Example QMCM_A_8

1. Introduction

The register-based employment statistics of Norway is disseminated annually by Statistics Norway. The employment variable measures whether or not a person is employed during the third week of November, and is constructed for each person in the target population by means of micro-integration. During the micro-integration process, several administrative registers are linked on micro-level and the information is harmonised before the integration ends in a classification of a person as employed or not employed.

We want to evaluate the accuracy of register-based employment statistics, REG-employment, for small areas such as municipalities, of which there are about 430 in Norway, and more than half of them have less than 5,000 inhabitants. The REG-employment data can be considered as surrogate data for a traditional census, and in Norway the traditional census was replaced by REG-employment for the Population Census 2001.

In the presence of a traditional census, the census would then have the role of target data, since it is a natural target or "benchmark" for evaluating the quality of estimates based on REG-employment data. In the absence of a traditional census, we will use the Labour Force Survey (LFS) to evaluate the quality of the REG-employment data. In particular, we will use the LFS-data to estimate the bias of the REG-employment estimator. This REG-employment estimator simply uses the REG-employment data to count the employed and unemployed persons per municipality.

Since the REG-employment data (in principle) cover the entire population, we can assume that there is no sampling variance in the REG-employment estimator. The estimator may be biased, though. We aim to estimate the bias of the REG-employment estimator.

2. Calculating the bias of REG-employment

Following QMCM_A_8, the estimated bias of the REG-employment estimator is given by

$$\tilde{b}_i = \gamma_i \bar{X}_i + (1-\gamma_i) u_i \hat{\beta}$$

where

$$\gamma_i = \frac{\sigma_v^2 c_i^2}{\psi_i + \sigma_v^2 c_i^2}$$

and

$$\hat{\beta} = \sum_i \frac{u_i \bar{X}_i}{\psi_i + \sigma_v^2 c_i^2} \left[\sum_i \frac{u_i^2}{\psi_i + \sigma_v^2 c_i^2} \right]^{-1}$$

Here

- i is the index for municipality i.
- $\bar{X}_i$ is the difference between the subgroup REG-employment estimator for municipality i, $\bar{Z}_i$, and survey-based output estimator for municipality i, $\bar{Y}_i$. That is, $\bar{X}_i = \bar{Z}_i - \bar{Y}_i$. Here, $\bar{X}_i = \frac{1}{N_i}\sum_{k\in i} X_{ik}$, $\bar{Y}_i = \frac{1}{N_i}\sum_{k\in i} Y_{ik}$ and $\bar{Z}_i = \frac{1}{N_i}\sum_{k\in i} Z_{ik}$, where Y_{ik} and Z_{ik} are values of the k-th person in municipality i, $X_{ik} = Z_{ik} - Y_{ik}$, and N_i is the number of persons in municipality i.

- $c_i = u_i = 1$ if intra-municipality quality heterogeneity is not taken into account, and $c_i = u_i = N_{gi}/N_i$ (where N_{gi} is the number of persons in group g of municipality i) if intra-municipality quality heterogeneity is taken into account.
- ψ_i is the variance of the survey-based output estimator for municipality i. This is assumed known and has to be estimated before the procedure described in QMCM_A_8 and the current document is applied.
- σ_v^2 is estimated using the iterative Fay-Herriot method suggested in Rao (2003; Section 7.1.2), where the $(a+1)$-th iteration is

$$\sigma_{v;(a+1)}^2 = \sigma_{v;(a)}^2 + \frac{m - p - h\left(\sigma_{v;(a)}^2\right)}{h'\left(\sigma_{v;(a)}^2\right)}$$

with m the number of municipalities (430 in this case), p the number of area level covariates (in this case $p = 0$),

$$h\left(\sigma_{v;(a)}^2\right) = \sum_i \left(\bar{X}_i - u_i\hat{\beta}\right)^2 \left(\psi_i + \sigma_v^2 c_i^2\right)^{-1}$$

and

$$h'\left(\sigma_{v;(a)}^2\right) = -c_i^2 \sum_i \left(\bar{X}_i - u_i\hat{\beta}\right)^2 \left(\psi_i + \sigma_v^2 c_i^2\right)^{-2}$$

3. Results

We use the REG-employment data and the LFS data for all municipalities for which the LFS net sample size is at least two persons to estimate the model parameters $\hat{\beta}$ and $\hat{\sigma}_v$.

When we assume no intra-municipality quality heterogeneity, i.e. when we use $c_i = u_i = 1$, the parameter estimates become $\hat{\beta} = 0.00302$ and $\hat{\sigma}_v = 0.03825$. This means that the estimated bias of the REG-employment estimator for municipality i is given by

$$\tilde{b}_i = \frac{(0.03825)^2}{\hat{\psi}_i + (0.03825)^2}\bar{X}_i + \left(1 - \frac{(0.03825)^2}{\hat{\psi}_i + (0.03825)^2}\right)0.00302 \qquad (1)$$

When we do assume intra-municipality quality heterogeneity, i.e. when we use $c_i = u_i = N_{gi}/N_i$, the parameter estimates become $\hat{\beta} = 0.00914$ and $\hat{\sigma}_v = 0.09228$. This means that the estimated bias of the REG-employment estimator for municipality i in the case of intra-municipality quality heterogeneity is given by

$$\tilde{b}_i^H = \frac{(0.03825)^2}{\hat{\psi}_i + (0.03825)^2}\bar{X}_i + \left(1 - \frac{(0.03825)^2}{\hat{\psi}_i + (0.03825)^2}\right)0.00302 \qquad (2)$$

where the superscript indicates that we assume intra-municipality quality heterogeneity.

Given values for $\hat{\psi}_i$ and $\bar{X}_i$, which can be easily be calculated from the REG-employed data and the LFS data, the estimated bias of the REG-employment estimator for municipality i can be found by plugging these values into (1), respectively (2).

References

Fosen, J. and L.-C. Zhang (2011), *Quality Assessment of Register-Based Census Employment Status*. Proceedings of the International Statistical Institute, World Congress, Dublin.

QMCM_A_8 (2018), Deliverable of WP 3 of the ESSnet on Quality of Multisource Statistics (SGA 2).

Rao, J.N.K. (2003), *Small Area Estimation*. Wiley Series in Survey Methodology, Wiley.