Benchmarking Provincial Poverty Incidences

Shirlee R. Ocampo
Vladimir II Christian R. Ocampo
De La Salle University
**shirlee.ocampo@dlsu.edu.ph*

Abstract: The first among the seventeen Sustainable Development Goals (SDGs) is to eradicate poverty. To achieve this goal, it is imperative to look at the estimation of poverty incidences in small areas as well the spatial distribution. One potential difficulty with small area estimates is that the combined estimate from all small areas does not usually match the value of the single estimate on the large area. Benchmarking is applied to modify these estimates to get the same aggregate estimate for the larger area. This is done by applying a constraint to ensure that the combined estimate of the small areas matches the larger area estimate. This research aims to incorporate benchmarking for provincial poverty incidence estimates under Bayesian beta-binomial hierarchical model using Philippine 2012 Family Income and Expenditure Survey (FIES) data. Estimates of provincial poverty incidences were generated from standard and benchmarked beta-binomial models. Twenty-three beta-binomial models were generated using Beta priors with varying hyperparameters along with different initializing values. Eleven of these models have incorporated constant and uniformly distributed benchmarking constraints. The eleven benchmarked beta-binomial models resulted to posterior estimates with slight differences depending on the benchmarking constraints but with much lower standard errors. The incorporation of prior information in both standard and benchmarked beta-binomial models showed an increase in the precision of the provincial poverty incidence estimates. There could be gains in precision when benchmarking constraint is incorporated in the beta-binomial model. Cluster analysis showed clusters of provinces with high poverty incidences in Eastern Visayas and Western Mindanao. It further indicated that cities in Metro Manila and its nearby provinces have low poverty incidences.

Key Words: benchmarking; beta-binomial; poverty incidence; cluster analysis; Bayesian hierarchical model

1. INTRODUCTION

Poverty can be described as not having adequate resources to satisfy the basic needs of a person or the inability of a household to meet the poverty line threshold (Albacea, 2009). Eradication of extreme poverty and hunger is the first of the Millennium Development Goals (MDG) established by the United Nations (UN), with specific target of halving the proportion of people who are in poverty and hunger between the years 1990 and 2015. In 2015, the United Nations (UN) General Assembly agreed on resolutions which include the Sustainable Development Goals (SDGs) which consist of seventeen set goals for the year 2030. The first of these SDGs "No poverty" aims to end extreme poverty in all forms globally by year 2030. In relation to these SDGs, the availability of the most possible accurate information concerning the living conditions of people at possible smallest domains is essential for targeting policies and programs aiming at the reduction of poverty. However, information collected from national surveys is limited and allows estimation only at larger domains or larger population subgroups.

The Philippine Statistical Authority (PSA) has for some years been producing estimates of the incidence of poverty at regional level. There has been however an increasing demand from policy makers and planners for a more disaggregated set of poverty statistics so that aid programs could be more effectively targeted to the areas in most need. In response to this, estimates of poverty at provincial level were released based on FIES. However, the standard errors of these estimates were sometimes quite large because of the small sample sizes at provincial level. Moreover, information collected from national surveys is limited and allows estimation only at larger domains like regional level. Hence, small area estimation techniques were utilized. One potential difficulty with small area estimates is that the combined estimate from all small areas does not usually match the value of the single estimate on the large area. The problem can be more severe in the event of model failure.

Benchmarking is applied to modify these model-based estimates to get the same aggregate estimate for the larger geographical area. This is done by applying a constraint, internally or externally, to ensure that the "total" of the small areas matches the "grand total". Internal benchmarking occurs when the pre-specified estimator can be a weighted average of the direct small area estimators. External benchmarking occurs when the pre-specified estimator is obtained from external sources, such as a different survey census, or other administrative records. Through benchmarking, the model-based estimates are modified to get the same aggregate estimate for the larger area. The combined small area estimates are forced to match the direct estimate of the large area obtained when the small areas are collapsed into a single area. This can be achieved either with respect to some weighted mean or with respect to both weighted mean and weighted variability (Nandram & Sanyit, 2011). Benchmarking can help to prevent model failure, an important issue in small area estimation. It provides some guard against model failure because the implementation of benchmarking corrects for some bias. It also shifts the small area estimators to accommodate the benchmark constraint. In doing so, it can provide some increase in the precision of the small area estimators of the finite population means or totals because the sample space is reduced by the constraint. An overall agreement with the direct estimates at an aggregate level with the small area estimates is essential in policy formulation and program implementation.

In the Philippines, numerous studies have been conducted on poverty statistics (Albert and Collado, 2004; Arcilla, Co and Ocampo, 2011). These studies had used nationwide surveys such as the Family Income and Expenditure Surveys (FIES), and the Annual Poverty Indicators Survey (APIS) as well as Community Based Monitoring System (CBMS) data, employing statistical techniques such as multiple linear regression and logistic regression analyses. However, there is no study yet that incorporated benchmarking constraints to model for both continuous and binary data. Many sample surveys require binary response (e.g. poor or nonpoor, hungry or not) from small areas. Direct estimators are not reliable because of the sparseness of the data, and hence, there is a need to use data from other areas to improve inference for a specific area. A Bayesian hierarchical model for this situation is the standard beta-binomial model. The beta-binomial model for small areas can be modified by incorporating prior information on the linear combination of the probabilities, called benchmarking constraint.

This research incorporated benchmarking for provincial level poverty incidences under Bayesian hierarchical beta-binomial models. The obtained provincial poverty level estimates were compared with those generated from standard beta-binomial models and small area estimates from PSA. Ranking and clustering of provinces were also performed.

2. METHODOLOGY

2.1 Data

The data used in this study is from Family Income and Expenditure Survey (FIES) 2012. For some models, FIES 2009 data was also used for initial estimates. This survey is conducted every three years and is used in generating official statistics which include poverty and hunger incidences. The target population of 2012 FIES included all households and members of households nationwide. A household is defined as an aggregate of persons, generally but not necessarily bound by ties of kinship, who live together under the same roof and eat together or share in common the household food (PSA, 2014). The sampling design used for 2012 FIES is the 2003 Master Sample (MS). Out of around 44,604 potential HHs interviewed, 90.1% of them responded to both FIES visits, and hence, the total number of responding households is 40,171 (Ericta & Collado, 2012). Variables included in this study are enumerated in Table 1.

Table 1. Variables and their Descriptions

Variable	Description
Region (regn)	A sub national administrative unit comprising of several provinces having homogenous characteristics
Province (prov)	States what specific province a particular household belongs in
Household Size (hh_size)	Total number of family member enumerated
Raising Factor (rfact)	Factor by which the number in the sample must be multiplied to give the total numbers in the population sampled
Poverty threshold (povth)	Represents the poverty threshold of the corresponding province
Total Family Income (toinc)	Includes primary income and receipts from other sources received by all family members
Per Capita Income (PCI)	Average or the income per person of the member of a household unit
Poor Household Classification (poor)	Classifies whether a household is poor or non-poor (0 = non-poor, 1 = poor)

A household (HH) is classified as poor if its per capita income is below poverty threshold. Poverty threshold is the minimum income required to meet the basic food and non-food needs such as clothing, fuel, light and water, housing, rental of occupied dwelling units, transportation and communication, health and education expenses, non-durable furnishing, household operations and personal care and effects (PSA, 2019). The poor are individuals and families whose income fall below the poverty threshold as defined by the NEDA and/or cannot afford in a sustained manner to provide their minimum basic needs of food, health, education, housing, and other essential amenities of life. Poverty incidence is the proportion of families/individuals with per capita income less than the per capita poverty threshold to the total number of families/individuals (PSA, 2019).

2.2 Methodology

2.2.1 Standard Beta-Binomial Model

Let n_i denote the total number of HH, s_i the number of poor HH , $f_i = n_i - s_i$ be the number of nonpoor HH, and p_i the proportion of poor HH in the i^{th} area, $i = 1, 2, 3, \ldots, \ell$. The unrestricted one-fold beta-binomial model is

$$s_i / p_i \sim_{iid} binomial\ (n_i, p_i)$$
$$p_i / \theta, \gamma \sim_{iid} Beta\left(\theta\left(\frac{1-\gamma}{\gamma}\right), (1-\theta)\left(\frac{1-\gamma}{\gamma}\right)\right).$$

The beta priors were given specified hyperparameter values for θ and γ or assigned uniform hyperprior distributions.

The posterior density of standard beta-binomial model is

$$g(p_i \mid s_i, \theta, \gamma) \propto \prod_{i=1}^{\ell} \frac{p_i^{s_i+\theta\left(\frac{1-\gamma}{\gamma}\right)-1}(1-p_i)^{f_i+(1-\theta)\left(\frac{1-\gamma}{\gamma}\right)-1}}{Beta\{\theta\left(\frac{1-\gamma}{\gamma}\right),(1-\theta)\left(\frac{1-\gamma}{\gamma}\right)\}} \quad \text{.(Eq.1)}$$

2.2.2 Benchmarked Beta-Binomial Model

The standard beta-binomial model can be extended to benchmarked beta-binomial model by the incorporation of benchmarking constraints. (Nadram and Sayit, 2011). The constraint incorporated is a weighted average of area probabilities given by $\phi = \sum_{i=1}^{\ell} \omega_i p_i - a$. If $\phi = 0$, this is equivalent to $\sum_{i=1}^{\ell} \omega_i p_i = a$. Without loss of generality, choose p_ℓ among $p_1, p_2, \ldots, p_\ell$. Solving for p_ℓ,

$$p_\ell = \frac{\phi + a - \sum_{i=1}^{\ell-1} \omega_i p_i}{\omega_\ell}.$$

Incorporating the constraint $\phi = 0$ in Eq.1, the joint posterior density of the benchmarked beta-binomial model is

$$g(\boldsymbol{p}_{(\ell)}, \theta, \gamma, a, \,/\, \boldsymbol{s}, \phi=0) \quad \text{(Eq. 2)}$$

$$\propto \prod_{i=1}^{\ell-1} \frac{p_i^{s_i+\theta\left(\frac{1-\gamma}{\gamma}\right)-1}(1-p_i)^{f_i+(1-\theta)\left(\frac{1-\gamma}{\gamma}\right)-1}}{Beta\{\theta\left(\frac{1-\gamma}{\gamma}\right),(1-\theta)\left(\frac{1-\gamma}{\gamma}\right)\}}$$

$$\cdot \frac{\left[\frac{a - \sum_{i=1}^{\ell-1} \omega_i\, p_i}{\omega_\ell}\right]^{s_\ell+\theta\left(\frac{1-\gamma}{\gamma}\right)-1} \left[1 - \frac{a - \sum_{i=1}^{\ell-1} \omega_i\, p_i}{\omega_\ell}\right]^{f_\ell+(1-\theta)\left(\frac{1-\gamma}{\gamma}\right)-1}}{Beta\{\theta\left(\frac{1-\gamma}{\gamma}\right),(1-\theta)\left(\frac{1-\gamma}{\gamma}\right)\}}$$

where $0 < p_i < 1$, $i = 1, 2, 3, \ldots \ell$, $0 < \theta < 1$, $0 < \gamma < 1$, $0 < p_\ell < 1$, and $0 < \phi + a - \sum_{i=1}^{\ell-1} \omega_i p_i < \omega_\ell$.

2.2.3 Estimation of Poverty Incidence

To estimate provincial poverty incidences using the standard and benchmarked beta-binomial models, Markov Chain Monte Carlo (MCMC) Gibb's sampling was applied using R integrated with Bayesian using Gibbs Sampling (BUGS). To achieve convergence, 10,000 iterations were performed. The beta distribution which is the conjugate prior to binomial distribution was used (Kurschke, 2011). Beta family of different specified hyperparameters such as Beta(1,1), Beta(2,5), and Beta(α, β) where the values of hyperparameters are random. In addition, hyperpriors of α and β are assigned to be uniformly distributed (Nandram & Sanyit, 2011). Numerous initializations were used such as all zeroes, all ones, random values, 2009 provincial poverty incidence estimates, and 2012 provincial poverty incidence estimates. The beta priors along with binomial likelihood resulted to beta-binomial posterior proportions which are the poverty incidence estimates. Posterior inference about the binomial probabilities includes posterior mean (PM), posterior standard deviation (PSD), and MCMC error.

The standard beta-binomial model is extended to a benchmarked beta-binomial model by incorporating a known or unknown benchmarking constraint which is a weighted average of the provincial probabilities as shown in Eq. 2. When assumed known, overall probability can be specified using prior information like a prior survey, census, larger domain estimates, or administrative records. In this research, constant and known benchmarking constraints were incorporated in the beta-binomial model. These fixed benchmarking constraints use information from nationwide 2012 poverty incidence (internal benchmarking) and prior survey 2009 FIES (external benchmarking). In addition, this research explored a benchmarking constraint which is unknown and has a uniform distribution under the Bayesian paradigm.

2.2.4 Measures of Complexity and Fit

The posterior mean deviance is suggested as a Bayesian measure of fit or adequacy. (Spiegelhalter et al, 2002) Deviance is computed by $D(\theta) = -2 \log p(y|\theta)$ for a likelihood $p(y|\theta)$. The posterior mean deviance $Dev = E[D(\theta)]$ is suggested as a measure of fit since it is robust and converges well. More complex models will fit the data well, and so, they will have smaller deviance. If a measure of complexity $P_d = E\theta|y[D] - D(E\theta|y[\theta]) = E\theta|y[-2 \log p(y|\theta)] + 2 \log p(y|\theta\tilde{}(y))$ is added to the measure of fit, *Dev*, then Deviance Information Criterion (*DIC*) results. Thus, *DIC* = 'goodness of fit' + 'complexity. This is analogous to Akaike Information Criterion (*AIC*) wherein models with smaller *DIC* are better supported by the data.

Nandram and Sanyit (2011) suggested Bayesian Root Mean Square Error (*RMSE*) which can be computed by

$$RMSE = \sqrt{(\hat{\pi} - PM)^2 + PSD^2}$$

where $\hat{\pi}$ is the direct estimate, *PM* is the posterior mean and *PSD* is the posterior standard deviation.

2.2.5 Cluster Analysis

Hierarchical and K-means clustering of provinces based on poverty incidences were performed using Statistica. Computations in this study were done using different software such as R integrated with BUGS, SAS, Statistica, and Microsoft Excel.

3. RESULTS AND DISCUSSION

Twelve standard beta-binomial models were simulated. Ten combinations of beta priors with varying hyperparameters and different sets on initial values generated provincial poverty incidence estimates which are almost the same but with very low standard and MCMC errors. The posterior estimates of the provincial poverty incidences of the standard beta-binomial models are robust to varying hyperparameters of the beta priors and different initial values. Two standard beta-binomial models were simulated where α and β have uniform hyperpriors. The obtained provincial poverty incidence estimates are almost the same with those generated from the ten models with very small standard and MCMC errors.

Furthermore, these 12 models resulted to the same ranking of provinces based on the posterior estimates of provincial poverty incidences. The three poorest provinces based on Bayesian beta-binomial poverty incidence estimates in 2012 were Lanao del Sur (66.93%), Eastern Samar (55.43%), and Apayao (54.69%) as shown in Table 2.

Table 2: Ten Provinces with Standard Beta-Binomial Provincial Poverty Incidence Estimates

Province	Poverty Incidence (%)	Standard Errors (%)	95% Interval Estimate (%)	
Lanao del Sur	66.93	0.125	66.67	67.17
Eastern Samar	55.43	0.1634	55.11	55.75
Apayao	54.69	0.3089	54.07	55.29
Maguindanao	54.36	0.1302	54.10	54.60
Zamboanga del Norte	48.14	0.1065	47.93	48.36
Saranggani	47.66	0.1456	47.37	47.96
North Cotabato	45.05	0.0936	44.86	45.23
Northern Samar	44.36	0.1435	44.08	44.63
Negros Oriental	43.95	0.0890	43.78	44.11
Western Samar	43.49	0.1252	43.26	43.73

Two known constant benchmarking constraints $\sum_{i=1}^{\ell} \omega_i p_i = a$ were incorporated in the beta-binomial model. The 2009 national poverty incidence 20.5% which is an external benchmark and the 2012 national poverty incidence 19.7% which is an internal benchmark were used along with different beta priors and different initialization values. Eight benchmarked beta-binomial models with constant benchmarking constraints were MCMC simulated. The choice of p_l was shown to be arbitrary, that is, p_l can be any data point. Under the benchmarked beta-binomial models with the same prior distribution and constant benchmarking constraint even with different sets of initial values, different choices of p_l resulted to the same poverty incidence estimates and very close standard errors. However, different constant benchmarking constraints produce different posterior provincial poverty incidence estimates. The obtained poverty incidence estimates using a benchmarking constraint of 20.5% are different from those obtained using a benchmarking constraint of 19.7%. The preferable benchmarking constraint is the internal 2012 FIES national poverty index of 19.7%. This matched the direct estimate 19.7% of a larger area (national level) when the small areas (provincial level) poverty incidence estimates are combined into one or single value. This overall agreement of a large area (national level) direct estimate at an aggregate level with provincial poverty incidence estimates is important in policy making and program implementation.

Three additional benchmarked beta-binomial models were explored to study the gain in precision. The study further explored the imposition of a proper but noninformative prior, and hence, the benchmarking constraint $\sum_{i=1}^{\ell} \omega_i p_i = a$ is imposed with a uniform distribution, that is, $a \sim Uniform\,(\sum_{i=1}^{\ell} \omega_i p_i\,, w_l + \sum_{i=1}^{\ell} \omega_i p_i)$. The generated poverty incidence estimates under these benchmarked models with uniformly distributed benchmarking constraints are the same with those obtained under the internally benchmarked beta-binomial model, but with very small standard errors. These benchmarked models resulted to the same ranking of provinces based on poverty incidence as shown in Table 3 with Lanao del Sur as the province having the highest poverty incidence estimate followed by Eastern Samar. It can be observed that most of these provinces are in Western Mindanao and Eastern Visayas.

Table 3: Ten Provinces with Highest Benchmarked Beta-binomial Poverty Incidence Estimates

Province	Poverty Incidence Estimates (%) Using 2012 Benchmarking Constraint	Poverty Incidence Estimates (%) Using 2009 Benchmarking Constraint	Poverty Incidence Estimates (%) Using Uniformly Distributed Benchmarking Constraint
Lanao del Sur	66.94	68.17	66.94
Eastern Samar	55.43	56.82	55.42
Apayao	54.69	56.07	54.68
Maguindanao	54.36	55.75	54.36
Zamboanga del Norte	48.14	49.55	48.13
Saranggani	47.67	49.07	47.66
North Cotabato	45.05	46.46	45.05
Northern Samar	44.36	45.77	44.36
Negros Oriental	43.95	45.34	43.95
Western Samar	43.5	44.89	43.49

Table 4 compares the beta-binomial models in terms of DIC and RMSE. The benchmarked beta-binomial models using the internal benchmark 19.7% which is the 2012 national poverty incidence and uniformly distributed benchmarking yielded the lowest DIC and RMSE and are the most preferred models.

Table 4: Measures of Fit and Complexity

Bayesian Hierarchical Models		DIC	RMSE
Standard beta-binomial (beta priors)	Any initial values	1183.48	0.041
Standard beta-binomial	2012 initial values	1183.32	0.041

(Uniform Hyperpriors)	2009 initial values	1183.07	0.041
Internal 2012 Benchmark	Beta (2,5)	1182.63	0.041
	Beta (1,1)	1181.08	0.041
External 2009 Benchmark	Beta (2,5)	10984.7	0.108
	Beta (1,1)	10987.9	0.108
Uniform distribution Benchmark	2012 initial values	1183.79	0.041
	2009 initial values	1182.57	0.041

4. CONCLUSIONS

One of the objectives of implementing a benchmarking constraint is the overall agreement of the national level poverty index direct estimate at an aggregate level with the provincial poverty incidence estimates. Hence, the internal and the uniform distribution benchmarking constraints optimize this aggregation. All 23 Bayesian hierarchical standard and benchmarked beta-binomial models (12 under standard beta-binomial models and 11 under benchmarked beta-binomial models) resulted to the same ranking of the 85 provinces based on the provincial poverty incidence estimates. This ranking is almost the same as the ranking of PSA but with small standard and MCMC errors.

The advantages of using standard and benchmarked beta-binomial models over the direct method in estimating the proportions are:

(1) the standard and MCMC errors are much lower compared to those in direct estimation;

(2) the estimates were derived from a posterior distribution which are derived from a likelihood function based on the sample data at hand and prior distribution based on prior information;

(3) the estimates are more precise compared to the direct method when the sample sizes are small, and the spatial units are small;

(4) the incorporation of prior information in both standard and benchmarked beta-binomial models showed an increase in the precision of the provincial poverty incidence estimates; and

(5) there could be gains in precision when benchmarking constraint which includes extra information is incorporated into the Bayesian hierarchical beta-binomial model.

Cluster analysis confirmed clusters of provinces with high poverty incidences in Eastern Visayas and Western Mindanao. It also showed that the National Capital Region cities and surrounding provinces have low poverty incidences.

5. ACKNOWLEDGMENTS

The researcher would like to acknowledge DLSU-University Research and Coordination Office (URCO) for funding this research, and DLSU Science Foundation for funding the research, DLSU-Angelo King Institute (AKI) Office for providing the data, and DLSU-Mathematics and Statistics Department for the support.

6. REFERENCES

Albacea, Z. V. J. Philippine Institute for Development Studies, (2009). Small area estimation of poverty statistics (No.2009-16).

Albert, J. & Collado, P. (2004). Profile and Determinants of Poverty in the Philippines. *9th National Convention on Statistics Proceedings.*

Arcilla, R., Co, F. & Ocampo, S. (2011). Correlates of Poverty. *DLSU Business and Economics Review* 20.2, 33-43.

Ericta, C.N., & Collado, P.M.G. (2012). *2012 Family Income and Expenditure Survey*. Philippines: National Statistics Office.

Garcia, H., Uy, M. & Ocampo, S. (2015). Exploring Hierarchical Bayesian Beta-Binomial and Dirichlet-Multinomial Models in Estimating Small Area Proportions of Poverty Incidence and Socio-economic Classes. *Proceedings of DLSU Research Congress.*

Kruschke, J. (2011). Doing Bayesian Data Analysis. UK: Elsevier

Nandram B. & Sayit H. (2011). "A Bayesian Analysis of Small Area Probabilities under a Constraint." *Survey Methodology*. 37, 137-152.

PSA (2014). 2012 *Family Income and Expenditure Survey Final Report*. Retrieved from https://psa.gov.ph/sites/default/files/2012%20FIES%20%28Final%20Report%29.pdf (June 20, 2020).

PSA (2019). Poverty threshold, Retrieved from https://psa.gov.ph/content/highlights-2012-full-year-official-poverty-statistics#:~:text=For%20the%20full%20year%202012,thresholds%20between%202009%20and%202012. (June 20, 2020)

Spiegelhalter, et. al. (2002). "Bayesian Measures of Model Complexity and Fit." *Journal of Royal Statistical Society*. 64, Part 4, pp. 583-639