International Journal of

Poverty, Investment and Development

(IJPID)

Disaggregating Poverty Estimates to Sub-County Level using Small Area Model: Application of EBP

CARI Journals

International Journal of Poverty, Investment and Development

2958-2458 (Online)

Vol. 6, Issue No. 1, pp 1 - 17, 2026 www.carijournals.org

Disaggregating Poverty Estimates to Sub-County Level using Small Area Model: Application of EBP

[1*]**Denis O. Onchomba**, [2]**Dr.Timothy K. K. Kamanu, PhD**

[1]Post Graduate Student

[2]Lecturer

[1,2]School of Mathematics, University of Nairobi

https://orcid.org/0009-0003-4314-2914

Accepted: 12th Jan, 2026, Received in Revised Form: 30th Jan, 2026, Published: 9th Feb, 2026

ABSTRACT

Purpose: The purpose of this study was to integrate Census data with sample survey data to disaggregate poverty estimates at the Sub County level. The aim was to produce estimates that provide more granular data, as current estimates are based on County level estimates.

Methodology: Small Area Estimation (SAE) was utilized by applying the Empirical Best Prediction (EBP) Model. This model assumes a unit-level approach, making use of data sets collected at the household (individual/unit) level. To achieve this, household-level secondary data from the Kenya Continuous Household Survey (KCHS, 2019) were combined with auxiliary covariates from the 2019 Kenya Population and Housing Census (KPHC). Both descriptive and inferential analyses were performed. The data was analyzed using Stata and R-software, and the results have been presented using tables, diagrams, and charts.

Findings: The results demonstrate similar poverty patterns as per county data findings on Kenya poverty report, 2022. The coefficients of variation (CV) for Small Area Estimation (SAE) estimates were consistently lower across most sub-counties. The bootstrap-based measures of uncertainty, including CV and mean squared error (MSE), confirmed that the Empirical Best Predictor (EBP) estimates were more precise.

Unique Contribution to Theory, Policy and Practice: The study recommended that National Statistical Organizations (NSO) implement SAE to routinely produce disaggregated poverty and development indicators at the lowest administrative levels, including Enumeration Areas. By adopting this methodology, the high costs associated with large-scale surveys currently required to achieve sufficient sample sizes for such estimates can reduce significantly. Furthermore, the study emphasizes that strengthening capacity in SAE through targeted training is essential to ensure methodological rigor, maintain data quality, and uphold high statistical standards.

Keywords: *Empirical Best Predictor (EBP), Empirical Best Linear Unbiased Predictor (EBLUP), Least Absolute Shrinkage and Selection Operator (LASSO)*

1.0 INTRODUCTION

1.1 Background

It is worth mentioning that SAE method is regarded as a game changer in the world of statistics. This method is essential because it allows us to address a significant challenge: breaking down indicators from the sub-national level to even lower levels. In the past, household surveys had no capability of producing statistically precise or representative estimates at these lower levels due to small sample sizes and the resulting significant variance. This new approach makes it possible to generate the detailed data needed where sample surveys have fallen short (Rao & Molina, 2021).

According to Rao & Molina (2021), one popular method for SAE estimation is known as Empirical Best Linear Unbiased Prediction (EBLUP). This method applies linear mixed models to merge information gathered both through surveys and other auxiliary sources. Such is done with the aim of borrowing strength across regions with the intention of improving precision. However, this method is vulnerable to assumptions concerning normality in random components and linearity. Such assumptions might not be held, especially when dealing with complex data.

Parker, P. A., Holan, S. H., & Janicki, R. (2022). Computationally efficient Bayesian methods for small area estimation with non-Gaussian outcomes. The Annals of Applied Statistics, 16(2), 887–904. This work utilizes Bayesian hierarchical methods to address the computational challenges associated with traditional Markov Chain Monte Carlo, offering more efficient alternatives for small area estimation. These challenges have motivated the use of the Empirical Best Prediction (EBP) method, which extends the strengths of traditional EBLUP by accommodating transformations and non-linear link functions. This makes it especially suitable for outcomes such as poverty indicators expressed as proportions.

By efficiently combining survey and auxiliary data through unit-level models, EBP offers greater flexibility and often improved performance over linear estimators for binary or non-normal outcomes. Contemporary research, such as that by Guadarrama, Molina, and Rao (2021), demonstrates that time-stable EBPs under unit-level mixed models can effectively estimate small area proportions while balancing efficiency with flexibility. Gradually, the incorporation of SAE approaches in national statistical systems has been expanded, with global organizations such as the World Bank and UN, among others, advocating for such approaches in estimating and assessing social indicators such as poverty (Pratesi, 2016).

The field was significantly advanced by the innovative work of Elbers, Lanjouw, and Lanjouw (2003), who developed a method to merge census and survey information for detailed poverty estimates. Since then, the World Bank has codified more recent guidelines on the use of SAE for estimating poverty and producing poverty maps, consolidating evidence-based methods and practical tools for practitioners in diverse data contexts (Corral, Molina, Cojocaru, & Segovia, 2022).

Poverty has been the major issue and challenge in Kenya and of high interest with varying patterns exhibited across counties. Kenya National Bureau of Statistics has been carrying out KIHBS and KCHS surveys with target of producing poverty indicator at the level of counties. This has been due to these surveys falling short of enough information as per geographical areas that can help in determining or calculating the estimates.

1.2 Statement of the Problem

Data on income and wealth distribution is critical for analysts who would like to forecast growth, development, and inequality in society. However, the process for processing information on income and wealth distribution is constrained by current limitations on accessibility and quality. Household surveys are considered the main source for information on income and consumption. However, the samples used in these surveys are small and regarded as insufficient even for larger geographical areas. Population censuses, on the other hand, cover all areas in a country but do not collect information on individual income and consumption. Thus, it is essential to have models for processing information that can combine both sources for enhancing accuracy on estimates for both poverty and inequality for a local geographical area (United Nations Statistics Division, 2022) Despite policymakers' demands for information on levels of poverty, for example, on a sub-county and ward level, it would be inefficient and unreliable for comprehensive surveys to take place on a wide scale (Paul Corral *et al*, (2022).

As for Kenya, little has been done with regards to estimating levels of poverty until it reaches levels for counties. This can be improved by employing concepts such as Small Area Estimation (SAE), where ultimately the aim is to obtain information, which is gathered by both surveying and censuses, with a focus on enhancing levels of estimation with regards to poverty levels. The above would ultimately allow one to have a clear understanding with regards to levels of distribution with regards to this matter. All this notwithstanding, it is important to note that it has not been done on levels with regards to sub-county levels. Thus, this research will aim to improve this by highlighting how it can be done with regards to enhancing levels of estimation with regards to levels of poverty on levels related to sub-county.

2.0 LITERATURE REVIEW

2.1 Empirical Literature

The unit-level SAE model has been improved and generalized by subsequent work to deal with limitations inherent in ELL. Strategies have been proposed, which involve Empirical Best (EB) and Empirical Best Prediction (EBP) for nested error models. These models showed flexibility for unit levels in nonlinear indicators such as Headcount Poverty Rate. These strategies used bootstrap samples for Mean Squared Error (MSE) estimation. Their approaches have shown improved accuracy with small-area estimation compared to conventional estimators and those used in ELL (Isabella Molina, 2019).

Recent experiences with empirical implementation of Small Area Estimation (SAE) in developing countries, particularly through World Bank poverty mapping projects, have been found to be successful for producing small-area estimation for data on poverty. Best practices on this issue have been cited by Corral et al. (2022) and Corral (2022) to include harmonization for variables, transformation for variables related to welfare, and bootstrap mean square error estimation. Despite all these, it has been found that problems related to temporal incongruence between survey and census data for small area estimation, quality of auxiliary variables, and model misspecification can impact predictivity (Corral, Molina, Cojocaru & Segovia, 2022).

In Kenya, given that indirect estimation using KCHS 2022 may not produce estimations for sub-counties because of small samples, SAE is a robust technique for producing exact estimations (Corral et al., 2022). According to this theory, this proposed work is informed by the need to produce sub-county estimations for poverty by combining information on consumption through KCHS 2022 with auxiliary variables for KPHC 2019 because disaggregated and cost-effective statistics are required for designing programs for alleviating poverty (Corral et al., 2022). Kenya Poverty and Inequality Assessment (2018) by World Bank and Kenya Poverty Report (2022) by KNBS are both significant background texts for this research work. The levels of both national and county poverty have been discussed in these texts, which have been shown with a considerable amount of inequality and a low level of precision for smaller administrative levels. Kenya Continuous Household Survey (KCHS 2022) is regarded as the main source for consumption, which is conducted every quarter to identify levels for welfare in households. Contrary to this, Kenya Population and Housing Census (KPHC 2019) is considered a source with significant demographic, housing, and asset information. Combining both these sources for this subject matter in the context of EBP will allow this research work to deal with multiple levels of issues with data discussed in both these national texts.

2.2 Theoretical Review

Many studies have been conducted on various approaches used for levels estimation on issues relating to monetary approaches as well as multidimensional approaches. The contribution on this issue by Sen (1976) developed a comprehensive framework for levels estimation on this issue by focusing on levels on this issue’s incidence, depth, and severity. That contribution is regarded as the root for levels estimation by instruments used on a class developed by Foster, Greer, and Thorbecke (1984) on levels estimation on issues relating to this matter. At present, even more contemporary approaches for levels estimation on this matter, encapsulated by the Multidimensional Poverty Index (MPI) developed by Alkire and Santos (2010), have taken levels on this matter with non-income dimensions such as education, health, and overall living.

The estimation for Kenya's poverty is normally done by applying a consumption based monetary approach that approximates household welfare by comparing it with average cost for each adult equivalence for pre-specified amounts generalized for food, overall, and hardcore poverty lines. The monetary method used in estimating Kenya's poverty is done by surveys across Kenya such as KIHBS, KCHS, which provide estimates such as Headcount Ratio, Poverty Gap, and Severity Index. Recently, Kenya used a Multidimensional Poverty Index (MPI) with non-monetary dimensions such as education, healthcare, and standard of living, which is aligned with Sustainable Development Goals (SDG) and Kenya's vision 2030. This is necessary because there is a need for appropriate estimation for Kenya's rate for policy purposes. Finally, it can be noted that only small samples can be done for small areas by national surveys, which is unreliable for counties. Of late, this is supplemented by emerging interest in appropriate estimates for small Area Estimation (SAE), which is recommended by (KNBS, 2024; Alkire & Foster, 2010; Rao & Molina, 2011).

As discussed by Rao and Molina (2015), model performance testing typically utilizes the Coefficient of Variation (CV), Mean Squared Error (MSE), and confidence intervals. These metrics are used to determine the levels of precision and accuracy for poverty estimates at the sub-county level. This is particularly relevant for direct estimates from sources like the KCHS 2022, which are known to exhibit high levels of variation. Additionally, Brown et al. (2001) propose a Goodness-of-Fit test to compare the performance of SAE estimates against Direct Survey Estimates. This involves testing the Wald Statistic to determine if there is a significant fit between the two. In this context, the null hypothesis is generally considered not statistically significant if the model is performing adequately.

Contribution of Elbers et al. (2003), which utilizes Small Area Estimation (SAE) and GIS technology to produce disaggregated maps. These maps are essential for the appropriate allocation of resources to address poverty. Moreover, Anselin's (2003) work on spatial autocorrelation to account for clustering effects in bordering sub-counties. By using KPHC Enumeration Blocks and SAE estimates, this research aims to provide cost-effective, disaggregated statistics. This methodology aligns with the goals stressed by Corral et al. (2022) to improve poverty alleviation activities and sustainably meet the Sustainable Development Goals (SDGs).

2.2 Conceptual framework

The conceptual framework for this specific adopted here is based on pre-existing approaches. Indeed, this conceptual framework is largely based on a guide by Corral et al. (2022) meant for small area estimation and poverty mapping with emphasis on unit-Level Empirical Best Prediction (EBP) for the disaggregation of household data, and with great emphasis on sub-county levels for estimation, testing, and subsequent mapping. This conceptual framework for this research is largely adapted, inspired, and developed on EBP frame, as discussed by Guadarrama, Molina, Rao (2016, 2018).

Studies such as Guadarrama et al. (2016, 2018), which applied the Empirical Best Predictor (EBP) model to Spanish labor market variables, the research of Corral et al. (2021) and Masaki et al. (2022) demonstrates that this model is universally adaptable. By following the comprehensive and flexible template developed by Corral et al. (2022), this research will adapt the EBP model using Kenya's Population and Housing Census (KPHC 2019) and the Kenya Continuous Household Survey (KCHS 2022). This approach ensures consistency, replicability, and integrity in estimating poverty at the sub-county level.

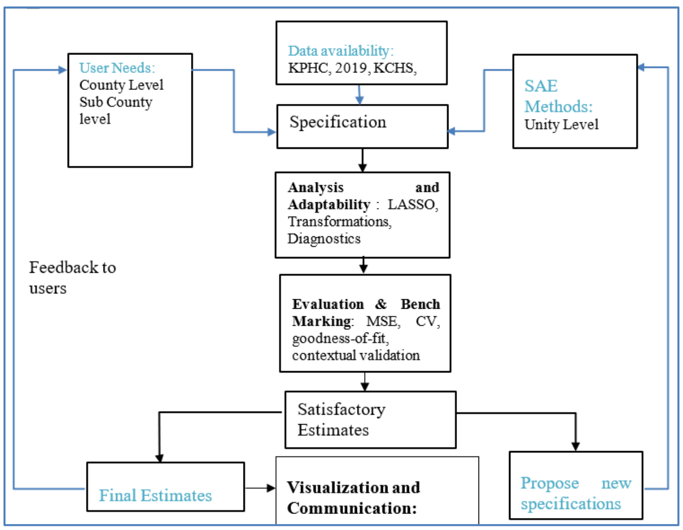


Figure 1: Framework to produce small area estimates

Source: Corral SAE Framework, 2022

3.0 METHODOLOGY

3.1 Survey Design

The study utilized a quantitative, cross-sectional ex-post facto research design, as it relied on existing secondary data from the 2019 Kenya Population and Housing Census and the 2022 Kenya Continuous Household Survey without the manipulation of any variables. The data were strictly quantitative, covering household- and individual-level socioeconomic and demographic variables. For the analysis, we employed model-based statistical methods, specifically the Empirical Best Prediction (EBP) approach under the Small Area Estimation (SAE) framework, to generate and validate sub-county-level poverty estimates.

3.2 Sampling and Sampling techniques

The study utilizes a 10% anonymized random sample from the 2019 Kenya Population and Housing Census (KPHC), covering approximately 4.7 million individuals and 1.23 million households. This sample was stratified by county and residence type into 92 strata, using Enumeration Areas (EAs) as primary sampling units selected via PPS sampling. Household weights have been calibrated using post-stratification adjustments to match known county-level totals. In addition, we are incorporating data from the 2022 Kenya Continuous Household Survey (KCHS). This survey followed a two-stage stratified cluster design based on the Kenya Household Master Sample Frame. Out of 24,000 targeted households, 17,850 completed the consumption module (an 85.5% response rate). Weights for this survey were adjusted for non-response and calibrated against the 2019 KPHC population totals. Table 3.2 for the detailed stratification and sample distribution across counties. This table provides the necessary context for understanding the calculation of sampling weights and sample allocation, which serve as the foundation for the implementation of the SAE model.

Table 1: Data sets Comparisons

Old Province	Population survey: KCHS 2022	census: KPHC 2019	survey/ census
National	**50,622,914**	**48,442,769**	**1.045**
Coast	4,591,475	4,388,794	1.046
North Eastern	2,652,291	2,536,169	1.046
Eastern	7,258,030	6,936,903	1.046
Central	5,846,635	5,595,501	1.045
Rift Valley	13,573,407	12,981,653	1.046
Western	5,368,910	5,141,037	1.044
Nyanza	6,685,711	6,399,001	1.045
Nairobi	4,646,456	4,463,711	1.041

Source: KPHC2019 and KCHS2022

3.4 Data Collection

Data collection in KCHS used Computer-Assisted Personal Interviewing. The consumption module looked at seven-day recall periods for food items and twelve months for non-food items. Specifically, we extracted a 10% anonymized sample from the KPHC 2019, which corresponds to about 1.2 million households, and the entire KCHS 2022 consumption module was harmonized for model estimation.

3.5 Data variable preparation and analysis

Data was cleaned and prepared data in two steps. Using Stata, to convert them to binary variables. Next, in R, we merged survey and census data using sub-county codes; we filled in missing values where possible and cleaned out inconsistencies in the data.

This study carefully compared questions in the survey and census to make sure they were measuring the same things across both sets of datasets. This yielded 529 variables, which were grouped into five main buckets entailing: household decision-maker characteristics (126 variables), household composition - 301 variables, housing conditions (76 variables), asset ownership (10 variables, agricultural activities(16 variables)

Covariate Selection was done using LASSO to select the most important predictors for this study. This penalized regression approach selects the most important covariates while shrinking less important coefficients toward zero, thus minimizing overfitting and handling multicollinearity:

$$\widehat{\beta^{(\lambda)}} = ar\,g\,m\,in_{\beta}\left\{\frac{1}{2M}\sum_{i=1}^{M}(y_i - x_i^{\top}\beta)^2 \;+\; \lambda\sum_{j=1}^{K}\left|\beta_j\right|\right\},$$

20 key variables were isolated from the 529 auxiliary predictors.

poor ~ weight + +W_night_lights +…………….H_lburni….to n variable where n=20

When working with headcount rates in our SAE framework, a common statistical challenge was identified. Since these rates fall between 0 and 1 in each area, they couldn't simply be plugged into a linear mixed-effects model - doing so would mess with the basic assumptions about normality and consistent variance. That informed us of the use of logit transformation instead.

$$\boldsymbol{yy_i = ln(P_i/1 - P_i)}$$

This transformation stretches across all real numbers, helping to keep the sampling variance steady and controlled. Afterwards the fitting of the nested-error regression model was carried out as shown

$$y_{ij} = x_{ij}\beta \;+ \mu_j + e_{ij}, \quad \mu_i j = N(0,,\sigma^2), e_i j = N(0\sigma_e^2),$$

Where:

$\boldsymbol{y_{ij}}$ is the log-transformed consumption of household in sub-county .
$\boldsymbol{x_{ij}}$ is the vector of KPHC 2019 auxiliary covariates.
$\boldsymbol{\beta}$ is the vector of fixed-effect coefficients.
$\boldsymbol{\mu_j}$ is the random effect for sub-county .
$\boldsymbol{e_{ij} \sim iidN(0, \sigma^2{}_e)}$ is the household-level error.
$\boldsymbol{n_j}$ is the sample size in sub-county , $\boldsymbol{i}$ and $\boldsymbol{j}$ is the total number of sub-counties

A nested error model proposed by Molina and Rao in 2010, later improved by Corral and colleagues in 2021, was adapted for this study. The logit method was chosen because it effectively stabilizes variance when working with proportions. After applying the model with restricted maximum likelihood, the predictions were converted from the logit scale back to regular proportions. This made the results more practical and easier to understand.

$$\widehat{p_i} \;=\; \frac{1}{1 + \exp(-\widehat{y_i})}$$

The accuracy of each estimate was evaluated using a parametric bootstrap to calculate the mean squared error (MSE) on the transformed scale. After converting the results back to the original

International Journal of Poverty, Investment and Development
2958-2458 (Online)
Vol. 6, Issue No. 1, pp 1 - 17, 2026 www.carijournals.org

scale using the delta method, reliable small-area poverty estimates, and their corresponding coefficients of variation were obtained.

The results were again checked using a Q-Q plot to compare the Pearson residuals to what we'd expect from a standard normal distribution (Figure 3.5.2.4). The points fell quite closely to the 45-degree reference line, only deviating at extreme ends. This visual check helped to confirm the transformed residuals were roughly normally distributed, a key requirement for valid EBP predictions and variance estimates.

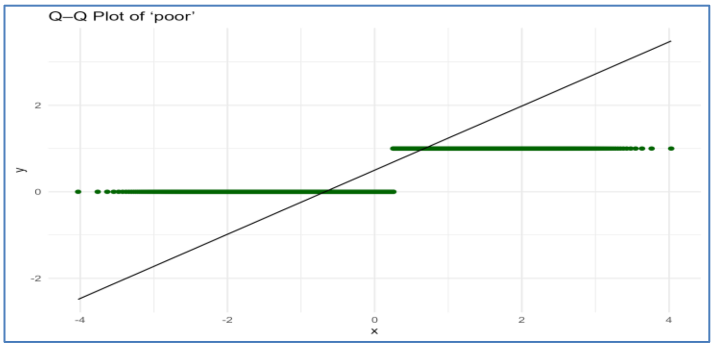


Figure 2: Q–Q Plot of Model Residuals for the Logit-LMM

Source: Author

The full model was built and tested in R's 'emdi' package, implementing REML to estimate variance. To get meaningful results, predictions were transformed back into household consumption figures that could assist in the calculation of the rates of poverty for every sub-county. Bootstrap method was used to estimate MSE, which helped to understand how accurate the predictions were. The lower the MSE, the better the model performed.

$$MSE(\widehat{\theta_j}) = \mathbb{E}\left[\left(\widehat{\theta_j} - \theta_j\right)^2\right]$$

The Coefficient of Variation (CV) showed how precise the measurements were by comparing the standard error to the actual estimate. In other words, it’s a way to measure how "scattered" the data is relative to the average. The smaller the CV, the more reliable the results.

$$CV(\widehat{\theta_j}) = \frac{\sqrt{\widehat{Var}(\widehat{\theta_j})}}{\widehat{\theta_j}} \times 100\%$$

4.0 FINDINGS

4.1 Produce sub-county poverty estimates and map to visualize spatial disparities

The results indicate that poverty rates vary significantly across Kenya. The highest levels were recorded in areas such as Turkana East (73%), Marsabit South (74%), and Kibish (84%). Conversely, rates in Westlands (12%), Embakasi (15%), and Meru South (16%) were comparatively lower. While these estimates are slightly higher, they align with the general trends exhibited in the Kenya Poverty Reports, 2022. Furthermore, most of these estimates proved to be highly reliable, with most areas showing standard errors of less than 5%

Table 2: Poverty Indicators by Sub-County

Subcounty Name	Poverty _SAE	SE	Subcounty Name	Poverty _SAE	SE	Subcounty _Name	Poverty _SAE	SE	Subcounty _Name	Poverty _SAE	SE	Subcounty _Name	Poverty _SAE	SE	Subcounty _Name	poverty_ SAE	SE
Changamwe	0.40	0.06	Sololo	0.73	0.03	Makindu	0.39	0.05	Turkana W.	0.94	0.03	Narok E.	0.31	0.04	Busia	0.48	0.05
Jomvu	0.34	0.05	Garbatulla	0.58	0.03	Makueni	0.62	0.04	Kipkomo	0.66	0.03	Narok N.	0.25	0.03	Butula	0.61	0.04
Kisauni	0.36	0.04	Isiolo	0.53	0.02	Mbooni E.	0.37	0.05	Pokot C.	0.76	0.04	Narok .	0.32	0.04	Nambale	0.45	0.04
Likoni	0.25	0.04	Marti	0.52	0.03	Mbooni W.	0.47	0.05	Pokot N.	0.48	0.04	Narok W.	0.26	0.04	Samia	0.56	0.04
Mvita	0.26	0.07	Buuri E.	0.45	0.06	Mukaa	0.46	0.04	Pokot S.	0.42	0.03	Trans Mara E.	0.40	0.05	Teso North	0.51	0.04
Nyali	0.25	0.05	Buuri W.	0.34	0.06	Nzaui	0.44	0.04	West P.	0.62	0.03	Trans Mara W.	0.35	0.04	Teso South	0.47	0.04
Kinango	0.60	0.04	Igembe C.	0.42	0.04	Kinangop	0.31	0.04	Samburu C.	0.52	0.03	Isinya	0.29	0.04	Siaya	0.42	0.03
Lunga Lunga	0.41	0.03	Igembe N.	0.49	0.03	Nyandarua S.	0.33	0.06	Samburu E.	0.57	0.04	Kajiado C.	0.59	0.05	Gem	0.35	0.04
Matuga	0.38	0.03	Igembe S.	0.75	0.06	Mirangine	0.33	0.05	Samburu N.	0.72	0.05	Kajiado N.	0.26	0.05	Ugenya	0.43	0.04
Msambweni	0.42	0.03	Imenti N.	0.32	0.05	Kipipiri	0.25	0.06	Trans Nzoia W.	0.31	0.04	Kajiado W.	0.48	0.03	Ugunja	0.31	0.05
Samburu	0.57	0.03	Imenti S.	0.29	0.04	Nyandarua C.	0.44	0.05	Trans Nzoia E.	0.27	0.05	Loitokitok	0.55	0.04	Bondo	0.31	0.04
Chonyi	0.61	0.04	Meru C.	0.29	0.05	Nyandarua W.	0.36	0.04	Kwanza	0.35	0.04	Mashuuru	0.36	0.05	Rarieda	0.34	0.04
Ganze	0.80	0.04	Tigania C.	0.41	0.05	Nyandarua N.	0.42	0.04	Endebess	0.65	0.05	Belgut	0.39	0.05	Kisumu East	0.34	0.04
Kaloleni	0.70	0.04	Tigania E.	0.63	0.07	Tetu	0.17	0.05	Kiminini	0.35	0.04	Burati	0.39	0.04	Kisumu Central	0.27	0.06
Kilifi N.	0.46	0.04	Tigania W.	0.43	0.05	Kieni E.	0.41	0.04	Ainabkoi	0.32	0.04	Kericho E.	0.35	0.04	Kisumu West	0.33	0.04
Kilifi S.	0.51	0.04	Igambang'ombe	0.48	0.04	Kieni W.	0.26	0.04	Kapseret	0.35	0.04	Kipkelion	0.48	0.05	Seme	0.62	0.04
Magarini	0.85	0.04	Maara	0.30	0.03	Mathira E.	0.33	0.05	Kesses	0.44	0.05	Londiani	0.47	0.05	Muhoroni	0.41	0.05
Malindi	0.53	0.03	Meru S.	0.16	0.04	Mathira W.	0.37	0.06	Moiben	0.24	0.05	Soin Sigowet	0.49	0.04	Nyando	0.47	0.05
Rabai	0.31	0.05	Tharaka N.	0.52	0.05	Nyeri South	0.27	0.05	Soy	0.42	0.04	Bomet East	0.45	0.04	Nyakach	0.53	0.05
Tana Delta	0.66	0.04	Tharaka S.	0.54	0.04	Mukurwe-Ini	0.37	0.04	Turbo	0.49	0.03	Chepalungu	0.49	0.04	Homa Bay	0.43	0.06
Tana N.	0.60	0.03	Embu E.	0.32	0.04	Nyeri C.	0.32	0.04	Keiyo N.	0.50	0.03	Konoin	0.41	0.04	Ndhiwa	0.33	0.04
Tana River	0.54	0.03	Embu N.	0.23	0.04	Kirinyaga C.	0.23	0.04	Keiyo S.	0.56	0.03	Sotik	0.44	0.03	Rachuonyo North	0.32	0.04
Lamu East	0.20	0.04	Embu W.	0.27	0.04	Kirinyaga E.	0.15	0.03	Marakwet E.	0.59	0.04	Bomet C.	0.53	0.04	Rachuonyo East	0.41	0.05
Lamu West	0.32	0.03	Mbeere S.	0.34	0.04	Kirinyaga W.	0.23	0.04	Marakwet W.	0.60	0.02	Butere	0.38	0.05	Rachuonyo South	0.29	0.05
Mwatate	0.35	0.04	Mbeere North	0.36	0.03	Mwea E.	0.40	0.03	Cheptumei	0.48	0.04	Kakamega C.	0.53	0.04	Rangwe	0.39	0.04
Taita	0.43	0.04	Ikutha	0.53	0.05	Mwea West	0.40	0.05	Nandi C.	0.42	0.04	Kakamega E.	0.37	0.04	Suba North	0.38	0.04
Taveta	0.34	0.04	Katulani	0.70	0.07	Murang'a East	0.43	0.05	Nandi East	0.35	0.05	Kakamega N.	0.36	0.04	Suba South	0.40	0.04
Voi	0.41	0.03	Kisasi	0.55	0.05	Kangema	0.32	0.07	Nandi North	0.32	0.04	Kakamega S.	0.28	0.05	Awendo	0.38	0.04
Balambala	0.94	0.03	Kitui Central	0.41	0.04	Mathioya	0.30	0.05	Nandi South	0.40	0.04	Khwisero	0.56	0.04	Kuria East	0.43	0.05
Dadaab	0.79	0.03	Kitui West	0.40	0.05	Kahuro	0.34	0.05	Tinderet	0.48	0.04	Likuyani	0.52	0.06	Kuria West	0.57	0.04
Garissa	0.63	0.03	Lower Yatta	0.29	0.06	Gatanga	0.26	0.04	Baringo North	0.35	0.04	Matete	0.51	0.07	Rongo	0.42	0.04
Ijara	0.85	0.03	Matinyani	0.39	0.07	Kigumo	0.32	0.05	East Pokot	0.52	0.03	Matungu	0.32	0.04	Suna East	0.51	0.04
Lagdera	0.47	0.04	Migwani	0.40	0.05	Kandara	0.38	0.04	Koibatek	0.34	0.04	Mumias East	0.57	0.07	Suna West	0.61	0.03
Buna	0.50	0.06	Mutitu	0.56	0.05	Gatundu North	0.27	0.06	Marigat	0.40	0.04	Mumias West	0.26	0.07	Uriri	0.38	0.05

Subcounty Name	Poverty _SAE	SE	Subcounty Name	Poverty _SAE	SE	Subcounty _Name	Poverty _SAE	SE	Subcounty _Name	Poverty _SAE	SE	Subcounty _Name	Poverty _SAE	SE	Subcounty _Name	poverty_ SAE	SE
Eldas	0.88	0.04	Mutitu North	0.43	0.07	Gatundu South	0.29	0.06	Mogotio	0.55	0.05	Navakholo	0.26	0.04	Etago	0.53	0.06
Habaswein	0.64	0.03	Mutomo	0.50	0.04	Githunguri	0.34	0.05	Tiaty East	0.47	0.05	Emuhaya	0.48	0.04	Gucha	0.29	0.04
Wajir East	0.86	0.03	Mwingi Central	0.44	0.04	Juja	0.27	0.06	Laikipia Central	0.27	0.05	Vihiga	0.51	0.04	Gucha South	0.28	0.05
Wajir North	0.46	0.07	Mwingi East	0.57	0.05	Kabete	0.36	0.07	Laikipia East	0.27	0.04	Sabatia	0.40	0.03	Kenyenya	0.41	0.05
Wajir South	0.49	0.04	Nzambani	0.69	0.06	Kiambaa	0.32	0.05	Laikipia North	0.64	0.05	Luanda	0.57	0.03	Kisii Central	0.35	0.05
Wajir West	0.69	0.03	Thagicu	0.54	0.06	Kiambu	0.28	0.07	Laikipia West	0.44	0.04	Hamisi	0.41	0.03	Kisii South	0.36	0.05
Mandera West	0.71	0.04	Athi River	0.31	0.05	Kikuyu	0.17	0.08	Nyahururu	0.31	0.03	Bumula	0.43	0.05	Kitutu Central	0.41	0.04
Banisa	0.68	0.04	Kalama	0.57	0.06	Lari	0.35	0.07	Gilgil	0.29	0.06	Bungoma Central	0.40	0.05	Marani	0.28	0.05
Lafey	0.74	0.06	Kangundo	0.41	0.05	Limuru	0.35	0.06	Kuresoi North	0.36	0.05	Bungoma East	0.71	0.05	Masaba South	0.48	0.06
Mandera Central	0.78	0.03	Kathiani	0.42	0.05	Ruiru	0.22	0.06	Kuresoi South	0.42	0.06	Bungoma North	0.32	0.05	Nyamache	0.40	0.05
Mandera East	0.71	0.02	Machakos	0.26	0.05	Thika East	0.38	0.09	Molo	0.46	0.06	Bungoma South	0.54	0.04	Sameta	0.29	0.07
Mandera North	0.72	0.04	Masinga	0.49	0.04	Thika West	0.23	0.07	Naivasha	0.33	0.05	Cheptais	0.50	0.04	Borabu	0.38	0.05
Loiyangalani	0.72	0.04	Matungulu	0.47	0.05	Kibish	0.84	0.03	Nakuru East	0.40	0.06	Kimilili	0.51	0.04	Manga	0.48	0.05
Marsabit Central	0.48	0.03	Mwala	0.49	0.04	Loima	0.73	0.03	Nakuru North	0.39	0.06	Mt Elgon	0.50	0.05	Masaba North	0.52	0.04
Marsabit North	0.69	0.03	Yatta	0.32	0.05	Turkana Central	0.78	0.03	Nakuru West	0.41	0.06	Bungoma West	0.58	0.06	Nyamira North	0.29	0.03
Marsabit South	0.74	0.03	Kathonzweni	0.40	0.05	Turkana East	0.73	0.03	Njoro	0.42	0.04	Tongaren	0.40	0.06	Nyamira South	0.41	0.03
Moyale	0.55	0.03	Kibwezi	0.40	0.03	Turkana North	0.83	0.04	Rongai	0.34	0.06	Webuye West	0.61	0.05	Dagoretti	0.29	0.06
North Horr	0.74	0.03	Kilungu	0.52	0.06	Turkana South	0.71	0.02	Subukia	0.38	0.08	Bunyala	0.49	0.04	Embakasi	0.15	0.05

Subcounty _Name	poverty_ SAE	SE
Kamukunji	0.27	0.07
Kasarani	0.34	0.06
Kibra	0.32	0.06
Lang'ata	0.30	0.10
Makadara	0.37	0.08
Mathare	0.25	0.11
Njiru	0.38	0.05
Starehe	0.17	0.07
Westlands	0.12	0.09

Source: Author

The poverty map created (Figure 3) clearly marks the dark red zones as high-poverty areas, while the lighter colors indicate lower poverty rates. What is interesting, however, is how starkly poverty can vary even within the same county—something that was not clearly visible in earlier analyses conducted at the county level. The use of EBP (SAE) helped balance out extreme values in the data, making this variation clear. This method also improved precision, mirroring the results seen in the county estimates. Similar improvements were observed in counties with smaller survey samples, such as Garissa, where data reliability improved by 3.75 times.

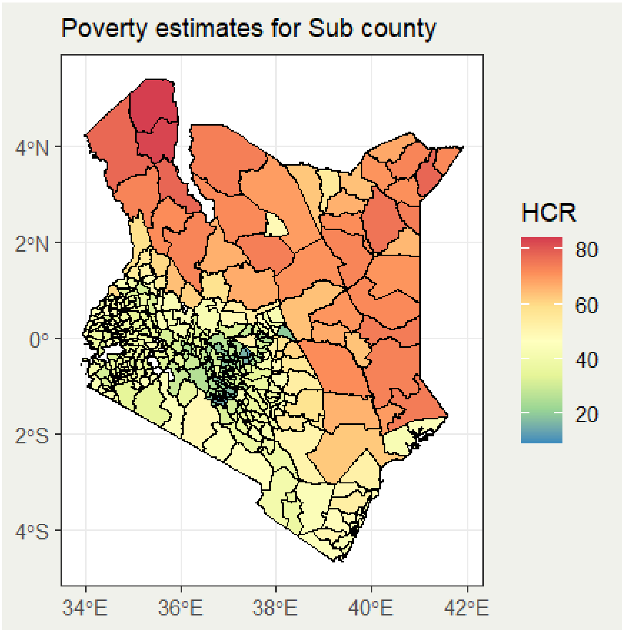


Figure 3: Sub-County Poverty Map
Source: Author

4.2 Assessing model diagnostics and precision gains through bootstrap Mean Squared Error (MSE).

4.2.1 Model Evaluation summary

Figure 4 shows the error term and random effects for the normality assumption. The random effects align well with this assumption, while the error terms exhibit slight deviations from normality, particularly in the tails, which should be considered during model diagnostics or inference. Figure 5 shows the density of Pearson residuals, which appear to be approximately normally distributed. This is a positive indication for model fit and supports the validity of methods that assume normally distributed residuals. The standardized random effects also appear to be approximately normally distributed, validating the assumption of normality for random effects, which is important for inference and prediction in mixed models or small area estimation (Figure 6).

International Journal of Poverty, Investment and Development

2958-2458 (Online)

Vol. 6, Issue No. 1, pp 1 - 17, 2026 www.carijournals.org

Figure 4 : Normal Q-Q Plot for Error Term and Random Effects

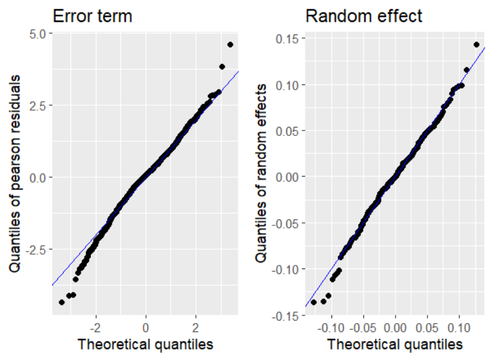


Figure 5: Density Pearson Residuals

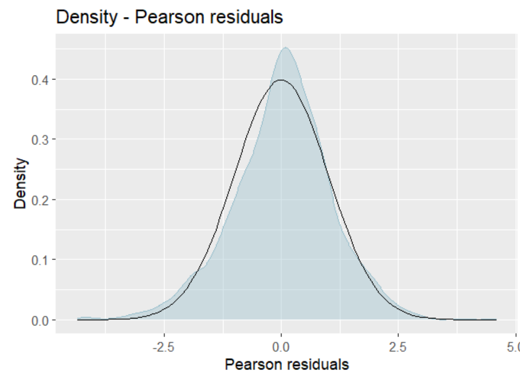
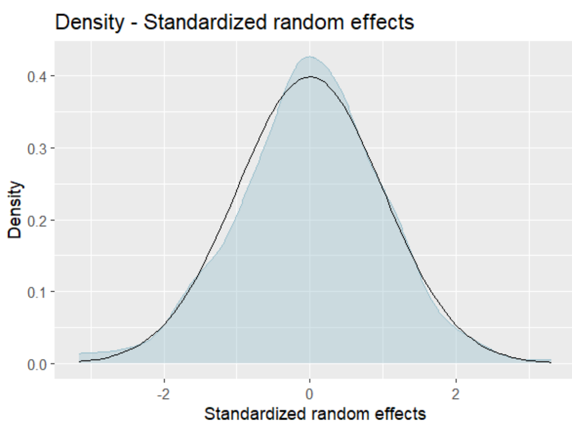


Figure 6: Density -Standardized Random Effects



Source: Author

Table 3 shows the model summary. The sample domains are much smaller in size compared to the population domains, which is typical in SAE where we have limited sample data but want to estimate for a larger population.

Arcsin transformation- suggests that the data was transformed using the arcsine transformation, often used for proportion or rate data to stabilize variance. Skewness near zero indicates roughly symmetric distributions. Kurtosis around 3.5 indicates slightly heavier tails than a normal distribution (which has kurtosis 3). For Shapiro-Wilk test, for Error, p-value is very small (4.19e-5), indicating Error residuals are not normally distributed. For Random_effect, p-value is 0.1843 > 0.05, so Random effects are approximately normally distributed. The variance due to error is much larger than the variance due to random effects, indicating that most variability is within the sampling error rather than between domains. Overall, the model explains a moderate to high proportion of variance, especially at the area level.

Table 3: Model Summary

row.names	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Sample_domains	1	3	4	4.144171779	5	19
Population_domains	1	17	24	25.98542274	32	73

Transformation	Shift_parameter
arcsin	0

row.names	Skewness	Kurtosis	Shapiro_W	Shapiro_p
Error	-0.125485445	3.52377406	0.994210197	4.18962E-05
Random_effect	-0.182855334	3.535232634	0.993615245	0.184326423

row.names	Variance
Error	0.051286217
Random_effect	0.006278105

Marginal_R2	Conditional_R2	Marginal_Area_R2	Conditional_Area_R2
0.373142791	0.441509365	0.579397205	0.774295293

row.names	gamma.Min.	gamma.1st.Qu.	gamma.Median	gamma.Mean	gamma.3rd.Qu.	gamma.Max.
PCCSS	0.109062434	0.229719671	0.293457179	0.294063669	0.363886148	0.63694047

Source: Author

4.3.2 Diagonistic, precison gains through bootstrap MSE

To assess the precision of these estimates, a bootstrap Mean Squared Error (MSE) approach was utilized, which involved multiple models runs and resampling of the survey data.

Key highlights from our analysis include:

Precision Gains: The results show a substantial improvement in precision across nearly all sub-counties. Most MSE ratio values were significantly less than one, indicating that the model-based estimates are far more stable than direct estimates. Sub-counties such as Njiru and Embakasi recorded the lowest MSE ratios (2.22×10^{-5} and 2.82×10^{-5}, respectively), resulting in precision gains exceeding 99%.

Robustness: Even in areas like Narok South and Habaswein, where MSE ratios were slightly higher, precision gains remained above 97%.

Efficiency: These findings confirm that the EBP model successfully leverages auxiliary information from census data to reduce variability, particularly in areas with small survey sample sizes.

International Journal of Poverty, Investment and Development

2958-2458 (Online)

Vol. 6, Issue No. 1, pp 1 - 17, 2026 www.carijournals.org

Ultimately, these results underscore that the adopted Small Area Estimation (SAE) approach enhances statistical reliability, making these estimates highly suitable for policy-relevant decision-making.

Table 4: Bootstrap MSE vs. Direct-Estimate Variance for the 10 Sub-counties with Greatest Precision Gain

Sub county_name	MSE_boot	SE_boot	Var_direct	SE_direct	MSE_ratio	Precision Gain
Njiru	4.83E-08	0.000219833	0.002179023	0.046680006	2.22E-05	0.995290643
Embakasi	1.14E-07	0.000338331	0.004061339	0.063728634	2.82E-05	0.994691067
Tana Delta	5.59E-07	0.000747435	0.015216016	0.123353216	3.67E-05	0.993940693
Butere	8.00E-07	0.000894495	0.012004929	0.109567008	6.66E-05	0.991836094
Aberdare National Park	1.83E-07	0.000427475	0.002617875	0.051165172	6.98E-05	0.991645201
Narok South	1.31E-05	0.003617349	0.038335262	0.195793928	0.000341336	0.981524715
Masinga	8.30E-07	0.00091111	0.002097946	0.045803343	0.000395683	0.980108218
Habaswein	3.56E-06	0.0018861	0.006464582	0.080402625	0.000550287	0.97654181
Balambala	8.13E-06	0.002851437	0.01160961	0.1077479	0.000700342	0.973536032
Nyeri South	6.16E-06	0.002481518	0.008139093	0.09021692	0.000756587	0.972493875

Source: author

Figure 7: Ratio of Bootstrap EBP MSE to Direct-Estimate Variance by Subcounty Interpretation

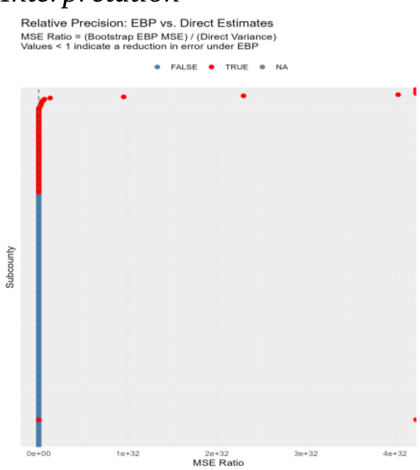
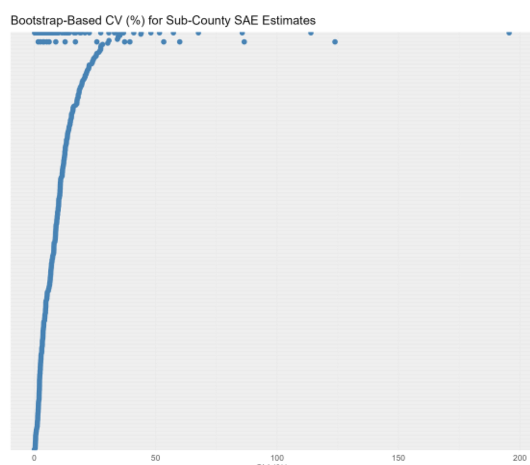


Figure 8: Bootstrap-Based Coefficients of Variation for Sub-County SAE Estimates



Source: Author

As shown in Figure 7, which compares the Bootstrap EBP MSE to Direct-Estimate Variance, most sub-counties are positioned to the left of the parity line. This indicates that the EBP model-based estimates generally offer higher precision and lower uncertainty than direct survey estimates, demonstrating the model's effectiveness in reducing sampling variability, particularly in areas with small sample sizes. Regarding Figure 8, the results further validate the model's precision. Most of the 346 sub-counties cluster with a Coefficient of Variation (CV) below 20-30%, signifying that the poverty estimates are reliably accurate. While a few outliers with higher CVs exist due to limited survey responses in specific areas, the SAE model overall provides credible data for most regions.

For practical application, estimates with a CV below 25% are considered reliable. For areas exceeding this threshold, it is recommended to combine them with neighbouring regions to achieve more credible figures, as suggested by Isabella Molina et al. (2022).

5.0 CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

The findings gave a clearly show that the small area estimation are reliable. The LASSO regression method was used to identify the most meaningful predictors from the 2019 Kenya Population and Housing Census. This helped in avoiding overcomplicating the model and gave a clearer insight into what drives poverty in different areas.

In sub counties estimates, the conclusion is clear that the estimates reflected the same pattern of poverty estimates as produced by Kenya poverty report, 2022. This is consistent with and clearly depicted in the map where the spread of the poverty estimates showed higher rates in counties like Turkana. . For policy purposes, sub-counties with error rates below 25% can be considered reliable, while those with higher error rates might need to be combined with neighboring areas for more stable results. The boostrap of MSE, CV is an enough indication to show that the model based EBP estimates have lower uncertainty to direct survey estimates and are reliable .

This study adds into contribution to previous studies such as Rao and Molina (2015 and corral *et al.*,(2022) whose conclusions realized the effectiveness of SAE. The findings having been also tested for Kenya situation, worked well hence great significance in internal understanding of the accuracy of measuring poverty as an indicator among other indicators.

5.2 Recommendations

The government National statistical agencies should establish small area statistical units to develop routine progress reports tracking poverty and making these indicators available up to the lowest administration levels such as Enumeration Areas. This will by far cut down the costs that would have been used in running comprehensive survey with a target of having sufficient sample for such indicators at lower administration levels.

Capacity building for statisticians in this area of SAE should be provided by providing appropriate foundations for such staff to venture in the area. This will act as a check of data quality and that the estimates produced are above standards and of quality.

For future research

EBP is not only the model that can produce SAE estimates, but researchers should also incorporate more ways such as area level models, Bayesian methods , geospatial big data models and machine learning to have even effective estimates. This is, however, in relation to the best model that fits the country's type of data available.

For effective planning the aspect of poverty maps for better visualization should be adopted and help administration regions such as sub counties in Kenya to direct resource where they are needed most. Small Area Estimation should not only be applied in producing poverty estimates, but just the indicator of interest. It should also cover a wide range of indicators such as health, job industry, school enrollment. The administrative data at county governments should also be stored in a way that can be useful in SAE in future.

6.0 REFERENCES

Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics, 95*(7–8), 476–487. https://doi.org/10.1016/j.jpubeco.2010.11.006

Brown, G., Chambers, R., Heady, P., & Heasman, D. (2001). Evaluation of small area estimation methods—An application to unemployment estimates from the UK Labour Force Survey. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 164*(1), 47–77. https://doi.org/10.1111/1467-985X.00193

Corral, P., Molina, I., Cojocaru, A., & Segovia, S. (2022). *Guidelines to small area estimation for poverty mapping*. Washington: World Bank.

Guadarrama, M., Molina, I., & Rao, J. N. K. (2016). Small area estimation of general parameters with application to poverty indicators: A hierarchical Bayes approach. *The Annals of Applied Statistics, 10*(2), 852–885. https://doi.org/10.1214/16-AOAS920

Pratesi, M. (Ed.). (2016). *Analysis of poverty data by small area estimation*. John Wiley & Sons.

Rao, J. N. K., & Molina, I. (2021). *Small area estimation* (3rd ed.). CRC Press. https://doi.org/10.1201/9780429061956.

World Bank. (2021). *Guidelines to small area estimation for poverty mapping*. World Bank. https://www.worldbank.org/en/events/2023/02/07/guidelines-to-small-area-estimation-for-poverty-mapping

KNBS (2020). Kenya Integrated Household Budget Survey 2015/16. Kenya National Bureau of Statistics.

Isabel Molina (2019) — *"Small area estimation methods for poverty mapping: a selective review"* (selective review / working paper / conference note, May 2019).

Rao, J. N. K., & Molina, I. (2015). *Small area estimation* (2nd ed.). Wiley

World Bank (2018). Kenya Poverty and Inequality Assessment. World Bank Group.

United Nations Statistics Division. (2022). *Handbook on Poverty Statistics: Concepts, Methods and Policy Use.*

Parker, P. A., Holan, S. H., & Janicki, R. (2022). Computationally efficient Bayesian methods for small area estimation with non-Gaussian outcomes. *The Annals of Applied Statistics, 16*(2), 887–904.