Forum Statistika dan Komputasi : *Indonesian Journal of Statistics*
ISSN : 0853-8115
Vol. 20 No.2, Oktober 2015, p: 1-4
available online at:
journal.ipb.ac.id/index.php/statistika

ESTIMATION OF UNEMPLOYMENT RATE USING SMALL AREA ESTIMATION MODEL BASED ON A ROTATING PANEL NATIONAL LABOR FORCE SURVEY

Siti Muchlisoh[1], Anang Kurnia[2], Khairil Anwar Notodiputro[2] ,I Wayan Mangku[2]

[1]Statistics Indonesia, Indonesia
[2] Department of Statistics, Bogor Agricultural University, Indonesia
E-mail : sitim@bps.go.id[1]

ABSTRACT

In Indonesia, labor force participation data are collected by Sakernas (National Labor Force Survey). Sakernas is conducted based on a quarterly rotating panel survey. Because of the groups differ according to their time-in-panel and observation strategy, it is possible to the presence of a bias. Besides, there are insufficiency problem of sample size to obtain an adequate precision of direct estimation at the district level. It is necessary to study how to estimate parameter based on a rotating panel survey when sample size is insufficient. Currently, a small area estimation (SAE) model that accomodates the bias component due to the rotation still only assume the effect over time which follows a random walk process, so it is necessary to develop a model that is more general. We propose a SAE model for rotation group level, its combined idea of the time-series multi-level model and the Rao-Yu model. The model will applied to Sakernas data to estimate a quarterly unemployment rate at the district level.

Key words : Sakernas, rotating panel survey, time-series multi-level model and Rao-Yu model

INTRODUCTION

Background

In many countries, including Indonesia, a rotating panel survey are used to collect data on labor force participation. The distinguishing feature of these survey with conventional panel survey is that the respondents (or household) divided into some parts called rotation group, each group is a subsample of the complete sample. Each periode one rotation group is introduced into the survey. Household in a rotation group is interviewed during some periods, then dropped at a certain periods and substituted by household from a new rotation group. There are existence of bias, its caused by rotating group is called rotation group bias (Bailar, 1975). Each country has different rotation design.

In Indonesia, labor force participation data are collected by Sakernas (National Labor Force Survey). Sakernas has been regularly conducted since 1976 by Statistics Indonesia (BPS), but the rotating panel survey started from 2011. Rotating panel survey of Sakernas are designed on quarterly basis by maintaining 3/4 groups of the previous quarter and add 1/4 of new groups of the current quarter. The main purpose of Sakernas is to obtain information about unemployment rate and its changes over time. The quarterly survey of Sakernas is designed only for estimating the parameters at the provincial level. Estimation of parameters for the district level only conducted in 3-rd quarter in each year. It requires 3 times more than the number of samples for each quarter which implies the requirement of additional cost and time. Also, the official quarterly unemployment rates is estimated based on only cross-sectional methods despite the fact that the data are collected under rotating panel survey.

There are three issues, that is 1) insufficient sample size, 2) possibility of the presence of bias caused by the rotation group and 3) the gaps in the parameter estimation. To solve these issues, it is necessary to study how to estimate parameter for the district level based on rotating panel survey when sample size is insufficient. Small area estimation (SAE) model is one of an

alternative to estimate parameter for small area. In the context of SAE, an area or domain becomes small when its sample size too small for direct estimation of adequate precision (Rao, 2003). One of SAE model for panel data is the Rao-Yu model. The model developed by Rao and Yu (1992, 1994). The model is not specifically developed for panel rotation data, so have not facilitated the possibility of the presence of bias that caused by the rotation. The model is an extension of the basic Fay-Herriot model (1979) by adding a random area-time component which follows an autoregressive process order 1 or AR (1).

Some researchers have been modify the Rao-Yu model related to the constraints of autoregressive coefficient. Datta et al. (2002) have used restricted maximum likelihood estimation (REML) under the random walk model. Fay and Diallo (2012) and Fay et al. (2013) have modified the Rao-Yu model by removing the stationarity requirement and modifying the random effects. They call the model as a dynamic model. They used the maximum likelihood (ML) estimation and restricted maximum likelihood (REML) estimation and they have successfully implemented an iterative Newton-Raphson algorithm in R software. Furthermore, Diallo (2014) also has been generalized the Rao-Yu model.

Boonstra (2014) developed a multi-level time-series models for small area estimation based on the data panel rotation. The model is used to estimate municipal unemployment based on the Dutch LFS at a quarterly frequency. The Dutch LFS uses a rotating panel design which in each quarter the total sample consists of five groups. The models add a component of bias due to the rotation. The rotation group biases are handled in the model by including measurement effects for the second to fifth waves. The model includes random municipality effects as well as random municipality-by-quarter effect. The municipality-by-quarter effects are modeled as random walk over time.

Related to three issues that described at 3rd paragraph, we propose a SAE model, its combine idea of Boonstra (2014) and the Rao-Yu model (1992, 1994). The focus of this study will be on Sakernas. The paper was structured as follows. Section 2 briefly described about Sakernas, section 3 described about the Rao-Yu model, section 3 described about the time-series multi-level model, section 4 described about the proposed model and section 5 described about methodology.

Sakernas

Started in 2011, Sakernas was designed based on three stages sampling design. First, select some primary sampling unit (PSU) by probability proportional to size (PPS) sampling. Second, select a census block by PPS sampling from the each PSU that was selected at the first stage. Third, select some household by linear systematic sampling from the each census block that was selected at the second stage. The total sample for West Java province was 400 census blocks (or 4000 households) for each quarter. The sample was allocated proportionally to 26 districts. The total sample for each district is about 8-21 census blocks (or 80-210 households) for each quarter. The sample was only about 0.3 percent of population (133,162 census blocks).

Sakernas is household survey conducted according to a rotating panel design. in the framework of the rotation panel, the total sample of census blocks are separated into four sample package (1,2,3,4). Every package, at each census block is formed 4 household groups. The groups at package-1 is A, E, I and M, package-2 is B, F, J and N, package-3 is C, G, K and O, and package-4 is D, H, L and P. So that in every quarter the total sample consists of four groups coming from different packages. Each group in the census block consists of about 10 households. Sakernas rotation panel designed on quarterly basis by maintaining 3/4 groups of the previous quarter and add 1/4 of new groups of the current quarter. FIGURE 1 illustrates the rotating panel design for February 2011 till February 2014. The same color identifies that the groups derived from the same census block. (BPS, 2014)

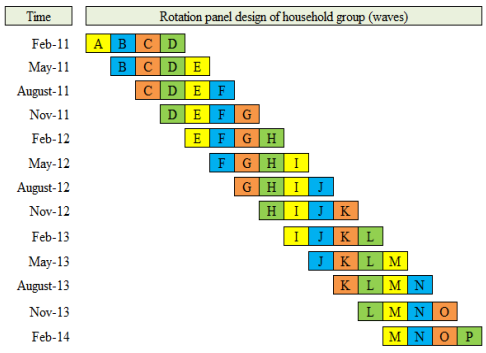

Figure 1. Sakernas rotating panel design

LITERATURE REVIEW

A. The Rao-Yu Model

Rao and Yu (1992, 1994) developed a SAE model which call as Rao-Yu model. The Rao-Yu model used for a combination of time series and cross-sectional data. The model is an extension of the basic Fay-Herriot model (1979) by adding a random area-by-time component which follows an autoregressive process order 1 or AR (1).

Let y_{it} be a direct survey estimator of i-th small area mean at time point t, say θ_{it} and y_{it} is assumed unbiased for θ_{it}. The θ_{it}'s are related to:

$$y_{it} = \theta_{it} + e_{it}. \quad (1)$$

The e_{it}'s are sampling errors that normally distributed with $E(e_{it}) = 0$ and known block diagonal covariance matrix $\boldsymbol{\Sigma}$. A vector of auxiliary variables, $\mathbf{x}_{it}$ related to θ_{it} is available such that:

$$\theta_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + v_i + u_{it}. \quad (2)$$

The $\boldsymbol{\beta}$ is a vector of regression coefficients, v_i is a random effect for small area i and u_{it} is a random effect for small area i at time point t. The Rao-Yu model is combining part of model (1) and (2):

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + v_i + u_{it} + e_{it}, \quad (3)$$
$$i = 1,2,\dots,m, \quad t = 1,2,\dots,T.$$

The $\boldsymbol{x}_{it}' = \left(x_{it1},\dots,x_{itp}\right)'$ is a vector of p fixed auxiliary variables for small area i at time t and $v_i \sim iidN(0,\sigma_v^2)$. The u_{it} 's are assumed to follow a first order autoregressive process within each area i,

$$u_{it} = \rho u_{i,t-1} + \varepsilon_{it}, \ |\rho| < 1, \quad (4)$$

with $\varepsilon_{it} \sim iidN(0,\sigma^2)$. The errors v_i, ε_{it} and e_{it} are assumed to be independent of each other. The condition $|\rho| < 1$ ensures stationarity of the series defined by (4) in order to obtain an autoregressive process of order 1.

B. The Time-Series Multi-Level Model

Boonstra (2014) developed a time-series multi-level model to estimate municipal unemployment based on the Dutch Labour Force Survey (LFS) at a quarterly frequency. The Dutch LFS is a household survey conducted according to a rotating panel design in which the respondents are interviewed five times at quarterly intervals, so that in each quarter the total sample consists of five waves. In the first wave of the panel, data are collected by means of computer assisted personal interviewing (CAPI), whereas in the four-follow-up waves, data are collected by means of computer assisted telephone interviewing (CATI).

Let y_{itp} be a direct survey estimator of *i*-th small area and *p*-th wave at time point *t*. There are 5 measurements y_{itp}, $p = 1,\dots,5$ for each small area estimated θ_{it}. The θ_{it}'s are related to:

$$y_{itp} = \theta_{it} + \mathbf{z}_{itp}'\boldsymbol{\alpha} + e_{itp}, \quad (5)$$
$$i = 1,\dots,m; \ t = 1,\dots,T; \ p = 1,\dots,5.$$

$$\theta_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + v_{it}. \quad (6)$$

The $\mathbf{z}_{itp}$ is only a measurement bias in the observed values y_{itp} and assumed that it does not predict the true variable θ_{it}. The $\mathbf{z}_{itp}$ to be a vector of dummy variables for the second to fifth waves, such an assumption is warranted. The corresponding coefficients, $\boldsymbol{\alpha}$ can be interpreted as the rotation group bias relative to the first wave, controlling for the covariates $\mathbf{x}_{it}$. The e_{itp}'s are sampling errors that normally distributed with $E(e_{itp}) = 0$ and known block diagonal covariance matrix $\boldsymbol{\Sigma}$. The v_{it} is an error term composed of multiple random effects.

The random effects term is taken to be a combination of some or all of 1) independent area effects v_i, 2) independent area-by-time effects w_{it}, and 3) area-by-time effects u_{it}, such that

$$v_{it} = v_i + w_{it} + u_{it}, \quad \text{with} \quad (7)$$
$$v_i \sim iidN(0,\sigma_v^2) \text{ and } w_{it} \sim iidN(0,\sigma_w^2).$$

$$u_{it} = u_{i;t-1} + \varepsilon_{it}, \quad \text{with} \quad (8)$$
$$\varepsilon_{it} \sim indN(0,\sigma_u^2).$$

For each area *i*, the u_{it} follow a random walk model over time with a common variance σ_u^2. Combining the measurement and structural parts of model (5) and (6) is:

$$y_{itp} = \mathbf{z}_{itp}'\boldsymbol{\alpha} + \mathbf{x}_{it}'\boldsymbol{\beta} + v_i + w_{it} + u_{it} + e_{itp}, \quad \text{with } u_{it} = u_{i;t-1} + \varepsilon_{it} \quad (9)$$

C. Proposed Model

As well as the Dutch LFS, Sakernas is also a household survey conducted according to a rotating panel design, but the respondents are interviewed four times at quarterly intervals, so that in each quarter the total sample consists of four waves. In all waves of panel, data are collected by same method that is direct interviewed. Cause of similarity in design panel rotation, we propose a model to estimate district unemployment based on the Indonesian Labour Force Survey (Sakernas) at a quarterly frequency. The proposed model

is combining idea of Boonstra (2014) and the Rao-Yu model (1992, 1994).

Let y_{itp} be a direct survey estimator of *i*-th small area and *p*-th wave at time point *t*. There are 4 measurements y_{itp}, $p = 1, \dots, 4$ for each small area estimated θ_{it}. The θ_{it}'s are related to:

$$y_{itp} = \theta_{it} + \mathbf{z}'_{itp}\boldsymbol{\alpha} + e_{itp}, \quad i = 1, \dots, m; \; t = 1, \dots, T; \; p = 1, \dots, 4. \tag{10}$$

$$\theta_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + v_{it}. \tag{11}$$

The $\mathbf{z}_{itp}$ to be a matrix design for the first to fourth waves, not as defined by Boonstra (2014). The e_{itp}'s are sampling errors that normally distributed with $E(e_{itp}) = 0$ and known block diagonal covariance matrix $\boldsymbol{\Sigma}$. The v_{it} is an error term composed of two random effects, such that,

$$v_{it} = v_i + u_{it}, \quad \text{with} \quad v_i \sim iidN(0, \sigma_v^2) \tag{12}$$

$$u_{it} = \rho u_{i,t-1} + \varepsilon_{it}, \quad |\rho| < 1, \tag{13}$$

with $\varepsilon_{it} \sim iidN(0, \sigma^2)$.

The v_i is independent area effects and the u_{it} is area-by-time effects. For each area *i*, the u_{it} follow a first order autoregressive process or AR(1). Combining the measurement and structural parts of model (10) and (11) is

$$y_{itp} = \mathbf{z}'_{itp}\boldsymbol{\alpha} + \mathbf{x}'_{it}\boldsymbol{\beta} + v_i + u_{it} + e_{itp}, \quad \text{with } u_{it} = \rho u_{i,t-1} + \varepsilon_{it}, \; |\rho| < 1 \tag{14}$$

METHODOLOGY

The data used was Sakernas data of West Java Province from February 2011 till February 2014 (13 quarters), data of village potential 2011 and census data 2010. The response variable is unemployment rate that estimated directly from Sakernas data. The auxiliaries variables come from data of village potential and census data. We will use the restricted maximum likelihood (REML) to estimate parameter.

REFERENCES

Bailar BA. 1975. The effects of rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association*. 70: 23-30.

Boonstra H. 2014. Time series small area estimation for unemployment based on a rotating panel surveys. *Discussion Paper*. Statistics Netherland.

[BPS] Badan Pusat Statistik. 2014. Pedoman pengawas survei angkatan kerja nasional (Sakernas) triwulanan. Jakarta (ID).

Datta GS, Lahiri P, Maiti T. 2002. Emperical bayes estimation of median income of four person families by state using time series and cross-sectional data. *Journal of Statistical Planning and inference*. 102: 83-97.

Diallo MS. 2014. Small area estimation under skew-normal nested error models [Ph.D. Thesis]. Carleton University. Canada.

Fay RE, Diallo MS. 2012. Small area estimation alternatives for the national crime victimization survey. *Proceedings of the Section on Survey Research Methods. American Statistical Association*. 3742–3756.

Fay RE, Planty M, Diallo MS. 2013. Small area estimates from the national crime victimization survey. *Proceedings of the Section on Survey Research Methods. American Statistical Association*. 1544–1557.

Fay RE, Herriot RA. 1979. Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*. 74: 269- 277.

Rao JNK, Yu M. 1992. Small area estimation combining time series and cross-sectional data. *Proceedings of the Section on Survey Research Methods. American Statistical Association.* 1-9.

Rao JNK, Yu M. 1994. Small area estimation combining time series and cross-sectional data. *Canadian Journal of Statistics*. 22(4), 511-528,

Rao JNK. 2003. *Small Area Estimation*. New York: John Wiley and Sons.