Asian Journal of Probability and Statistics

4(1): 1-17, 2019; Article no.AJPAS.49705

Small Area Procedures for Estimating Income and Poverty in Egypt

Mai M. Kamal El Saied[1*], Amal A. Talat[1] and Mervat M. El Gohary[1]

[1]Department of Statistics, Faculty of Commerce, Al Azhar University (Girls' Branch), Cairo, Egypt.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2019/v4i130106

Editor(s):

(1) Dr. Manuel Alberto M. Ferreira, Professor, Department of Mathematics, ISTA-School of Technology and Architecture, Lisbon University, Portugal.

Reviewers:

(1) Atilla Akbaba, İzmir Katip Celebi University, Turkey.

(2) Olumide Adesina, Olabisi Onabanjo University, Nigeria.

(3) Juliet Libanda, Jiangsu University of Science and Technology, China.

Complete Peer review History: http://www.sdiarticle3.com/review-history/49705

Original Research Article

Received: 26 March 2019
Accepted: 12 June 2019
Published: 19 June 2019

Abstract

In recent years, the demand for small area statistics has greatly increased worldwide. A recent application of small area estimation (SAE) techniques is in estimating local level poverty measures in Third World countries which is necessary to achieve the Millennium Development Goals. The aim of this research is to study SAE procedures for estimating the mean income and poverty indicators for the Egyptian provinces. For this goal the direct estimators of mean income and (FGT) poverty indicators for all the Egyptian provinces are presented. Also this study applies the empirical best/Bayes (EB) and the pseudo empirical best/Bayes (PEB) methods based on the unit level - nested error - model to estimate mean income and (FGT) poverty indicators for the Egyptian border provinces with (2012-2013) income, expenditure and consumption survey (IECS) data. The (MSEs) and coefficient of variations (C.Vs) are calculated for comparative purposes. Finally the conclusions are introduced. The results show that EB estimators for poverty incidence and poverty gap are smaller than PEB for all selected provinces. EB figures indicate that the largest poverty incidence and gap are for the selected municipality at the scope of the border south west of Egypt (New Valley). The PEB figures indicate that the largest poverty incidence and gap are for the selected municipality at the scope of the border north east of Egypt (North Sinai). As expected, estimated C.Vs for EB of poverty incidence and poverty gap estimators are noticeably larger than those of PEB estimators in all selected provinces.

**Corresponding author: E-mail: maikamaal@azhar.edu.eg, maikamaal@yahoo.com, maikamaal@hotmail.com;*

Keywords: SAE techniques; FGT poverty indicators; nested error model; empirical best/Bayes (EB); the pseudo empirical best/Bayes (PEB).

1 Introduction

For effective planning of health, social and other services, and for rationalizing government funds, there is a growing demand among various government agencies such as the U.S. Census Bureau, U.K. Central Statistical Office, and Statistics Canada to produce reliable estimates for smaller sub-populations, called small areas [1]. Small area estimation (SAE) was first studied at Statistics Canada in the seventies, Small area estimates have been produced using administrative files or surveys enhanced with administrative auxiliary data since the early eighties [2]. The terms "small area" and "local area" are commonly used to denote a small geographical area, such as a county, municipality or a census division. They may also describe a "small domain", a small subpopulation such as a specific age-sex-race group of people within a large geographical area [3]. Small area estimating quantities of interest for subpopulations (also known as domains) with survey data is a common practice. Domains can be defined by any characteristics that partition the population into a set of mutually exclusive subpopulations. Domain estimators that are computed using only the sample data from the domain are known as Direct Estimators (design-based estimators). [4] introduced one of the common approaches in direct estimation, Horvitz- Thompson estimator. Direct estimates often lack precision when domain sample sizes are small [5]. Due to cost and other considerations, sample surveys are typically designed to provide area-specific (or direct) estimators with small sampling coefficient of variation (CV) for large areas (or domains). In fact, survey practitioners often stress that non-sampling errors, including measurement and coverage errors, contribute much more than sampling errors to total mean squared error (MSE) which is often used as a measure of quality of estimators. In fact, sample sizes can be zero in many small areas of interest. Due to difficulties with direct estimators, it is often necessary to employ Indirect Estimates that borrow information from related areas through explicit (or implicit) linking models, using census and administrative data associated with the small areas [6]. Therefore the indirect estimation (model-based small area estimation) mainly uses two types of statistical models – implicit and explicit models. The implicit models provide a link to related small areas through supplementary data from census and/or administrative records; whereas the explicit models account for small area level variations through supplementary data [7]. Indirect estimation requires to go beyond the survey data analysis methods that are available [5]. The traditional indirect estimators are synthetic which introduced by [8], and composite which is a natural way to balance the potential bias of a synthetic estimator against the instability of a direct estimator by choosing an appropriate weight, see [3]. Synthetic and composite estimators, rely on implicit linking models. Indirect estimators based on explicit linking models have received a lot of attention in recent years because of the following advantages over the traditional indirect estimators based on implicit models:

(i) Explicit model-based methods make specific allowance for local variation through complex error structures in the model that link the small areas.
(ii) Models can be validated from the sample data.
(iii) Methods can handle complex cases such as cross-sectional and time series data, binary or count data, spatially-correlated data and multivariate data.
(iv) Area-specific measures of variability associated with the estimates may be obtained, unlike overall measures commonly used with the traditional indirect estimators [6].

So the explicit linking models provide significant improvements in techniques for indirect estimation. Based on mixed model methodology, these techniques incorporate random effects into the model. The random effects account for the between-area variation that cannot be explained by including auxiliary variables [5]. Explicit Linking Models are split into two main types; these types are known as area level model that is introduced by [9], and unit level model which is considered by [10], each type has many extension models that emerge from it. [11] provide an excellent account of the use of traditional and model-based indirect estimators in US Federal Statistical Programs. Text books on SAE have also appeared [12,13,14,15], and [16]. Good accounts of SAE theory are also given in the books by [17] and [18].

Both unit and area level models have been used extensively to estimate linear parameters such as totals and means. Poverty maps are an important source of information on the regional distribution of poverty and are currently used to support regional policy-making and to allocate funds to local jurisdictions. Good examples are the poverty and inequality maps produced by the World Bank for many countries all over the world [19]. Most poverty indicators are non-linear functions of a welfare variable such as income or expenditure. This makes many of the current small area estimation methods, typically developed for the estimation of linear characteristics, such as means, not applicable [20]. The first method designed to estimate general non-linear parameters in small areas is ELL method [21], used by the World Bank (WB) to construct poverty maps at local level. This method assumes a (unit level) linear mixed model which is presented by [10] for the log income or other variable used to measure the wellbeing. [22] have shown that the poverty estimates obtained by the ELL method can have poor accuracy. The empirical best (EB) method of [22] gives an approximation to the best estimates in terms of mean squared error (MSE), provided that the log incomes (or other one-to-one transformation of the welfare variable) are normally distributed. For estimation of general non-linear parameters in small areas, [20] proposed pseudo empirical best (PEB) method that incorporates the sampling weights and reduces considerably the bias of the un-weighted empirical best (EB) estimators under informative selection mechanisms.

This research is organized as follows. Section 2 introduces the unit level – nested error – model. The direct method, Empirical Best / Bayes (EB) method, and Pseudo Empirical Best / Bayes (PEB) method are introduced in Sections 3, 4 and 5 respectively. The parametric bootstrap MSE estimator is reviewed in Section 6. Section 7 shows the measures of inequality that are used. The estimation of mean income and poverty indicators (poverty incidence and poverty gap) for the Egyptian provinces with (2012-2013) IECS data is presented in the application within Section 8. Finally the conclusions are introduced in Section 9.

2 The Unit Level Nested Error Model

Let U be a finite population partitioned into $U_{i,}\ =1,2,\dots,m$ areas or domains. Each domain U_i has population size $N_i = 1,\dots,m$ where $N=\sum_{i=1}^{m} N_i$ the total population size. We denote by Y_{ij} the measurement of the study variable for j^{th} unit within i^{th} domain. Let H_i be a possibly non-linear domain parameters of interest, in the sense that it can be expressed as

$$H_i = \frac{1}{N_i}\sum_{j=1}^{N_i} h(Y_{ij}) \qquad i = 1\,,2\,,\dots,m \tag{1}$$

Where $h(.)$ is a real measurable function. Suppose that the population measurements Y_{ij} follow the nested error model introduced by [4],

$$\boldsymbol{Y}_{ij} = \mathbf{x}_{ij}^{`}\boldsymbol{\beta} + v_i + e_{ij};\ v_{i,}\sim N(0,\sigma_v^2)\ ,\ e_{ij}\sim N(0,\sigma_e^2),\ j = 1,2,\dots,N_i\,,\ i = 1,2,\dots,m. \tag{2}$$

Where $\mathbf{x}_{ij}$ is a $p\times 1$ vector of auxiliary variables, $\boldsymbol{\beta}$ is the $p\times 1$ vector of regression coefficients, $v_{i,}$ is area-specific random effects of the domain i, and e_{ij} is the individual regression error, where domain effects and errors are all mutually independent. Under that model, the area vectors $\mathbf{y}_i = (Y_{i1},\dots,Y_{iN_i})'$, $\mathbf{X} = \left(\mathbf{X'}_{i1},\dots,\mathbf{X'}_{iN_i}\right)'$; $\mathbf{e}_i = \left(e_{i1},\dots,e_{iN_i}\right)'$, $i = 1,2,\dots,m.$ and $y_i\sim N(\mu_i,\mathbf{V}_i)$, where $\mu_i = \mathbf{X}_i\beta$ and $\mathbf{V}_i = \sigma_v^2\mathbf{1}_{N_i}\mathbf{1'}_{N_i} + \sigma_e^2\mathbf{I}_{N_i}$, , $\mathbf{1}_k$ denotes a column vector of ones of size k, and $\mathbf{I}_k$is the $k\times k$ identity matrix. $\mathbf{y}_i = \left(\mathbf{y'}_{\mathbf{1}},\dots,\mathbf{y'}_{\mathbf{m}}\right)'$ denotes the population vector of measurements, $\mathbf{X} = (\mathbf{X'}_1,\dots,\mathbf{X'}_m)'$, is the population design matrix and $\boldsymbol{\theta} = (\boldsymbol{\beta}',\boldsymbol{\sigma_v^2},\boldsymbol{\sigma_e^2})'$ is the vector of unknown model parameters.

3 Direct Method

A direct estimator for a small area uses only sample data from the target area and it is usually design based. The definition of direct (point and variance) estimators in this research follows [23]. The mean helps to describe the distribution of a target variable, especially for target variables with a skewed distribution like income. Direct estimator of the mean is defined as follows:

$$\hat{\bar{\mathbf{y}}}_i = \frac{\sum_{j=1}^{n_i} w_{ij} y_{ij}}{\sum_{j=1}^{n_i} w_{ij}} \;, i = 1, \dots, m \tag{3}$$

where w_{ij} be the sampling weight (inverse of the probability of inclusion) of individual j from area i.

Direct estimators of the poverty indicators FGT that are defined as in Equation (4) at $\alpha = 0$ for the poverty incidences to be as Equation (5), and at $\alpha = 1$ for poverty gaps to be as Equation (6)

$$f_{\alpha,i} = \frac{1}{\sum_{j=1}^{n_i} w_{ij}} \sum_{j\in s_i} w_{ij} \left(\frac{z - E_{ij}}{z}\right)^{\alpha} I\left(E_{ij} < z\right), \alpha \geq 0, \quad i = 1, \dots, m \tag{4}$$

$$f_{0,i} = \frac{1}{\sum_{j=1}^{n_i} w_{ij}} \sum_{j=1}^{n_i} w_{ij}\, I\left(E_{ij} < z\right), \quad i = 1, \dots, m, \tag{5}$$

$$f_{1,i} = \frac{1}{\sum_{j=1}^{n_i} w_{ij}} \sum_{j=1}^{n_i} w_{ij} \left(\frac{z - E_{ij}}{z}\right) \quad I\left(E_{ij} < z\right), \quad i = 1, \dots, m\,, \tag{6}$$

Where $I\left(E_{ij} < z\right)$=1 if $E_{ij} < z$ (person under poverty) and $I\left(E_{ij} < z\right)$= 0 if $E_{ij} \geq z$ (person not under poverty). Indeed, a common definition of poverty classifies a person as "under poverty" when the selected welfare variable for this person is below 60% of the median.

4 Empirical Best / Bayes (EB) Estimator

This method assumes that the sampling design is non-informative for inference about y. Then, the outcomes corresponding to sampled units, Y_{ij}; $j \in s_i$, preserve the same distribution as the outcomes for out-of-sample units, given by (2) under the considered nested error model. Let us decompose the domain vector $\mathbf{y_i}$ into sub vectors corresponding to sample and out-of-sample elements as $\mathbf{y}_i = (\mathbf{y}_{is}', \mathbf{y}_{ir}')'$, where the subscript s denotes the sample units and r the out-of-sample units. The sample data is then $\mathbf{y}_s = (\mathbf{y}_{1s}', \dots \;, \mathbf{y}_{ms}')'$. For a general domain parameter $H_i = h(\mathbf{y}_i)$, the best predictor is defined as the function of the sample observations y_s that minimizes the mean squared error (MSE) and is given by

$$\widetilde{H}_i^B(\boldsymbol{\theta}) = E_{\mathbf{y}_{ir}}(H_i | \mathbf{y}_{is}; \boldsymbol{\theta}) \tag{7}$$

Where the expectation is taken with respect to the distribution of $\mathrm{y}_{ir}|\mathrm{y}_{is}$, which depends on the true value of $\boldsymbol{\theta}$. For a domain parameter H_i that is additive as in (1), the best predictor is reduced to

$$\widetilde{H}_i^B(\boldsymbol{\theta}) = \frac{1}{N_i}\left[\sum_{j\in s_i} h(Y_{ij}) + \sum_{j\in r_i} \widetilde{H}_{ij}^B(\boldsymbol{\theta})\right] \tag{8}$$

where $\widetilde{H}_{ij}^B(\boldsymbol{\theta}) = E\left[h\left(Y_{ij}\right)|\mathbf{y}_{is}; \boldsymbol{\theta}\right]$ is also the best predictor of $H_{ij} = h(Y_{ij})$ for out-of-sample unit $j \in r_i$. The best predictor $\widetilde{H}_{ij}^B(\boldsymbol{\theta})$ is exactly model unbiased for H_i regardless of the complexity of the function $h(.)$. However, it cannot be calculated in practice since model parameters $\boldsymbol{\theta}$ are typically unknown. An empirical best predictor (EB) of H_i, denoted as $\widehat{H}_i^{EB}$, is then obtained by replacing $\boldsymbol{\theta}$ in $\widetilde{H}_i^B(\boldsymbol{\theta})$ by a consistent estimator $\widehat{\boldsymbol{\theta}}$, that is, $\widehat{H}_i^{EB} = \widetilde{H}_i^B(\widehat{\boldsymbol{\theta}})$. The EB predictor is not exactly unbiased, but the bias arising from the estimation of $\boldsymbol{\theta}$ is typically negligible when the overall sample size n is large. Given the nested error model specified in (2) and assuming non-informative selection, the out-of-sample vectors y_{ir} given the sample data vectors y_{is} are independent and follow exactly the same distribution as $\mathbf{y}_{ir}|\bar{\mathbf{y}}_{is}$, where $\bar{\mathbf{y}}_{is}$ is the un-weighted sample mean for area i. Thus, the best predictor of $H_{ij} = h(Y_{ij})$ is $\widetilde{H}_{ij}^B(\boldsymbol{\theta}) = E\left[h\left(Y_{ij}\right)|\bar{y}_{is}; \boldsymbol{\theta}\right]$. For an out-of-sample observation Y_{ij} , $j \in r_i$,we have

$$Y_{ij}|\bar{y}_{is} \sim N(\mu_{ij|s}, \sigma^2_{ij|s}), j \in r_i \quad (9)$$

$$\mu_{ij|s} = \mathbf{x'}_{ij}\beta + \gamma_{is}(\bar{y}_{is} - \bar{\mathrm{x}}'_{is}\beta),\ \sigma^2_{ij|s} = \sigma^2_v(1-\gamma_{is}) + \sigma^2_e \quad (10)$$

For $\bar{\mathrm{x}}_{is} = n_i^{-1}\sum_{j\in s_i}\mathrm{x}_{ij}$ and $\gamma_{is} = \sigma^2_v/(\sigma^2_v + \sigma^2_e/n_i)$.

Foster et al. [24] introduced the family of FGT poverty indicators, which contain several widely-used poverty measures and which are additive in the sense described above. In particular, the poverty maps released by World Bank are traditionally based on members of this family. Let E_{ij} be a welfare measure for individual j in area i and z be the poverty line. The family of FGT poverty indicators for domain i is given by Equation (4), where $I(E_{ij} < z) = 1$ if $E_{ij} < z$, and $I(E_{ij} < z) = 0$ otherwise. For $\alpha = 0$, we obtain the poverty incidence, measuring the frequency of poverty. For $\alpha = 1$, we get the poverty gap, measuring the poverty depth. Both indicators together give a good description of poverty.

Consider that the model (2) holds for $Y_{ij} = \log(E_{ij} + C)$, where $C \geq 0$ is a constant. Then, we can express $F_{\alpha ij}$ in terms of the response variable Yij as

$$F_{\alpha ij} = \left[\frac{z - \exp(Y_{ij}) + C}{z}\right]^{\alpha} I\left[\exp(Y_{ij}) - C < z\right] = h_\alpha(Y_{ij}), \quad (11)$$

Which shows that $F_{\alpha ij} = N_i^{-1}\sum_{j=1}^{N_i} h_\alpha(Y_{ij})$ is an additive parameter in the sense of (1).

According to (8), the best predictor of $H_i = F_{\alpha i}$ is then given by

$$\tilde{F}^B_{\alpha i}(\theta) = \frac{1}{N_i}\left(\sum_{j\in s_i} F_{\alpha ij} + \sum_{j\in r_i}\tilde{F}^B_{\alpha ij}(\theta)\right) \quad (12)$$

Where $\tilde{F}^B_{\alpha i}(\theta) = E\left[h_\alpha(Y_{ij})|\bar{y}_{is};\theta\right]$ is the best predictor of $F_{\alpha ij} = h_\alpha(Y_{ij})$. For $\alpha = 0$; 1, the best predictor $\tilde{F}^B_{\alpha ij}(\theta)$ can be calculated analytically. Let us define $\alpha_{ij} = \left[\log(z+c) - \mu_{ij|s}\right]/\sigma_{ij|s}$, for $\mu_{ij|s}$ and $\sigma^2_{ij|s}$ given in (9) and (10). Then, the best predictors of F_{0ij} and F_{1ij} are respectively given by

$$\tilde{F}^B_{0ij}(\theta) = \mathbf{\Phi}(\alpha_{ij}) \quad (13)$$

$$\tilde{F}^B_{0ij}(\theta) = \mathbf{\Phi}(\alpha_{ij})\left\{\mathbf{1} - \frac{\mathbf{1}}{\boldsymbol{z}}\left[\exp\left(\mu_{ij|s} + \frac{\sigma^2_{ij|s}}{2}\right)\frac{\mathbf{\Phi}(\alpha_{ij} - \sigma_{ij|s})}{\mathbf{\Phi}(\alpha_{ij})} - c\right]\right\} \quad (14)$$

where $\mathbf{\Phi}(.)$ is the c.d.f. of a standard Normal random variable.

For additive area parameters $H_i = N_i^{-1}\sum_{j=1}^{N_i} h(Y_{ij})$ with more complex $h(.)$, analytical expressions for the expectation $E\left[h(Y_{ij})|\bar{y}_{ij};\boldsymbol{\theta}\right]$ defining the best predictor may not be available. In any case, the EB predictor $\hat{H}^{EB}_{ij} = E\left[h(Y_{ij})|\bar{y}_{ij};\hat{\boldsymbol{\theta}}\right]$ of a general $H_{ij} = h(Y_{ij})$ can be approximated by Monte Carlo, similarly as in [5]. This is done by simulating L replicates $\{Y^{\ell}_{ij}; \ell = 1, \dots, L\}$ of Y_{ij} , $j \in r_i$, from the estimated conditional distribution of $Y_{ij}|\bar{y}_{ij}$ given in (9), calculating the corresponding $h(Y^{\ell}_{ij})$ for each ℓ and then averaging over the L replicates as

$$\hat{H}^{EB}_{ij} = L^{-1}\sum_{\ell=1}^{L} h(Y^{(\ell)}_{ij}) \quad (15)$$

5 Pseudo Empirical Best / Bayes (PEB) Estimator

As stated above, under the nested error model (2), $\mathbf{y}_{ir}|\bar{y}_{is}$ follows exactly the same distribution as $\mathbf{y}_{ir}|\mathbf{y}_{is}$ and the best predictor of $H_{ij} = h(Y_{ij})$, $j \in r_i$, can be expressed as $\widetilde{H}_{ij}^{B} = E[h(Y_{ij})|\bar{y}_{is}]$. When the sample selection mechanism is informative, to avoid a bias due to a non-representative sample, the estimation procedure should incorporate the sampling weights. Let w_{ij} be the sampling weight of j^{th} unit within i^{th} domain and $w_i = \sum_{j\in s_i} w_{ij}$. We consider the same conditioning idea of the EB estimator, but now we condition on the weighted sample mean $\bar{y}_{iw} = w_i^{-1}\sum_{j\in s_i} w_{ij}y_{ij}$ instead of the un-weighted sample mean$\bar{y}_{is}$. Thus, we define the pseudo best (PB) estimator of $H_{ij} = h(Y_{ij})$, as

$$\widetilde{H}_{ij}^{PB}(\boldsymbol{\theta}) = E\left[h(Y_{ij})|\bar{y}_{iw};\boldsymbol{\theta}\right] \tag{16}$$

The PB estimator of the additive area parameter H_i is as [25] where they used a similar approach in the special case of area means under the nested error model and also in the case of a binary response variable and a logit linking model. Their method is applicable only for area level covariates in the unit level models. For example, when using the area mean vector $\overline{\mathbf{X}}_i = N_i^{-1}\sum_{i=1}^{N_i} \mathrm{x}_{ij}$ as area level covariates in the unit level model.

Similarly as in EB method, the PB estimator (16) depends on the true values of the model parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma_v^2, \sigma_e^2)'$, which need to be estimated. The PEB predictor is defined as the PB predictor with θ replaced by a consistent estimator. The approach of [26] based on the sample likelihood can be used to find correct maximum likelihood (ML) estimates of the regression parameter $\boldsymbol{\beta}$ and of the variances σ_v^2 and σ_e^2. Alternatively, $\boldsymbol{\beta}$ can be estimated using the weighted method of moments used in [27] and using ML (or REML) estimators of σ_v^2 and σ_e^2. For an out-of-sample variable Y_{ij} , $j \in r_i$, under the nested error population model (2), we have

$$Y_{ij}|\bar{y}_{iw} \sim N(\mu_{ij|s}^{w}, \sigma_{ir|s}^{2w}) \text{ , } \mu_{ij|s}^{w} = \mathbf{X}'_{ij}\beta + \gamma_{iw}(\bar{y}_{iw} - \bar{\mathbf{x}}'_{iw}\boldsymbol{\beta}), \sigma_{ij|s}^{2w} = \sigma_v^2(1-\gamma_{iw}) + \sigma_e^2, \tag{17}$$

where $\bar{\mathbf{x}}_{ij} = w_i^{-1}\sum_{j\in s_i} w_{ij}\mathbf{x}_{ij}$ and $\gamma_{iw} = \sigma_v^2/(\sigma_v^2 + \sigma_e^2\delta_i^2)$, for $\delta_i^2 = w_i^2\sum_{j\in s_i} w_{ij}^2$. Observe that the mean $\mu_{ij|s}^{w}$ is obtained from $\mu_{ij|s}$ given in (9) by replacing the un-weighted best predictor of the domain effect by its weighted version. Even if the conditional distribution (17) is obtained assuming that the sample units satisfy the same population model (2) (i.e. non-informative sampling), we will see that conditioning on the weighted sample mean $\bar{y}_{iw}$ protects against informative sampling.

For the FGT poverty indicators of order α = 0, 1, the PB are given by (13) and (14) with $\mu_{ij|s}$and $\sigma_{ij|s}^2$ replaced by the weighted versions $\mu_{ij|s}^{w}$ and $\sigma_{ij|s}^{2w}$. For more complex additive parameters, such as the FGT indicators for $\alpha > 1$, we can apply a Monte Carlo procedure to approximate the PEB predictor of $H_{ij} = h(Y_{ij})$ similarly as done for the EB predictor. We generate L replicates $\{Y_{ij}^{\ell};\ \ell = 1, \dots\ , L\}$ of $Y_{ij}, j \in r_i$ from the estimated conditional distribution of $Y_{ij}|\bar{y}_{iw}$ given in (17), calculate $h\ (Y_{ij}^{(\ell)})$ for each ℓ and then average over the L replicates as $\widehat{H}_{ij}^{PEB} = L^{-1}\sum_{\ell=1}^{L} h(Y_{ij}^{(\ell)})$.

6 Parametric Bootstrap MSE Estimator

Even though the PEB estimators that are presented in Section 5 incorporate the sampling weights, they are essentially model-based. Thus, estimators of the MSE of PEB estimators under the model are proposed here. The considered procedure is a similar bootstrap procedure as in [20], based on the parametric bootstrap method for finite populations introduced by [28]. The parametric bootstrap estimator of the model MSE of $\widehat{H}_i^{PEB}$ is obtained as follows: i) Fit the model (2) to the sample data $(\mathrm{y}_s, \mathbf{X}_s)$ and obtain estimators

$\hat{\boldsymbol{\beta}}, \hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$ of $\boldsymbol{\beta}, \sigma_v^2$ and σ_e^2 respectively. ii) For b = 1,…,B, with B large, generate $v_i^{*(b)} \sim N(0, \hat{\sigma}_v^2)$ and $e_{ij}^{*(b)} \sim N(0, \hat{\sigma}_e^2), j = 1, \dots, N_i, i = 1, \dots, m,$ independently. iii) Construct B iid bootstrap population vectors $\mathbf{y}^{*(b)}$, b = *1, … , B*, with elements $Y_{ij}^{*(b)}$ generated as

$$Y_{ij}^{*(b)} = \mathbf{x}\grave{}_{ij}\hat{\beta}_w + v_i^{*(b)} + e_{ij}^{*(b)}; \quad j = 1,2, \dots, N_i, \quad i = 1,2, \dots, m \tag{18}$$

From each bootstrap population b, calculate the true value of the domain parameter $H_i^{*(b)} = N_i^{-1} \sum_{j=1}^{N_i} h(Y_{ij}^{*(b)})$, b = *1, … ,B*. iv) From each bootstrap population b, take the sample with the same indices as the initial sample s and, using the sample elements $y_s^{*(b)}$of $y^{*(b)}$ and the known population vectors $x_{ij}, j \in U_i,$ calculate the bootstrap PEB predictors of H_i, denoted as $\hat{H}_i^{PEB*(b)}, b = 1, \dots, B$. v) A bootstrap estimator of the model MSE of the PEB estimator, $MSE_m(\hat{H}_i^{PEB})$.

$$MSE_m(\hat{H}_i^{PEB}) = \frac{1}{B} \sum_{b=1}^{B} \left(\hat{H}_i^{PEB*(b)} - H_i^{*(b)}\right)^2 \tag{19}$$

7 Measures of Inequality

One of the inequality measures for direct estimation is the inequality indicator Gini, which is defined as a ratio between 0 and 1 and is estimated by

$$\widehat{\text{Gini}} = \left[\frac{2\sum_{j=1}^{n_i} w_{ij} y_{ij} - \sum_{j=1}^{n_i} w_{ij}^2 y_{ij}}{\sum_{j=1}^{n_i} w_{ij} \sum_{j=1}^{n_i} w_{ij} y_{ij}} - 1\right] \tag{20}$$

The higher the value, the higher the inequality is. The extreme values of 0 and 1 indicate perfect equality and inequality, respectively. On the other hand, another important measure which is used to indicate the reliability of the estimators is the coefficient of variation (CV). It is a measure for showing the extent of the variability of the estimate [29]. The CV is used, for instance, by National statistical institutes (NSI) for quantifying the uncertainty associated with the estimates and is defined as follows,

$$CV = \frac{\sqrt{\widehat{MSE}\,(\hat{\zeta}_i)}}{\hat{\zeta}_i} \tag{21}$$

Where $\hat{\zeta}_i$ is an estimate of an indicator ζ_i for domain i and $\widehat{MSE}\,(\hat{\zeta}_i)$ is the corresponding mean squared error. Often, the coefficient of variation (CV), defined as the standard error of an estimate expressed as a ratio or a percent of the estimate, is used to decide whether an estimate is reliable or not. For instance, Statistics Canada follows the general rule which considers an estimate with a coefficient of variation less than 15% to be reliable for general use while estimates with a coefficient of variation greater than 35% are deemed to be unreliable (unacceptable quality). Statistics Canada recommends not publishing unreliable estimates (CV > 35%) and if published informing the public that the estimates are not reliable [30].

8 The Application

The aim of this study is to estimate the mean income and the poverty indicators which are the poverty incidences and the poverty gaps for the Egyptian provinces with (2012-2013) IECS data. The poverty incidence for a province is the province mean of a binary variable E_{ij} taking value 1 when the person's income is below the poverty line z and 0 otherwise. The considered welfare measure is 60% of the median for the annual total income. For that year, the calculated poverty line is 14946 EGP. The FGT measure in Equation (4) the poverty incidence at α = 0, and for α = 1 is called poverty gap which measure the area mean of the relative distance to non-poverty (the poverty gap) of each individual.

$$F_{\alpha i} = \frac{1}{N_i} \sum_{j=1}^{N_i} \left(\frac{z - E_{ij}}{z}\right)^{\alpha} I(E_{ij} < z), \alpha \leq 0, \quad i = 1, \dots, m,$$

The Central Agency for Public Mobilization and Statistics (CAPMAS) is preparing the income, expenditure and consumption surveys (IECS), which is considered one of the most important family surveys carried out by statistical agencies in different countries of the world. CAPMAS conducts survey every two years periodically. The (2012-2013) IECS - survey under study - was conducted to cover all governorates of the Arab Republic of Egypt. The sample design for (2012-2013) IECS used a two-stage stratified clustered sampling technique. Survey data collected over 12 months period from 1 July 2012 to the end of June 2013 through survey questionnaire. The survey included a sample of 7528 households (survey unit) distributed in 27 governorates. The data include basic information about members of household (such as gender, age, educational statue, labor statue...etc), data about the household expenditure and consumption behavior, data about the household sources of income, and finally the sample weights. For the purposes of the study, some modifications were made to the auxiliary variables classification. First of all the auxiliary variables related to the head of the household were used instead of all members of it. The data set contains unit-level data on income and other sociological variables in the Egyptian provinces. The statistical packages software, such as SPSS version 22, STAT version 12, SAS University Edition, Excel 2010, Access 2010 have been used for data preparation, data cleaning, imputation and summarizing.

8.1 Direct estimation results

Tables 1 and 2 show the results of the direct estimation which uses the sample data only. The R software with version 3.5.1 through package **emdi** with version 1.1.3 for 64 bit windows has been used to get the results of direct estimation parameters, (see [29]). These results can give a general review about the estimators under study for all the Egyptian provinces. Although we can recognize from Table 1 that the mean income has very large variance, but the C.V still small and less than 15%. The range of Gini coefficient is small and fall between 0.21 and 0.36.

The poverty indicators are presented in Table 2, we can note that both indicators either incidences or gaps have small variances.

8.2 Model based estimation results

The PEB estimates and EB of province poverty incidences and poverty gap based on nested error model are obtained for the variable income. The R statistical package **sea** with version 1.2 for 64 bit windows has been used to estimate model parameters, mean squared errors of estimates, model selection, diagnostics, graphical plots and other statistical analysis (R Core Team, 2018) according to [31]. Note that the PEB and EB methods assume that the response variable considered in nested error model is (approximately) normally distributed.

Normal Q-Q plot of EB and PEB residuals are included in Fig. 1 shows that the distributions of PEB residuals (on the left side) and EB residuals (on the right side) have slightly heavier tail than the normal distribution.

Fig. 2 shows normal Q-Q plot of estimates of weighted and unweighted area effects υ_i^{PEB} (in the left) and υ_i^{EB} (in the right) for each sampled municipality respectively. The distribution of estimated area effects is approximately similar to a normal distribution in the two plots.

To save computation efforts and time of the study, the PEB, EB estimates and their corresponding MSE estimates will be presented here only for 5 provinces. To uphold the concept of borrow strength from neighbors; the selected provinces are with the smallest sample sizes. These provinces are the Egyptian border provinces which include Red Sea, New Valley, Matrouh, North Sinai and South Sinai governorates.

The values of the dummy indicators are not known for the out-of-sample units, but the PEB and EB methods can be derived by the knowledge of the total number of people with the same x-values as in [22]. These totals were estimated using the sampling weights attached to the sample units in the IECS.

The PEB and the EB estimates for the mean income separated by the selected provinces with their MSEs and (C.Vs) are listed in Tables 3 and 4 respectively.

Table 1. Direct estimation of mean income with Egyptian pound (EGP)

ID	Province	Mean income ($\bar{y}_i$) EGP	Gini	Var.($\bar{y}_i$)	C.V.($\bar{y}_i$)
1	Cairo	37354.66	0.3591201	2030958.4	3.815098244
2	Alexandria	37038.65	0.3241597	1935915.1	3.756539888
3	Port Said	35125.45	0.2844535	9801400.6	8.912964426
4	Suez	51968.21	0.2894889	18770846	8.336891802
5	Damietta	29162.87	0.2653453	1859334.5	4.675720064
6	Dakahlia	27929.15	0.2639792	427292	2.340478534
7	Sharqia	30008.35	0.2408574	472837.3	2.291467742
8	Qalyubia	26622.78	0.2476557	468676.9	2.571481303
9	Kafr El Sheikh	32414.59	0.3091761	1768008.5	4.102056535
10	Gharbia	30791.32	0.2700054	547239.8	2.402484147
12	Beheira	28625.52	0.2511179	416510.6	2.254548823
12	Monufia	31302.76	0.2879109	1622056.5	4.068650239
13	Ismailia	34892.08	0.2656844	2254222.4	4.303001733
14	Giza	30567.54	0.3180021	947269.6	3.18402384
15	Beni Suef	27960.66	0.3039866	1434562.5	4.28363362
16	Faiyum	26821.85	0.2704065	1076489.2	3.868264028
17	Minya	29124.17	0.2873607	3319787.9	6.256070173
18	Asyut	25622.28	0.3198718	848382.3	3.594827263
19	Sohag	21457.39	0.2731217	351326	2.762347114
20	Qena	23508.65	0.297188	896471.7	4.027546858
21	Aswan	28738.33	0.2510043	1485542.8	4.241124846
22	Luxor	26651.04	0.2676741	1920577.7	5.199981276
23	Red Sea	45791.52	0.2617929	37531166.8	13.37860937
24	New Valley	37613.46	0.2123923	16678768.1	10.85772155
25	Matruh	38660.05	0.285335	14901690.1	9.98516745
26	North Sinai	31056.41	0.22474	3560481.4	6.075794958
27	South Sinai	33094.29	0.2207474	8857598.3	8.993006807

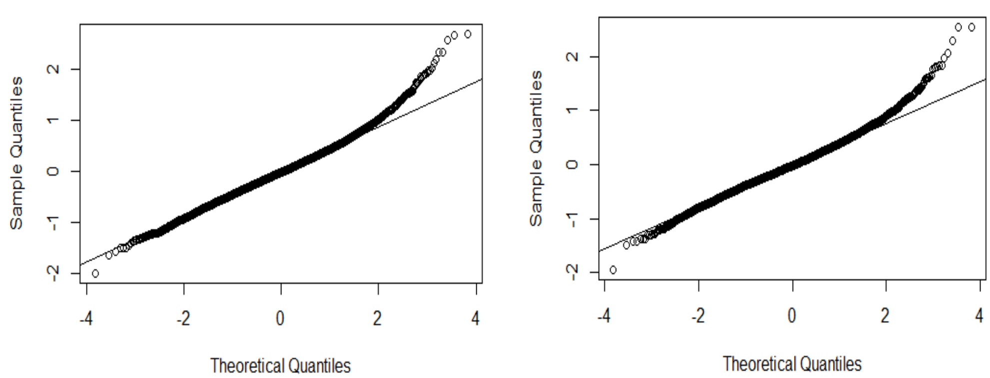

Fig. 1. Normal Q-Q plots of PEB and EB residuals

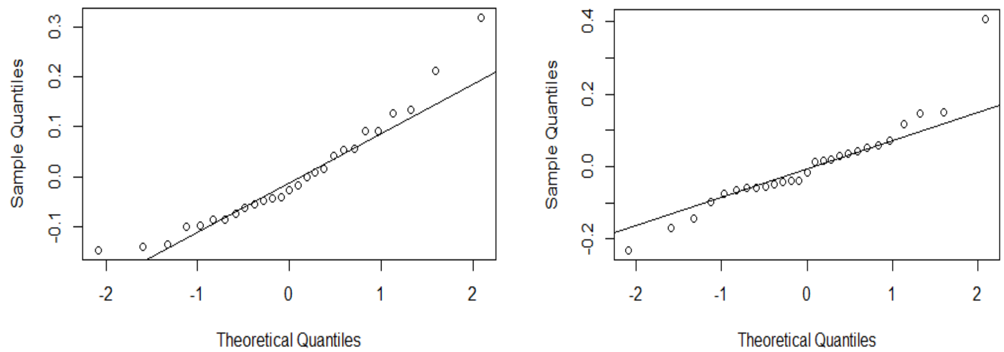

Fig. 2. Normal Q-Q plot of PEB and EB predicted municipality effects

Table 2. Direct estimation of poverty incidences and poverty gaps

ID	Province	Poverty incidence	Var. poverty incidence	Poverty gap	Var. poverty gap
1	Cairo	0.1221936	0.000117828	0.027358332	8.31E-06
2	Alexandria	0.09795272	0.000162555	0.020516961	9.01E-06
3	Port Said	0.04479355	0.000608363	0.0095196	2.77E-05
4	Suez	0.01614794	0.000221581	0.002270803	4.38E-06
5	Damietta	0.12089942	0.000932207	0.033379813	8.63E-05
6	Dakahlia	0.14966522	0.000250662	0.03609568	2.52E-05
7	Sharqia	0.08189572	0.000107115	0.021608298	1.19E-05
8	Qalyubia	0.14786936	0.000300165	0.028364387	1.16E-05
9	Kafr El Sheikh	0.12549267	0.000464397	0.029533177	5.27E-05
10	Gharbia	0.10539727	0.000353964	0.0245335	2.21E-05
12	Beheira	0.09678696	0.000134541	0.022919835	1.36E-05
12	Monufia	0.13238527	0.00031322	0.026038753	1.40E-05
13	Ismailia	0.05608687	0.000478576	0.017382378	6.66E-05
14	Giza	0.14632627	0.000224751	0.034122752	1.83E-05
15	Beni Suef	0.18891616	0.000564935	0.045846482	9.22E-05
16	Faiyum	0.15700208	0.000685377	0.032624829	4.15E-05
17	Minya	0.13973196	0.000340319	0.036710746	3.42E-05
18	Asyut	0.26869029	0.000807846	0.08054702	1.40E-04
19	Sohag	0.32263846	0.000505966	0.085394064	7.91E-05
20	Qena	0.28027316	0.000916053	0.077674435	1.00E-04
21	Aswan	0.11274453	0.001000174	0.021833663	7.65E-05
22	Luxor	0.17600056	0.002186758	0.04720513	2.61E-04
23	Red Sea	0.0251024	0.000694178	0.003530025	1.37E-05
24	New Valley	0	0	0	0.00E+00
25	Matruh	0.04990479	0.001145011	0.017524721	1.53E-04
26	North Sinai	0.04627757	0.000804524	0.007334556	5.14E-05
27	South Sinai	0.14193932	0.005455162	0.057755257	2.28E-03

Fig. 3 shows the PEB and the EB of the mean income separated by the provinces sample sizes (on the left side). According to this figure there is no noticeable difference between PEB and EB for all provinces except for the third one in sample size (Red Sea), the PEB in it is greater than the EB.

Also Fig. 3 shows the C.Vs for PEB and EB separated by the provinces sample sizes (on the right side). According to this figure the C.Vs for PEB are smaller than the C.Vs for EB in all provinces except the second one in sample size (New Valley), the C.V for PEB on it is greater than the C.V for EB. The estimated C.Vs are still under 15% for both methods in all selected provinces.

Table 3. Estimated population size (households), sample size, PEB estimates of Mean Income, estimated MSE of PEB estimates and C.Vs of PEB estimates

Province Name	$\hat{N}_i$	n_i	$\hat{\bar{y}}_i^{PEB}$	MSE	C.V.
Red Sea	14101	21	42908.97EGP	10330793	7.490637
New Valley	8687	20	34097.82 EGP	12513246	10.37429
Matruh	16771	30	27361.34 EGP	6518819	9.331407
North Sinai	35103	41	26609.11 EGP	5227968	8.592826
South Sinai	2934	12	25943.74 EGP	8682530	11.357704

Table 4. Estimated population size (households), sample size, EB estimates of Mean Income, estimated MSE of PEB estimates and C.Vs of PEB estimates

Province Name	$\hat{N}_i$	n_i	$\hat{\bar{y}}_i^{EB}$	MSE	C.V.
Red Sea	14101	21	41016.53 EGP	10389271	7.858391
New Valley	8687	20	34133.1 EGP	9739876	9.143258
Matruh	16771	30	27625.04 EGP	7103047	9.647603
North Sinai	35103	41	26755.64 EGP	5406914	8.690794
South Sinai	2934	12	26122.83 EGP	9205278	11.614438

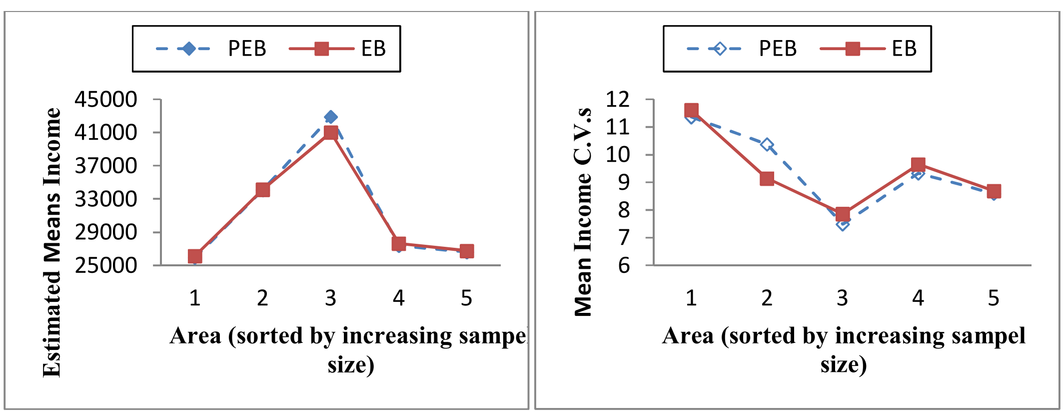

Fig. 3. The estimated mean income and coefficient of variations (C.Vs) for PEB and EB

The estimated mean income for the households under the poverty line for all of these five provinces is 10350 EGP with standard deviation 0.4773 EGP in PEB method, and 9026.83 EGP with standard deviation 0.4194 EGP in EB method.

Table 5. Estimated population size (households), sample size, PEB estimates of poverty incidence, estimated MSE of PEB estimates and C.Vs of PEB estimates. Estimated poverty incidence and C.Vs are in percentage

Province Name	$\hat{N}_i$	n_i	F_{0i}^{PEB}%-	$MSE\ F_{0i}^{PEB}$	$C.V\ F_{0i}^{PEB}$
Red Sea	14101	21	1.319537	0.00204277	3.4252154
New Valley	8687	20	2.566900	0.00018833	0.5346230
Matruh	16771	30	4.798941	0.00152137	0.8127790
North Sinai	35103	41	5.193091	0.00138425	0.7164413
South Sinai	2934	12	4.248642	0.00148297	0.9063906

The MSEs of the poverty measures for the selected domains were estimated by using the bootstrap procedure described in Section 6. Values of PEB estimates and (C.Vs) - in other words, estimated RRMSEs (Relative Root Mean Squared Error) - for the poverty incidence and the poverty gap are listed in Tables 5 and 6 respectively.

Table 6. Estimated population size (households), sample size, PEB estimates of poverty gap, estimated MSE of PEB estimates and C.Vs of PEB estimates. Estimated poverty gap and C.Vs are in percentage

Province Name	$\hat{N}_i$	n_i	F_{1i}^{PEB}%	$MSE\ F_{1i}^{PEB}$	$C.V\ F_{1i}^{PEB}$
Red Sea	14101	21	0.342435	0.000433333	6.079009
New Valley	8687	20	0.830101	0.000039691	0.758956
Matruh	16771	30	1.522847	0.000159927	0.830433
North Sinai	35103	41	1.747470	0.000151761	0.704969
South Sinai	2934	12	1.366086	0.000347382	1.364349

The EB estimates and (C.Vs) for the poverty incidence and the poverty gap are listed in Tables 7 and 8 respectively.

Table 7. Estimated population size (households), sample size, EB estimates of poverty incidence, estimated MSE of EB estimates and C.V of EB estimates. Estimated poverty incidence and C.V are in percentage

Province Name	$\hat{N}_i$	n_i	F_{0i}^{EB}%	$MSE\ F_{0i}^{EB}$	$C.V\ F_{0i}^{EB}$
Red Sea	14101	21	0.6017561	0.0018829813	7.211116
New Valley	8687	20	2.0752268	0.0001913543	0.666582
Matruh	16771	30	1.0695792	0.0004110843	1.895625
North Sinai	35103	41	1.6292255	0.000384532	1.203798
South Sinai	2934	12	1.3637135	0.0004198669	1.502563

Table 8. Estimated population size (households), sample size, EB estimates of poverty gap, estimated MSE of EB estimates and CV of EB estimates. Estimated poverty gap and CV are in percentage

Province Name	$\hat{N}_i$	n_i	F_{1i}^{EB}%	$MSE\ F_{1i}^{EB}$	$C.V\ F_{1i}^{EB}$
Red Sea	14101	21	0.1196754	0.00031345	14.7938865
New Valley	8687	20	0.5730147	0.00003267	0.9974976
Matruh	16771	30	0.2074845	0.00002191	2.2562051
North Sinai	35103	41	0.3441736	0.00002109	1.3341577
South Sinai	2934	12	0.3084063	0.00005441	2.3917750

Figs. 4 and 5 report the resulting estimates and the estimated coefficients of variation (C.Vs) for selected municipalities, obtained as estimated root MSE by the corresponding estimate in percentage.

The left side of these figures show that EB estimators for poverty incidence and poverty gap lie under PEB for all selected provinces. Additionally that the differences are large in three provinces (Matruh, North Sinai and South Sinai), and are small in two of them (Red Sea and New Valley).

As expected, the right side of Figs. 4 and 5 show that the estimated C.Vs of EB for poverty incidence and poverty gap estimators are noticeably larger than those of PEB estimators in all provinces. But the difference for the second province in sample size (New Valley) was small. In spite of the noticeable differences, the estimated C.Vs still under 15% for both methods in all selected provinces.

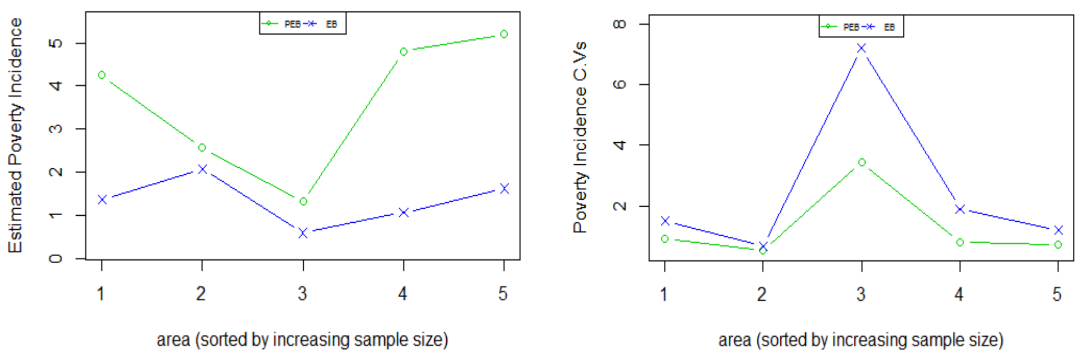

Fig. 4. The estimated poverty incidence and coefficient of variations (C.Vs) for PEB and EB

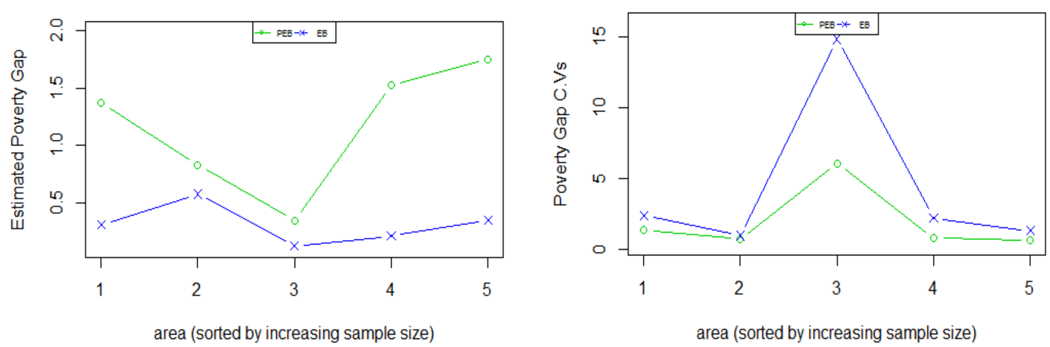

Fig. 5. The estimated poverty gap and coefficient of variations (C.Vs) for PEB and EB

Figs. 6 and 7 display cartograms of EB and PEB estimates of poverty incidence $F_{0,i}$ (on the left) for each of the selected municipalities. EB estimates provide a larger number of municipalities with poverty incidence in the third interval of poverty than PEB ones. EB figures indicate that the largest poverty incidence and gap are for the selected municipality at the scope of the border south west of Egypt (New Valley). The PEB estimates of poverty incidence are noticeably large from the third municipality in sample size to the last one (Matrouh, South Sinai, and North Sinai) respectively.

Figs. 6 and 7 show the analogous estimates for the poverty gap $F_{1,I}$ (on the right). The different poverty intervals and colors are considered for each method because the ranges of EB and PEB estimates were quite different. The PEB figures indicate that the largest poverty incidence and gap are for the selected municipality at the scope of the border north east of Egypt (North Sinai). We can see colors also tending to be darker for PEB estimates than for EB ones in the case of poverty incidence.

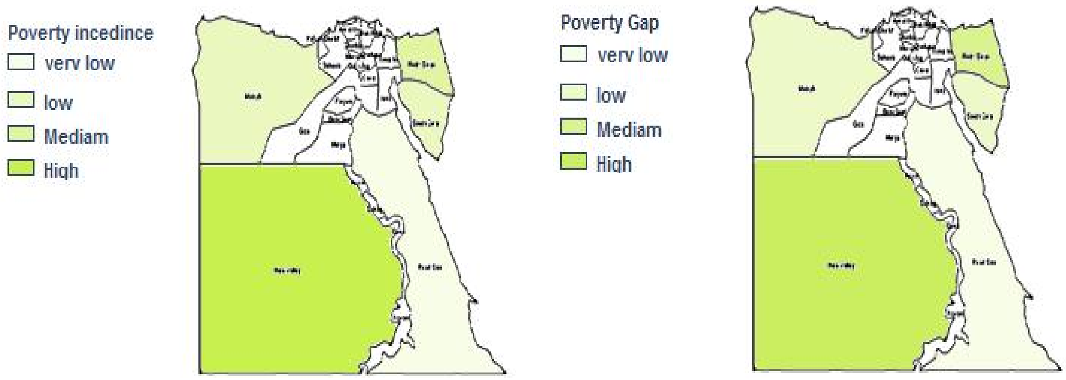

Fig. 6. Cartograms of Estimated Percent of Poverty Incidences and Gaps in the Selected Municipalities from Egypt, Obtained by EB Method

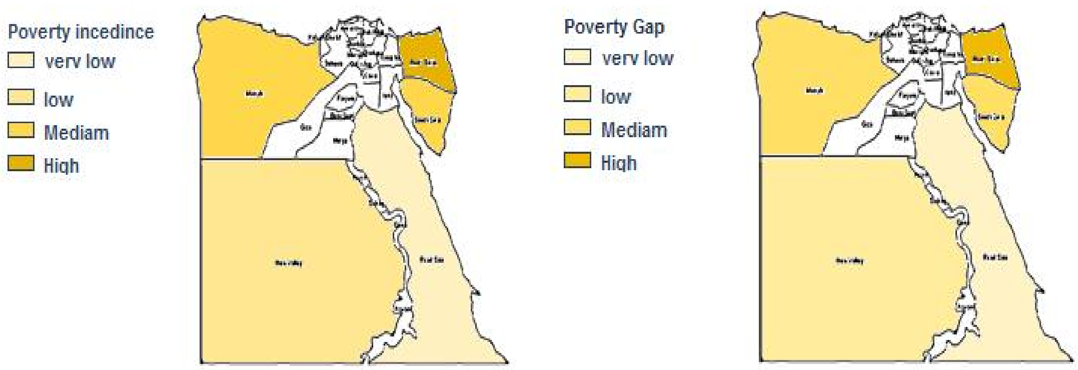

Fig. 7. Cartograms of Estimated Percent of Poverty Incidences and Gaps in the Selected Municipalities from Egypt, Obtained by PEB Method

9 Conclusion

The aim of this research is to study small area estimation procedures for estimating the mean income and poverty indicators (poverty incidences and gaps) for the Egyptian provinces with (2012-2013) IECS data. To make a general review about the estimators under study for all the Egyptian provinces, direct estimation was applied which uses the sample data only. Although that the estimated mean income with direct method has very large variance, but the C.Vs still small and less than 15%. The range of Gini coefficient of the estimated mean income is small and fall between 0.21 and 0.36. The estimated poverty incidence and gap by the direct method are calculated and have small variances. The results for estimated mean income show that PEB and the EB separated by the provinces sample sizes have no noticeable differences for all provinces except for the third one in sample size (Red Sea), the PEB in it is greater than the EB. The C.Vs for PEB are smaller than the C.Vs for EB in all selected provinces except the second one in sample size (New Valley), the C.V for PEB on it is greater than the C.V for EB. The estimated C.Vs are still under 15% for both methods in all selected provinces. EB estimates for poverty incidence and poverty gap are smaller than PEB for all selected provinces. Additionally that the differences are large in three provinces (Matruh, North Sinai and South Sinai), and are small in two of them (Red Sea and New Valley). As expected, estimated C.Vs for EB of poverty incidence and poverty gap estimates are noticeably larger than those of PEB estimates in all provinces. But the difference for the second province in sample size (New Valley) was small. In spite of the noticeably differences, the estimated C.Vs still under 15% for both methods in all selected provinces. The cartograms show that EB estimates provide a larger number of municipalities with poverty incidence in the third interval of poverty than PEB ones. EB figures indicate that the largest poverty incidence and gap are for the selected municipality at the scope of the border south west of Egypt (New Valley). The PEB estimates of poverty incidence are noticeably large from the third municipality in sample size to the last one (Matrouh, South Sinai, and North Sinai) respectively. The analogous estimates for the poverty gap are introduced. The different poverty intervals and colors are considered for each method because the ranges of EB and PEB estimates were quite different. The PEB figures indicate that the largest poverty incidence and gap are for the selected municipality at the scope of the border north east of Egypt (North Sinai).

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Ganesh N. Small area estimation and prediction problems: Spatial models, bayesian multiple comparisons, and robust MSE estimation. Ph.D. Dissertation, Department of Mathematics, University of Maryland. U.S.: College Park; 2007.

[2] Hidiroglou MA. Small-area estimation: Theory and practice. Joint Statistical Meetings Section on Survey Research Methods in Salt Lake City Utah, Innovation and Research Division, Statistics Canada; 2007.

[3] Ghosh M, Rao JNK. Small area estimation: An appraisal. Statistical Science. 1994;9(1):55-93.

[4] Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association. 1952;47(260):663-685.

[5] Mukhopadhyay PK, McDowell A. Small area estimation for survey data analysis using SAS software; 2011.
Available:http://support.sas.com/resources/papers/proceedings11/336-2011.pdf

[6] Rao JNK. Some new developments in small area estimation. Journal of the Iranian Statistical Society. 2003;2(2):145-169.

[7] Rahman A. A review of small area estimation problems and methodological developments. Online Discussion Paper Series -DP66, NATSEM. University of Canberra Australia; 2008.

[8] National Center for Health Statistics. Synthetic State Estimates of disability, Postal History of Society. Publications 1759, Government Printing Office, United Stats: Washington DC; 1968.

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

[10] Battese GE, Harter RM, Fuller WA. An error component model for prediction of county crop areas using survey and Satelite data. Journal of the American Statistical Association. 1988;83:28-36.

[11] Schaible WL. Indirect estimation in U.S. Federal Programs. New York: Springer-Verlag; 1996.

[12] Mukhopadhyay PK. Small area estimation in survey sampling. New Delhi: Narosa Publishing House; 1998.

[13] Rao JNK. Small area estimation. New Jersey: John Wiley & Sons Inc.; 2003.

[14] Longford NT. Missing data and small-area estimation. New York: Springer-Verlag; 2005.

[15] Chaudhuri A. Developing small domain statistics: Modelling in survey sampling, Saarbrücken: LAP LAMBERT Academic Publishing GMbH & Co. KG. Germany; 2012.

[16] Rao JNK, Molina I. Small area estimation, second edition, New Jersey: John Wiley & Sons Inc.; 2015.

[17] Fuller WA. Sampling statistics, New York: John Wiley & Sons Inc.; 2009.

[18] Chambers RL, Clark RG. An introduction to model-based survey sampling with applications, Oxford: Oxford University Press; 2012.

[19] Guadarrama M. Small area estimation methods under complex sampling designs. Ph.D., Departamento de Estadística, Madrid: Universidad Carlos III; 2017.

[20] Guadarrama M, Molina I, Rao JNK. Small area estimation of general parameters under complex sampling designs. Computational Statistics and Data Analysis. 2018;221:20-40.

[21] Elbers C, Lanjouw JO, Lanjouw P. Micro-level estimation of poverty and inequality. Econometrica. 2003;71:355-364.

[22] Molina I, Rao JNK. Small area estimation of poverty indicators. The Canadian Journal of Statistics. 2010;38:369-385.

[23] Alfons A, Templ M. Estimation of social exclusion indicators from complex surveys: The R package laeken. Journal of Statistical Software. 2013;54(15):1-25.

[24] Foster J, Greer J, Thorbecke E. A class of decomposable poverty measures. Econometrica. 1984; 52:761–766.

[25] Jiang J, Lahiri P. Mixed model prediction and small area estimation. Test. 2006;15:1–96.

[26] Pfeffermann D, Sverchkov M. Small-area estimation under informative probability sampling of areas and within the selected areas. Journal of the American Statistical Association. 2007;102:1427–1439.

[27] You Y, Rao JNK. A pseudo-empirical best linear unbiased predictor approach to small area estimation using survey weights. Canadian Journal of Statistics. 2002;30(3):431-439.

[28] González-Manteiga W, Lombardía MJ, Molina I, Morales D, Santamaría L. Bootstrap mean squared error of a small-area EBLUP. Journal of Statistical Computation and Simulation. 2008;78:443-462.

[29] Kreutzmann AK, Pannier S, Rojas-Perilla N, Schmid T, Templ M, Tzavidis N. Emdi: Estimating and Mapping Disaggregated Indicators. R Package Version 1.1.3; 2018.
Available:http://CRAN.R-project.org/package=emdi

[30] Diallo M. Small area estimation: Skew-normal distributions and time series. Unpublished Ph.D. Dissertation, Carleton University. Canada: Ottawa; 2014.

[31] Molina I, Marhuenda Y. Sae: An R package for small area estimation. The R Journal. 2015;**7**(1):81-98.
Available:http://journal.r-project.org/archive/2015-1/molina-marhuenda.pdf

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle3.com/review-history/49705