monthly survey.

For simplicity of notations, the two-stage sampling design is ignored, and a simple-random sampling is assumed. The actual computation in the R software, though, captures the true sampling design of the LFS which is a two-stage cluster design by using ReGenesees package.

The PSA releases two groups of labor statistics, in totals such as count of the employed, and in ratios such as employment rate. In a quarter, let Y_m be the labor variable in totals (e.g., employment totals) in a month, m which could be the regular and the two additional-month surveys. If the three survey rounds are pooled, the average employed in a quarter, q is,

$$\hat{\hat{Y}}_q = (1/3)\sum_{m=1}^{3}(\hat{Y}_m). \tag{1}$$

Since the samples of the two additional-month surveys are selected randomly from one of the four replicates (i.e., subsets) of the regularly survey, there is an overlapping of subpopulations (i.e., individuals that were interviewed twice) which introduces covariance to the estimates. To capture the regular-to-additional-month covariances, the variance of the averaged employed in a quarter is now composed of the monthly variances and two covariances as shown as,

$$\hat{V}(\hat{\hat{Y}}_q) = \left(\frac{1}{3}\right)^2 \left(\hat{V}(\hat{Y}_1) + \hat{V}(\hat{Y}_2) + \hat{V}(\hat{Y}_3) + 2cov(\hat{Y}_1, \hat{Y}_2) + 2cov(\hat{Y}_1, \hat{Y}_3)\right), \tag{2}$$

where the first covariance is between the first month and the second month, while the second covariance is between the first month and the third month. This entails additional step such as matching the data for individuals that were interviewed twice to capture the covariance.

Equation 1 can be extended to compute the average total in a year, A as,

$$\hat{\hat{\mathrm{Y}}}_A = (1/12)\sum_{m=1}^{12}(\hat{Y}_m). \tag{3}$$

The associated variance in a year, can also be computed by extending equation 2, as shown below,

$$\begin{aligned}\hat{V}(\hat{\hat{Y}}_A) = (1/12)^2(&\hat{V}(\hat{Y}_1) + \hat{V}(\hat{Y}_2) + \hat{V}(\hat{Y}_3) + 2cov(\hat{Y}_1, \hat{Y}_2) + 2cov(\hat{Y}_1, \hat{Y}_3) + \\ &\hat{V}(\hat{Y}_4) + \hat{V}(\hat{Y}_5) + \hat{V}(\hat{Y}_6) + 2cov(\hat{Y}_4, \hat{Y}_5) + 2cov(\hat{Y}_4, \hat{Y}_6) + \\ &\hat{V}(\hat{Y}_7) + \hat{V}(\hat{Y}_8) + \hat{V}(\hat{Y}_9) + 2cov(\hat{Y}_7, \hat{Y}_8) + 2cov(\hat{Y}_7, \hat{Y}_9) + \\ &\hat{V}(\hat{Y}_{10}) + \hat{V}(\hat{Y}_{11}) + \hat{V}(\hat{Y}_{12}) + 2cov(\hat{Y}_{10}, \hat{Y}_{11}) + 2cov(\hat{Y}_{10}, \hat{Y}_{12})).\end{aligned} \tag{4}$$

[4] PSA website (https://psada.psa.gov.ph/index.php/2013-master-sample-design)

Equation 4 contains twelve monthly variances and eight regular-to-additional-month covariances to capture the variance of average total correctly. This also requires matching the regular to month data for individuals that were interviewed twice.

For ratios, let X_m be another labor variable (e.g., labor force) in a month so that the ratio estimate (e.g, employment rate) in a quarter is,

$$\hat{R}_q = \hat{\hat{Y}}_q / \hat{\hat{X}}_q, \tag{5}$$

and its variance is approximated based on the Taylor series expansion as,

$$\hat{V}(\hat{R}_q) \approx (1/\hat{\hat{X}}_q^2)[\hat{V}(\hat{\hat{Y}}_q) + \hat{R}_q^2 \hat{V}(\hat{\hat{X}}_q) - 2\hat{R}_q cov(\hat{\hat{Y}}_q, \hat{\hat{X}}_q)]. \tag{6}$$

Equations 5 and 6 can be extended for annual ratios but are not shown here.

Results

The equations above are applied to the survey of the first three months of 2021 as an illustration. The table below shows the totals in the first quarter of 2021 2estimates. The coefficient of variation (CV) is slightly larger if the covariance is correctly accounted for, than if it is ignored.

Table 1. Totals in the first quarter, 2021, using proposed method

	Totals	**If covariance is accounted**		**If covariance is ignored**	
	(,000)	SE	CV	SE	CV
Total Population, ≥15	74,788	947.88	1.27%	810.28	1.08%
Labor Force	47,105	622.67	1.32%	535.09	1.14%
Employed	43,244	532.96	1.23%	497.53	1.15%
Unemployed	3,860	88.80	2.30%	91.68	2.37%
Underemployed	7,258	205.19	2.83%	189.20	2.61%

The same pattern can be observed for ratios, the CVs are generally higher if covariance is correctly accounted for. This means that in both totals and ratios ignoring the covariance of individuals that were interviewed twice would underestimate the variance and consequently the coefficient of variation.

Table 2. Rates in the first quarter, 2021, using the proposed method

	Rates	**If covariance is accounted**		**If covariance is ignored**	
		SE	CV	SE	CV
LF Participation Rate	62.98%	0.78%	1.24%	0.23%	0.36%
Employment Rate	91.80%	1.03%	1.12%	0.17%	0.19%
Unemployment Rate	8.20%	0.19%	2.35%	0.17%	2.08%
Underemployment Rate	16.78%	0.47%	2.81%	0.39%	2.35%
LF Participation Rate	62.98%	0.78%	1.24%	0.23%	0.36%

Summary and Recommendation

Using a sub-sample or a replicate of a regular survey for surveys of additional months offers an ease of implementation given a limited time to plan for preparation like during the pandemic. This, however, comes with a trade-off later once the monthly surveys are pooled quarterly or annually in the form of a complicated variance computation. This trade-off comes in matching the data of individuals that were interviewed more than once to capture the variance correctly.

The result above shows that if covariance is correctly accounted for, the resulting variance is generally larger, if the covariance is positive. This means that if there is no overlapping of samples between months, the covariances in equation 2 and 4 would be equal to zero which simplifies the computations. Therefore, it has been recommended that in the succeeding years, it would beneficial to use an independent sample per month to minimize the variance and simplify its computation.

References

Ericta, C., and Fabian. E [2009] “A Documentation of the Philippines Family Income and Expenditure Survey”, *Discussion Paper Series No. 2009-18*. Philippines: Philippine Institute for Development Studies.

Kish, L., [1987] *Statistical Design for Research*, New York: John Wiley & Sons.

Lohr, S.L., [199] *Sampling: Design and Analysis*. California: Duxbury Press, Pacific Grove.

Philippine Statistics Authority [2015] “Master Sample Design”, https://psada.psa.gov.ph/index.php/2013-master-sample-design. Accessed 15 June 2022.

R Core Team [2022] “R: A language and environment for statistical computing” R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed 15 June 2022.

Zardetto D. [2021] "ReGenesees: R Evolved Generalized Software for Sampling Estimates and Errors in Surveys". R package version 2.1, Istat, Italy.