WORLD BANK GROUP

THE FEDERAL GOVERNMENT OF SOMALIA

DISTRICT-LEVEL POVERTY ESTIMATES IN SOMALIA:

A FAY HERRIOT MODEL APPROACH

ACKNOWLEDGEMENTS

This note is the product of a collaboration between the Poverty and Equity Global Practice of the World Bank and the Somali National Bureau of Statistics (SNBS). The technical assistance provided by the World Bank was led by Ifeanyi Edochie, with support from Alastair Haynes and Alejandro de la Fuente. Abdisalam Mohamed and Mohamed Abdinur led the SNBS team, who kindly facilitated the Somali Integrated Household Budget Survey 2022 for this work. The note benefited from the review and inputs of Paul Corral and Alexandru Cojocaru.

INTRODUCTION

Household surveys are often representative at the national level or at the level of the first administrative division (region/state level). National Statistical Offices and government entities can benefit from poverty estimation at a higher level of resolution, such as the district level. This note describes the small area estimation (SAE) methodology implemented to estimate poverty rates in Somalia at the district level. SAE is a statistical method that can be used to improve the reliability of survey estimates by combining survey data with geographically comprehensive auxiliary data, such as census when available or geospatial, remotely sensed data. In Somalia, we show that SAE generates poverty estimates that are sufficiently precise to report at the district level instead of the regional level. This has the potential to improve the targeting and evaluation of interventions intended to achieve poverty reduction in the future. Ideally, SAE combines survey data with household-level data from a recent census. Countries often aim to collect census data every 10 years. However, many African countries take more years between consecutive censuses, and indeed Somalia's last census was conducted in 1974. Therefore, this exercise relies on contemporaneous geospatial data derived from a variety of sources. Battese, Harter, and Fuller (1988) were the first to use geospatial satellite data in the context of crop production. Georganos et al. (2019) and Chi et al. (2022) have used satellite imagery to predict wealth indices. However, the use of geo-referenced data is less ideal than the traditional microdata obtained from household surveys or administrative datasets. Van Der Weide et al. (2024) used satellite imagery to predict monetary poverty in Malawi and noted the less-than-ideal nature of geospatial data. Corral, Henderson, and Segovia (2025) also find that remotely sensed data may not be ideal for poverty mapping due to the relatively lower predictive power. However, applying the area-based modelling approach to the use of geospatial data improves the precision of the poverty estimates significantly over district level direct estimates.

In this note, we present the approach that models poverty rates at the district level in Somalia using the model of Fay III and Herriot (1979). The area level model approach allows us to relate district level direct survey poverty rates to auxiliary variables (geospatial indicators) to estimate poverty rates in all districts within Somalia. Seitz (2019) provides district level poverty rates in the Central Asia region using the Fay-Herriot modelling approach and auxiliary geospatial data. The World Bank has employed the SAE methodology extensively to estimate poverty and other socioeconomic indicators of interest at more granular levels and continues to produce these estimates in combination with other non-monetary measures of poverty. At this point, SAE has been applied in a wide variety of contexts across many developing countries. This note is subdivided as follows. In section 2, we present survey data (specifically the household consumption data) and why SAE is necessary for district level poverty estimation in Somalia. We also present the Fay-Herriot (FH) model as described by Seitz (2019). Section 3 describes the geospatial databases sourced and indicators created as well as the model selection process employed. Sections 4 and 5 describes the FH model results and the poverty maps for the country.

This note describes the small area estimation (SAE) methodology implemented to estimate poverty rates in Somalia at the district level.

DATA

For Somalia, the 2022 Somalia Integrated Household Budget Survey (SIHBS) is representative at the regional level. The development of the SIHBS-22 sampling frame followed a stratified multi-stage probability cluster sample design. Urban and rural areas followed a three-stage stratified cluster sample design, while in nomadic areas the design was a two-stage stratified cluster sample design. The primary sampling units (PSUs) were selected with a probability proportional to the number of dwelling structures. The secondary sampling units (SSUs) for rural and urban areas were selected with a probability proportional to the number of listed households which constituted the frame. The ultimate sampling units (USUs) for rural, urban, and nomadic areas were randomly selected from listed households in the cluster.

District level poverty rate estimates computed from this survey will be insufficiently precise and unreliable for publication. Table 1 illustrates why it is necessary to use SAE to report poverty rates at more disaggregated geographic levels in Somalia. We use the mean coefficient of variation (CV) as a standardized measure of precision (i.e., the square root of the estimated mean square error divided by the poverty rate). Differing thresholds for mean or median CVs, often ranging from 0.1 to 0.3, have been applied by National Statistics Offices (NSO) to determine if statistics are sufficiently reliable to report. The median and mean direct CVs in Somalia at the district level are approximately 0.098 and 0.099.[1] While this is within the acceptable range of reliability for some countries, it is not considered reliable enough to be published by the Somalia NSO.

Table 1: Descriptive Statistics

Indicator	Estimate	Source
Population (in millions)	13.6	HBS 2022
Population Number of HHs (in millions)	2.0	HBS 2022
Sample # of HHs	6221	HBS 2022
Poverty Rate (NPL)	0.514[2]	HBS 2022
Latest Census Year	1974	N/A
Number of Regions	17	SNBS Official Somalia boundary file (shapefile)
Region Median CV	0.098	HBS 2022
Region Mean CV	0.097	HBS 2022
Number of Targets (Population)	74	SNBS Official Somalia boundary file (shapefile)
Number of Targets (Sample)	48	HBS 2022

1 The coefficient of variation is the MSE divided by the poverty rate. In estimating direct variances (directs MSEs), we adopt the sampling design of the SIHBS (described in Section 2). We use the svydesign function of the survey R package in computing the variances.

2 The poverty rate computed only uses the rural and urban areas and does not include households in refuge encampments or IDPs. This also explains the disparity between the poverty rate displayed and the official national poverty rate.

We utilize freely available geospatial data for this small area estimation exercise since the last census carried out is outdated (from 1974). The goal of the SAE exercise is to estimate more reliable district level poverty rates in Somalia by using a Fay-Herriot model based on relating the target area direct estimate poverty rates and district level geospatial indicators. Given that any recent developments in Somalia might not be captured by its 50-year-old census, it would be difficult to make a case for area poverty rates estimated using the 1974 census particularly to guide current policy interventions.

METHODOLOGY

Corral et al. (2022) recommend implementing an area level Fay-Herriot model with geospatial indicators for poverty mapping. Imagine a finite population for Somalia, P, that consists of N households that are subdivided into D districts with sizes $N_1,\ldots,N_D$. A random sample of households can be drawn from the d^{th} commune (i.e., $n_1,\ldots,n_d$ s.t. $n < N$. The Fay-Herriot (FH) model comprises of two levels. The first is a sample model which assumes a direct survey estimator:

$$\hat{\theta}_i^{Dir} = \theta_i + e_i, \qquad \forall_i = 1_i, \ldots, D$$

$\hat{\theta}_i^{Dir}$ is design unbiased for the small area parameter, θ_i the population indicator of interest, in this case, the poverty rate each district, d_i. We assume a sample error e_i is normally distributed with a mean of zero and a variance of $\sigma^2_{e_i}$.

$$e_i \sim N(0, \sigma^2_{e_i})$$

In the second level, a linking model is assumed to relate θ_i to auxiliary variables $x_i = (x_{i1},\ldots x_{ic})'$ via a linear regression. Both levels of the model together are presented as follows:

$$\hat{\theta}_i^{Dir} = x_i^T \beta + \mu_i + e_i; \quad \mu_i \sim N(0, \sigma^2_\mu); \quad \forall_i = 1_i, \ldots, D$$

The empirical best linear unbiased estimators (EBLUP) β are computed by weighted least squares regression. The EBLUP of θ_i is obtained by substituting the variance parameter σ^2_μ with an estimate. The resulting estimator can then be written as:

$$\hat{\theta}_i^{FH} = x_i^T \hat{\beta} + \mu_i$$

$$\hat{\theta}_i^{FH} = \hat{\gamma}_i \theta_i^T + (1 - \hat{\gamma}_i) x_i \hat{\beta}$$

The EBLUP/FH estimator can be understood as a weighted average of the direct estimator and a regression synthetic part. The estimated shrinkage factor $\hat{\gamma}_i = \dfrac{\sigma^2_\mu}{\sigma^2_\mu + \sigma^2_{\varepsilon_i}}$ puts more weight on the direct estimator when the sampling variance is small and vice versa. Areas for which no direct estimation results are called out-of-sample domains. For those domains the prediction reduces the regression-synthetic component $\hat{\theta}_{i,out}^{FH} = x_i^T \hat{\beta}$ (Molina and Rao 2010).

This method is widely used by the Small Area Income and Poverty Estimates (SAIPE) program of the US census bureau and has been thoroughly validated in Corral, Rodas, Henderson, and Segovia (2023). This approach improves the error efficiency rates over the direct estimates at the target area level. Inter-area unexplained heterogeneous area effects are accounted for within the model. Section 3.3 in Corral et al. (2022) provides a full list of advantages and disadvantages of the Fay-Herriot modelling approach.

For this small area estimation exercise, 48 of the 74 Somalia districts are included in the SIHBS. As a result, these in-sample districts will benefit from the information available in the survey. In some cases with FH models, districts with low sample sizes can result in all households from a specific sample district being all poor or not poor ($\theta_i^{Dir} = 1$ or 0) or only one Enumeration Area in a district is sampled. The common practice of sample variance smoothing (You and Hidiroglou 2012; You and Hidiroglou 2023) in the SAE literature is typically implemented to solve this problem. The variance smoothing approach of You and Hidiroglou (2012) applies a log-linear model of the direct sampling variance $\{\widehat{V}_i\}$ as a function of the sampling size, n_i.

$$log\ (\widehat{V}_i) = \widehat{\phi}_0 + \widehat{\phi}_1 n_i + \varepsilon_i,\ \ i = 1_i, ..., D$$

Assuming $\widehat{\phi}_0$ and $\widehat{\phi}_1$ to be the simple OLS estimators for the regression coefficients $\widehat{\phi}_0$ and $\widehat{\phi}_1$. Applying the exponential of the equation produces the naive variance estimator (Dick 1995) as follows:

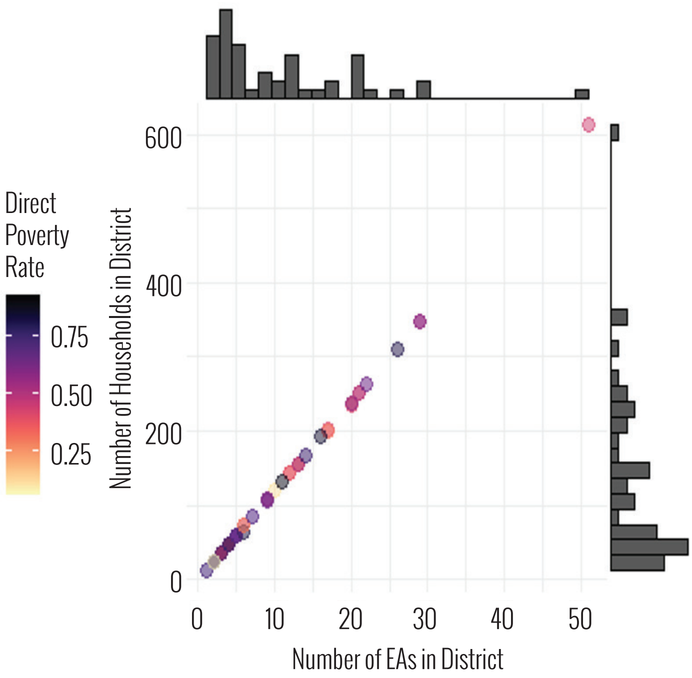
$$\widehat{V}_i = e^{\widehat{\phi}_0 + \widehat{\phi}_1 n_i}$$

Rivest and Belmonte (2000) show that the naive estimator can underestimate sampling variance. They propose correction as follows:

$$\widehat{V_i^{RB}} = \widehat{V}_i e^{\frac{\tau^2}{2}}$$

since the naive variance estimator can be easily shown to overestimate sampling variance by a factor of $e^{\frac{\tau^2}{2}}$. For the purposes of our analysis, there are no districts without variances or extreme case poverty rates ($\theta_i^{Dir} = 1$ or 0). Consequently, we do not remove districts from the analysis. However, the NSO flagged the initial predicted poverty rates in 3 districts as being too low. 2 of these 3 districts have low sample sizes and with 2 or less primary sampling units as well. In the supplementary section, we re-estimate the Fay-Herriot model without these areas and present the results. Below, we simply show the sample distribution of PSUs and households within districts and how this varies with the direct district poverty rates.

Figure 1: Joint Distribution of EAs and Households (District-Level)



GEOSPATIAL DATA AND MODEL SELECTION PROCESS

The process leading up to model selection involves sourcing freely available geospatial indicators that might be correlated with household welfare and poverty. The geospatial features were sourced at native resolution and then zonal statistics were computed at the target area level (districts). Table 2 shows all the geospatial features and the data sources employed.

Table 2: EBP Model (Regression Results)

Feature	Estimate	Original_Data_Resolution	Year
Built-settlement extent area	WorldPop Building Footprints	1km	2001 - 2020
Gridded Population & Density	WorldPop Gridded Population Counts & Density	90m	2020
Share of area planted by crop for banana, beans, cassava, maize, sesame seed, sorghum, sugar cane, temperature fruit, tropical fruit, vegetables	IFPRI Spatial Production Allocation Model (SPAM)	10km	2009, 2017, 2020
Production quantity for each crop for banana, beans, cassava, maize, sesame seed, sorghum, sugar cane, temperature fruit, tropical fruit, vegetables	IFPRI Spatial Production Allocation Model (SPAM)	10km	2009, 2017, 2020
% production as a total crop production for banana, beans, cassava, maize, sesame seed, sorghum, sugar cane, temperature fruit, tropical fruit, vegetables	IFPRI Spatial Production Allocation Model (SPAM)	10km	2009, 2017, 2020
Standardized precipitation evaporation index, 12 month	Global SPEI database, version 4.03	0.5 degrees	2020
Drought exposure, Drought hazard, Drought risk index, Drought vulnerability	(Carrao et al. 2018)	0.5 degrees	2000-2014
Drought hazard, risk for irrigated agricultural systems	Drought risk for rainfed, irrigated agric. systems aggregated as an average per polygon based on the data from (Meza et al. 2020)		2020
Percent of area with Vegetation Index below 40 for the Gu season (April - June)	STAR - Global Vegetation Health Products		2017-2022
Average travel time in nearest urban areas with a population of 5000, 20000 and 50000.	Computed based on population data from WorldPop and accessibility data from (Nelson et al. 2019)		2019

We begin by transforming all indicators as necessary to minimize the risk of divergence in model parameter estimation. For indicators with values greater than 100, we take the natural logarithm. We have avoided feature scaling to avoid excessive distortion or loss of information for the scaled variables.

The geospatial data listed under the previous header was used to construct candidate features, at the grid and target area level. In addition, we include regional dummy variables. In all, we created 157 potential geospatial candidate indicators. Using all these features in the linear mixed model risks potentially leads to over-fitting the survey sample and generates poor out-of-sample estimations.

Next, we employ a stepwise (both-ways method) selection approach which picks the most predictive set of indicators from the pool of candidate indicators. The both-ways method was used, enabling iterative testing of each variable's contribution by alternatively adding and removing variables based on statistical significance criteria at each step. This approach begins with a constant term and tests the inclusion of variables one-by-one; then it considers each for potential removal, thus optimizing the model's explanatory power while controlling for over-fitting. The both-ways method provides flexibility, more so than the forward or backward algorithm, to achieve an optimal balance of predictive power and model parsimony, ensuring that only variables with significant and robust relationships to the outcome are retained.

FAY-HERRIOT MODEL ESTIMATION RESULTS

The final selected set of variables suggests that spatial measures for urbanization, climatic factors (such as drought), agricultural productivity proxies and the regional dummies are the most predictive indicators from the pool of candidate predictors. The regression results are as follows:

Table 3: EBP Model (Regression Results)

Variables	Coefficients	Standard Error
Intercept	0.151	0.134
Rural Reachability Index	0.003**	0.001
Drought Hazard (Carrao et al. 2016 estimates)	0.206	0.124
% of area with vegetation health index (VHI) below 40 during the 2020 season	0.005**	0.002
Population density per sq. km of populated area	0.0000019	0.0000053
Harvest area for maize as a share of all crops	0.014***	0.004
Gedo Region	-0.23**	0.072
Production quantity for maize as a share of all crops	-0.0078***	0.002
Harvest area for vegetables as a share of all crops	-0.031***	0.009
Nugaal Region	-0.08	0.082
Hiraan Region	0.296**	0.106
Bakool Region	0.271**	0.089
Sample Size (n) = 48		

Statistical significance for each coefficient value, *** for 1%, ** for 5%

Table 4: Assessing Normality Assumptions

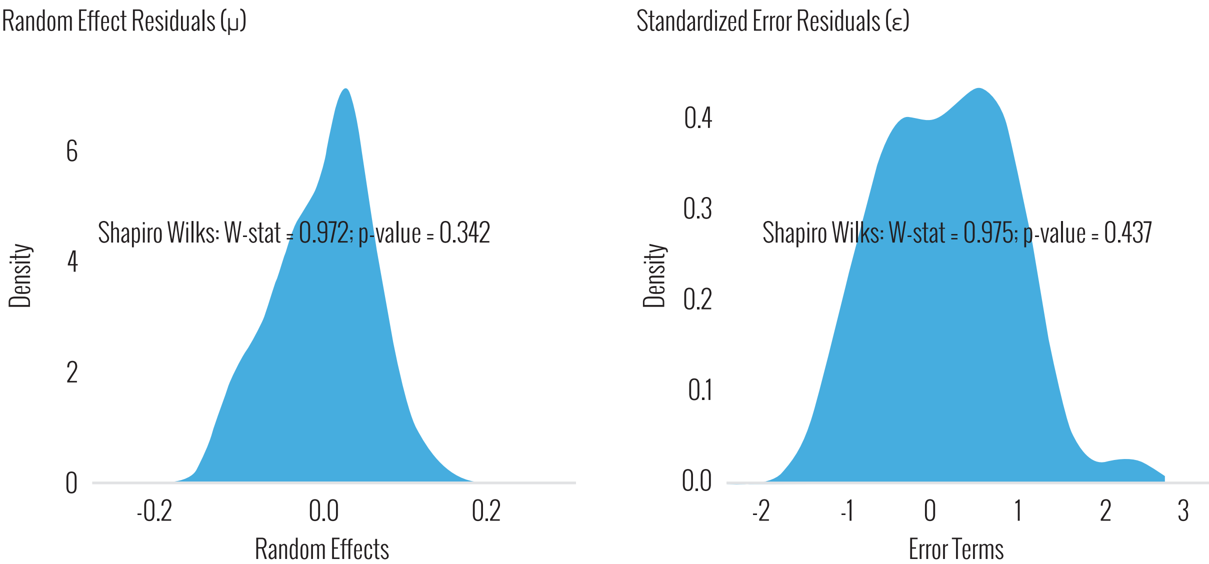
Model R^2		(Error Term) ε		(Random Effect) μ	
marginal	conditional	skewness	kurtosis	skewness	kurtosis
0.433	0.577	0.653	3.002	-0.215	3.054

The regression coefficients have the expected sign. The Hiraan and Bakool regions appear to be poorer on average than other regions. In contrast, the Nugaal region appears to be better-off than the average region, although this is not statistically significant. Several agriculture-related variables appear to be significant predictors of poverty rates in Somalia. The production quantity for maize is associated with lower poverty rates while larger harvest areas appear to be increasing in poverty rates. Maize cultivation, and most crop production in Somalia, is heavily rural, which might explain the direction of the sign for maize harvest area shares with respect to poverty. The proportion of an area with a vegetation health index (a geospatial proxy for crop production) below 40 (in 2020) is increasing in poverty rates. Consequently, less green areas are more likely to be worse off. In addition, drought hazards also appear to be unsurprisingly increasing in poverty rates, although this result is not statistically significant.

Several assumptions are made in this model which needed to be verified. The Fay-Herriot model R^2 equals 57.7% with an adjusted R^2 of 43.3% which is typical for the FH model particularly with only 48 in-sample districts used in the regression of only geospatial features. We assume independent normal distributions for the area effects as well as error terms. The table shows the skewness and kurtosis which should be approximately 0 and 3 for normally distributed random variables.

The normality assumptions proposed in the method section matter for the noise estimates but the EB methodology ensures that the poverty estimates are unbiased. The residual analysis suggests that the skewness and kurtosis of the idiosyncratic and district level area effects match the normality assumptions. However, there appears to be few outliers within the error term normal density plot. The residual plots for both the random error and idiosyncratic errors can be found below:

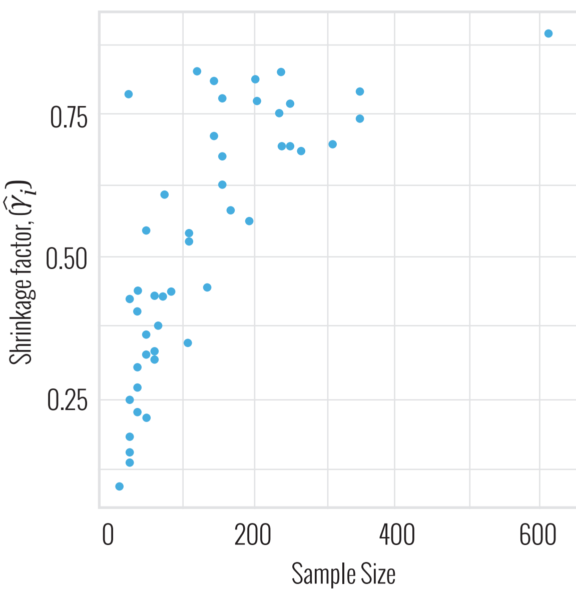
Figure 2: Fay-Herriot Residual Plots



We employ the Shapiro-Wilks measure to test the null hypothesis that random variables, e_i and μ_i, come from a normality distributed populations. The test results, W_{e_i} = 0.963 (p= 0.129); W_{c_i} = 0.989 (p= 0.916) suggest normally distributed random effects and idiosyncratic error terms. We cannot reject the null hypothesis at the 5% level (although the standardized error residuals are significant at 10%).

The Fay-Herriot model employs direct estimates in predicting poverty rates. The shrinkage factor measures the ratio of the random effect to the total variance within the model. Full shrinkage $\widehat{V}_i = 1$ means predicted poverty rates are simply the direct estimates while the other extreme uses a purely synthetic predictions, $\widehat{V}_i = 0$. We present a scatter plot of the $\widehat{V}_i$ as a function of sample size.

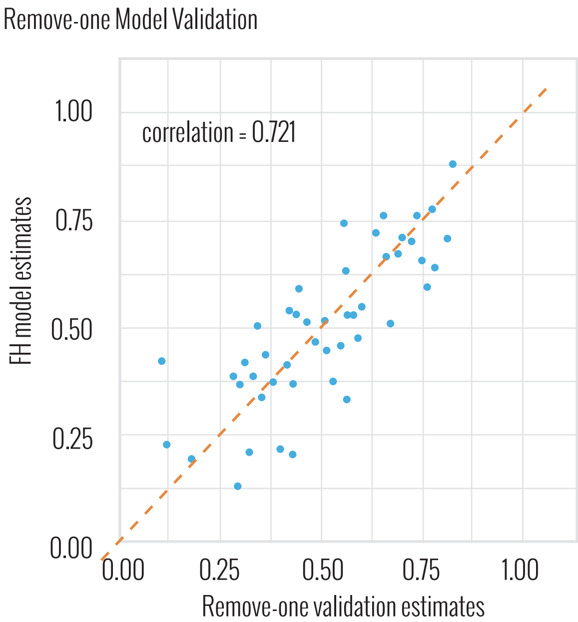
Figure 3: Scatter Plot of Shrinkage Factor as a function of sample size



In a final check, we attempt to validate the model by performing the Remove-One Model validation. Since our sample only contains 48 target areas, the typical n-fold validation process would have to split an already limited sample into 2 smaller training and test sets. Instead, the Remove-One validation process, trains a Fay-Herriot model on 47 districts and removes 1 until every district has been excluded once. We compare show a plot comparing model validation estimates with the actual FH model predictions to check the stability of the model.

Figure 4 plots the FH model estimates against their corresponding estimates as a result of the remove-one validation model. The correlation between the set of FH model estimates and the validation estimates stands at 0.75 as shown in above chart.

Figure 4: Remove-One Model Validation Plot

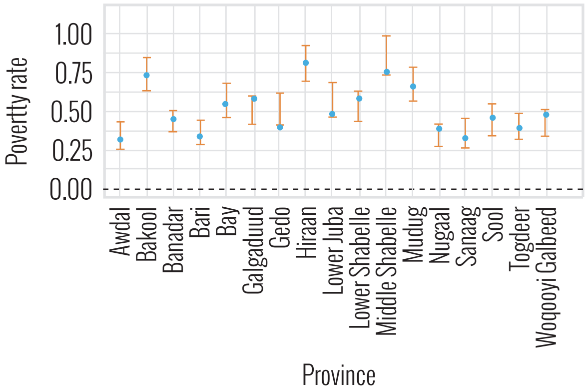


POVERTY MAPS

As a final check, the FH poverty rates at the district level are aggregated to the regional level to compare against the direct estimates. The regional level is the highest level of resolution at which the survey design reaches representativeness. The direct estimates in Figure 5 are shown as 95% confidence intervals (in red) which are plotted in comparison with Fay Herriot poverty estimates.

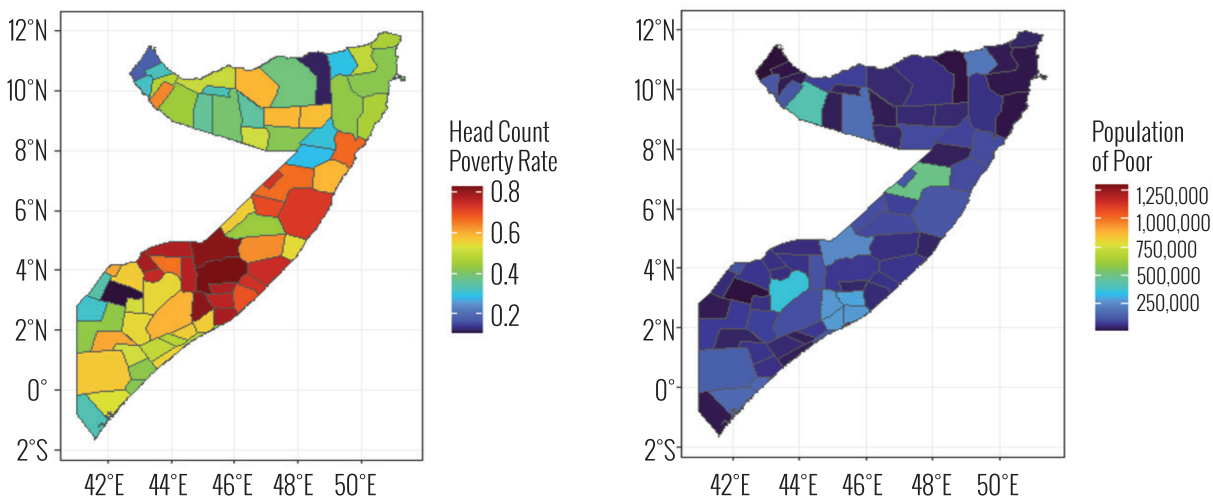
It should be noted that all the model based regional estimates fall within the direct estimate confidence intervals for all regions. Figure 6 presents the initial district level poverty estimates.

Figure 5: Comparison of FH and SIHBS Estimates



Note: Author calculations of Direct Estimate 95% CI at the province level (Red error bars) compared with FH estimates at same level (blue dots)

Figure 6: Initial Predicted District Poverty Rates and Number of Poor



RE-ESTIMATING FH MODEL WITHOUT GARBAHAAREY, LASSQORAY AND ZEYLAC IN SIHBS SAMPLE

However, upon review of the initial estimates in collaboration with the SNBS, 3 districts were identified as having unrealistically low poverty estimates as the surrounding districts within each region had significantly higher poverty rates. These districts were Garbahaarey, Lassqoray and Zeylac in Gedo, Sanaag and Awdal regions, respectively. This may have been driven in part by the limited number of enumeration areas in these districts (see Table A3). Three benchmarking approaches were implemented in attempt to solve the problem:

i. First, the raking benchmarking method iteratively adjusts district estimates until convergence is reached with the regional poverty rate. However, the FH model regional poverty rates are all within 5 percent of the direct estimates, as a result this had little effect on changing the district poverty rates.

ii. Next, the ratio method adjusts the district estimates using a constant factor.

iii. Finally, a method that incorporates the MSE estimates was also applied.

All three methods had minimal effect on the district poverty rates as they are all sensitive to the accuracy of sampling in the specific districts. The decision was made to treat all three districts as out of sample, which resulted in poverty rates more aligned with the neighboring districts.

We remove the 3 districts flagged by the SNBS and re-iterate the entire modeling exercise previously described, including both the model/variable selection process and the FH model estimation. The results are as follows:

Table 5: EBP Model (Regression Results)

Variables	Coefficients	Standard Error
Intercept	0.148	0.180
Rural Accessibility Index	0.004**	0.002
Production quantity for banana as a share of all crops	0.026***	0.008
Share of people living within 5km from conflicts	0.002**	0.001
Number of Schools	-0.0057***	0.002
Nugaal Region	-0.15**	0.069
Hiraan Region	0.350***	0.094
Bakool Region	0.207**	0.078

Statistical significance for each coefficient value, *** for 1%, ** for 5%

Table 6: Assessing Normality Assumptions

Model R^2		(Error Term) ε		(Random Effect) μ	
marginal	conditional	skewness	kurtosis	skewness	kurtosis
0.313	0.488	0.256	2.729	-0.251	2.524

Figure 7 below shows the updated district poverty estimates, while Table 8 presents the final regional and district poverty estimates and confidence intervals based on the FH model.

Figure 7: Adjusted Predicted District Poverty Rates and Number of Poor

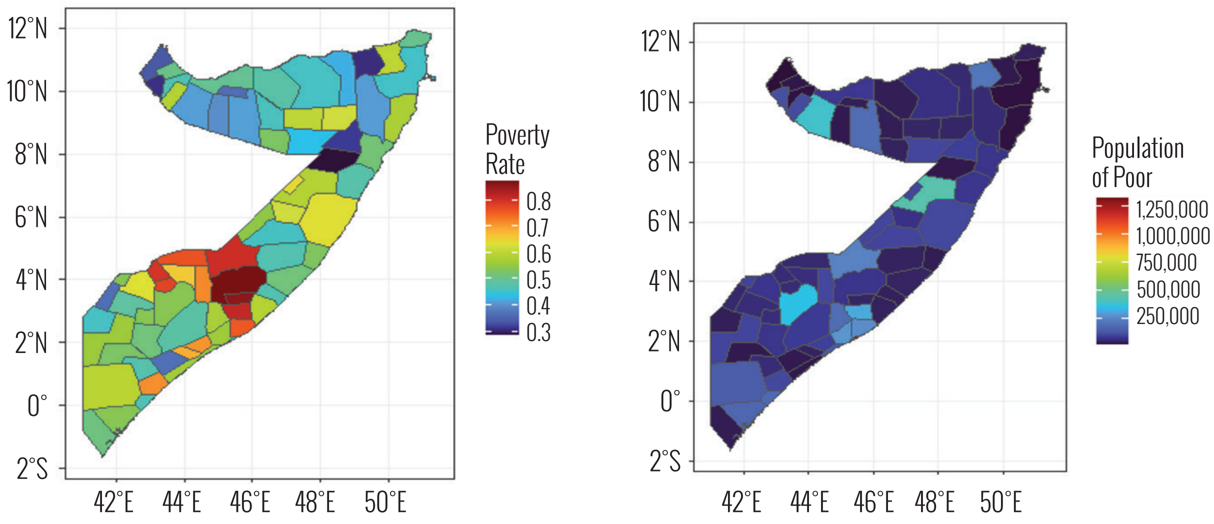


Table 8: Final FH Estimates

Region	Regional Poverty Rate (Direct)	Regional Poverty Rate CI (Direct)		District	District Poverty Rate (FH)	District Poverty Rate CI (FH)	
Awdal	0.357	0.263	0.452	Borama	0.308	0.224	0.391
				Baki	0.470	0.285	0.655
				Lughaye	0.504	0.309	0.699
				Zeylac	0.337	0.127	0.548
Woqooyi Galbeed	0.427	0.34	0.515	Hargeysa	0.411	0.306	0.515
				Berbera	0.497	0.384	0.610
				Gebiley	0.586	0.391	0.781
Togdheer	0.406	0.322	0.49	Burco	0.412	0.315	0.509
				Buuhoodle	0.528	0.372	0.684
				Owdweyne	0.398	0.236	0.560
				Sheikh	0.365	0.226	0.504
Sool	0.45	0.348	0.553	Laas Caanood	0.435	0.334	0.536
				Caynabo	0.460	0.278	0.641
				Taleex	0.620	0.422	0.818
				Xudun	0.601	0.445	0.757
Sanaag	0.436	0.314	0.558	Ceerigaabo	0.455	0.339	0.572
				Ceel Afweyn	0.480	0.298	0.662
				Laasqoray	0.423	0.258	0.589
Bari	0.366	0.285	0.447	Bossaso	0.305	0.217	0.394
				Bandarbeyla	0.577	0.423	0.732
				Caluula	0.500	0.322	0.678
				Iskushuban	0.453	0.290	0.617

Region	Regional Poverty Rate (Direct)	Regional Poverty Rate CI (Direct)		District	District Poverty Rate (FH)	District Poverty Rate CI (FH)	
				Qandala	0.600	0.432	0.769
				Qardho	0.409	0.250	0.568
Nugaal	0.349	0.276	0.422	Garoowe	0.315	0.216	0.413
				Burtinle	0.283	0.192	0.375
				Eyl	0.517	0.307	0.728
Mudug	0.679	0.568	0.789	Gaalkacyo	0.587	0.475	0.700
				Galdogob	0.640	0.466	0.814
				Hobyo	0.635	0.437	0.833
				Jariiban	0.476	0.329	0.623
				Xarardheere	0.532	0.389	0.674
Galgaduud	0.509	0.416	0.602	Dhuusamarreeb	0.457	0.348	0.566
				Cabudwaaq	0.547	0.421	0.672
				Cadaado	0.620	0.448	0.792
				Ceel Buur	0.470	0.328	0.612
				Ceel Dheer	0.512	0.365	0.659
Hiraan	0.815	0.7	0.931	Belet Weyne	0.804	0.677	0.931
				Bulo Burto	0.872	0.650	1.094
				Jalalaqsi	0.859	0.630	1.087
Middle Shabelle	0.866	0.739	0.992	Jowhar	0.813	0.666	0.960
				Adan Yabaal	0.472	0.324	0.619
				Balcad	0.755	0.589	0.921
				Cadale	0.588	0.403	0.773
Banadir	0.44	0.374	0.505	Banadir	0.458	0.393	0.523
Lower Shabelle	0.534	0.436	0.631	Marka	0.560	0.391	0.730
				Afgooye	0.562	0.468	0.657
				Baraawe	0.567	0.412	0.722
				Kurtunwaarey	0.685	0.520	0.850
				Qoryooley	0.707	0.529	0.884
				Sablaale	0.368	0.158	0.578
				Wanla Weyn	0.574	0.428	0.720
Bay	0.572	0.462	0.681	Baydhaba	0.539	0.438	0.641
				Buur Hakaba	0.482	0.342	0.621
				Diinsoor	0.513	0.376	0.650
				Qansax Dheere	0.540	0.402	0.677
Bakool	0.743	0.632	0.855	Xudur	0.658	0.500	0.816
				Ceel Barde	0.762	0.615	0.910
				Tayeeglow	0.716	0.521	0.911
				Waajid	0.779	0.592	0.965
				Rab Dhuure	0.795	0.594	0.997
Gedo	0.549	0.441	0.658	Garbahaarey	0.487	0.352	0.622
				Baardheere	0.558	0.390	0.725
				Belet Xaawo	0.367	0.237	0.496

Region	Regional Poverty Rate (Direct)	Regional Poverty Rate CI (Direct)		District	District Poverty Rate (FH)	District Poverty Rate CI (FH)	
				Ceel Waaq	0.456	0.298	0.615
				Doolow	0.538	0.370	0.706
				Luuq	0.632	0.477	0.787
Middle Juba				Bu'aale	0.473	0.301	0.645
				Jilib	0.712	0.523	0.901
				Saakow	0.527	0.387	0.668
Lower Juba	0.578	0.468	0.689	Kismaayo	0.539	0.426	0.652
				Afmadow	0.595	0.456	0.734
				Badhaadhe	0.510	0.372	0.648
				Jamaame	0.468	0.299	0.637

Figure 8: Correlation between FH Model Estimates and Direct Estimates at District Level

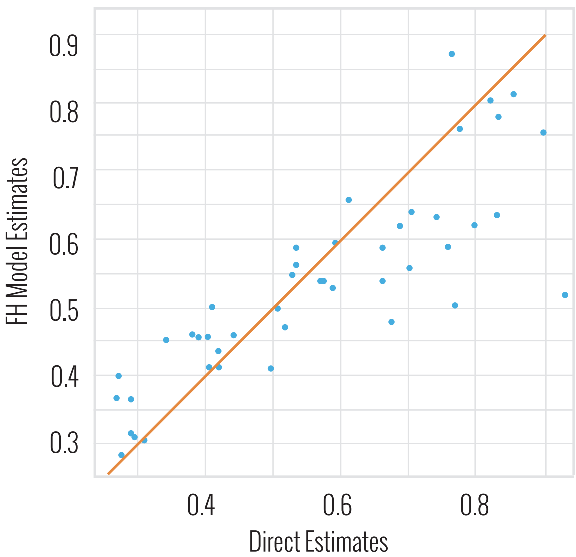
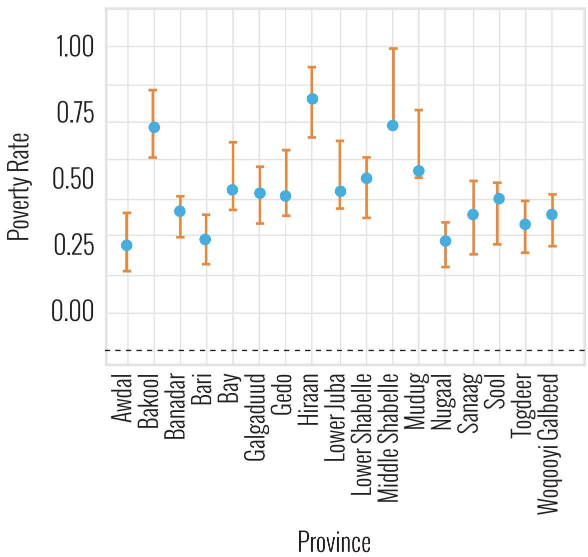


Figure 9: Comparison of FH and SIHBS Estimates (without the 3 SNBS flagged districts)



Note: Author calculations of Direct Estimate 95% CI at the province level (Red error bars) compared with FH estimates at same level (blue dots).

APPENDIX

Figure A1: Correlation between FH Model Estimates and Direct Estimates at District Level (Original Model)

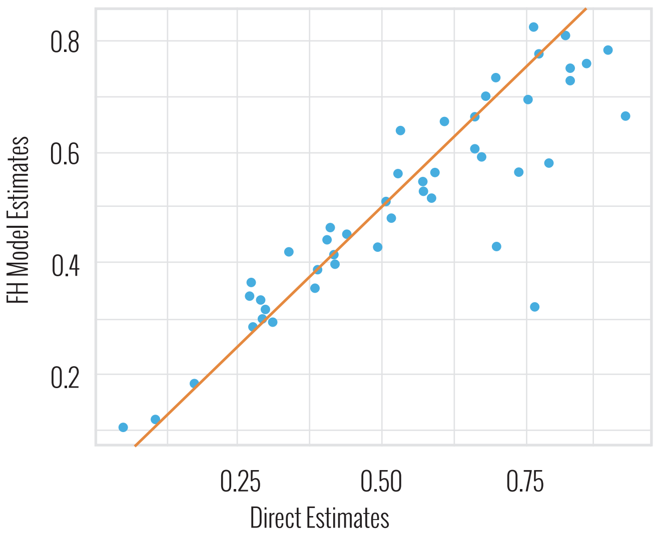


Table A1: Comparing FH Estimates to Direct Estimates at Regional Level (Original Model)

Province	Survey	FH Estimate 95% Confidence Intervals	
	Direct Estimate	Lower Bound	Upper Bound
Awdal	0.343	0.254	0.432
Woqooyi Galbeed	0.427	0.340	0.515
Togdheer	0.406	0.322	0.490
Sool	0.450	0.348	0.553
Sanaag	0.363	0.266	0.460
Bari	0.366	0.285	0.447
Nugaal	0.349	0.276	0.422
Mudug	0.679	0.568	0.789
Galgaduud	0.509	0.416	0.602
Hiraan	0.815	0.700	0.931
Middle Shabelle	0.866	0.739	0.992
Banadir	0.440	0.374	0.505
Lower Shabelle	0.534	0.436	0.631
Bay	0.572	0.462	0.681
Bakool	0.743	0.632	0.855
Gedo	0.521	0.419	0.624
Lower Juba	0.578	0.468	0.689

Table A2: District-Level Poverty Map Table

Region	Direct	Direct Estimate	FH Model Estimate
Awdal	Borama	0.298	0.316
Awdal	Baki	0.519	0.483
Awdal	Lughaye	0.768	0.321
Awdal	Zeylac	0.173	0.181
Woqooyi Galbeed	Hargeysa	0.406	0.443
Woqooyi Galbeed	Berbera	0.506	0.508
Woqooyi Galbeed	Gebiley	0.534	0.636
Togdheer	Burco	0.419	0.397
Togdheer	Buuhoodle	0.590	0.515
Togdheer	Owdweyne	0.273	0.361
Togdheer	Sheikh	0.291	0.329
Sool	Laas Caanood	0.418	0.417
Sool	Caynabo	0.383	0.352
Sool	Taleex	0.795	0.579
Sool	Xudun	NA	0.586
Sanaag	Ceerigaabo	0.388	0.384
Sanaag	Ceel Afweyn	0.675	0.589
Sanaag	Laasqoray	0.105	0.118
Bari	Bossaso	0.310	0.293
Bari	Bandarbeyla	NA	0.463
Bari	Caluula	0.412	0.464
Bari	Iskushuban	0.342	0.420
Bari	Qandala	NA	0.498
Bari	Qardho	0.496	0.429
Nugaal	Garoowe	0.292	0.300
Nugaal	Burtinle	0.276	0.286
Nugaal	Eyl	0.929	0.668
Mudug	Gaalkacyo	0.662	0.662
Mudug	Galdogob	0.703	0.734
Mudug	Hobyo	0.829	0.726
Mudug	Jariiban	NA	0.588
Mudug	Xarardheere	NA	0.541
Galgaduud	Dhuusamarreeb	0.405	0.438
Galgaduud	Cabudwaaq	0.528	0.559
Galgaduud	Cadaado	0.688	0.700
Galgaduud	Ceel Buur	NA	0.632
Galgaduud	Ceel Dheer	NA	0.746
Hiraan	Belet Weyne	0.821	0.812
Hiraan	Bulo Burto	0.765	0.824
Hiraan	Jalalaqsi	NA	0.775
Middle Shabelle	Jowhar	0.855	0.761

Region	Direct	Direct Estimate	FH Model Estimate
Middle Shabelle	Cadale	0.758	0.692
Banadir	Banadir	0.440	0.451
Lower Shabelle	Marka	NA	0.483
Lower Shabelle	Afgooye	0.534	0.558
Lower Shabelle	Baraawe	NA	0.551
Lower Shabelle	Kurtunwaarey	NA	0.485
Lower Shabelle	Qoryooley	NA	0.558
Lower Shabelle	Sablaale	NA	0.500
Lower Shabelle	Wanla Weyn	NA	0.810
Bay	Baydhaba	0.572	0.546
Bay	Buur Hakaba	NAa	0.593
Bay	Diinsoor	NA	0.544
Bay	Qansax Dheere	NA	0.520
Bakool	Xudur	0.613	0.656
Bakool	Ceel Barde	0.776	0.775
Bakool	Tayeeglow	NA	0.768
Bakool	Waajid	0.831	0.752
Bakool	Rab Dhuure	NA	0.784
Gedo	Garbahaarey	0.050	0.104
Gedo	Baardheere	0.702	0.428
Gedo	Belet Xaawo	0.272	0.340
Gedo	Ceel Waaq	NA	0.310
Gedo	Doolow	0.663	0.604
Gedo	Luuq	0.739	0.561
Lower Juba	Kismaayo	0.576	0.529
Lower Juba	Afmadow	0.592	0.565
Lower Juba	Badhaadhe	NA	0.330
Lower Juba	Jamaame	NA	0.417

Table A3: Sparsely Sampled Districts

District Code	District Name	Number of Households	Number of EAs
S01103	Lughaye	12	1
S01102	Baki	24	2
S01104	Zeylac	24	2
S01403	Taleex	24	2
S01803	Hobyo	24	2
S02104	Cadale	24	2
S02601	Garbahaarey	24	2
S01303	Owdweyne	36	3
S01402	Caynabo	36	3
S01502	Ceel Afweyn	36	3

Note: The colored districts were flagged by the NSO.

Table 9: District-Level Poverty Map Table (Comparing Poverty Rates with and without the 3 SIHBS flagged areas)

Region	District	Final FH Model Estimate	Initial FH Model Estimate
Awdal	Borama	0.308	0.3158937
Awdal	Baki	0.470	0.4834294
Awdal	Lughaye	0.504	0.3212016
Awdal	Zeylac	0.337	0.1814266
Woqooyi Galbeed	Hargeysa	0.411	0.4433516
Woqooyi Galbeed	Berbera	0.497	0.5079388
Woqooyi Galbeed	Gebiley	0.586	0.6362767
Togdheer	Burco	0.412	0.3970575
Togdheer	Buuhoodle	0.528	0.5153860
Togdheer	Owdweyne	0.398	0.3612637
Togdheer	Sheikh	0.365	0.3290785
Sool	Laas Caanood	0.435	0.4165817
Sool	Caynabo	0.460	0.3523711
Sool	Taleex	0.620	0.5787886
Sool	Xudun	0.601	0.5860947
Sanaag	Ceerigaabo	0.455	0.3843609
Sanaag	Ceel Afweyn	0.480	0.5890278
Sanaag	Laasqoray	0.423	0.1177883
Bari	Bossaso	0.305	0.2932197
Bari	Bandarbeyla	0.577	0.4632127
Bari	Caluula	0.500	0.4643615
Bari	Iskushuban	0.453	0.4197501
Bari	Qandala	0.600	0.4980277
Bari	Qardho	0.409	0.4291898
Nugaal	Garoowe	0.315	0.2997191
Nugaal	Burtinle	0.283	0.2860153
Nugaal	Eyl	0.517	0.6675905
Mudug	Gaalkacyo	0.587	0.6619848
Mudug	Galdogob	0.640	0.7338951
Mudug	Hobyo	0.635	0.7257320
Mudug	Jariiban	0.476	0.5881161
Mudug	Xarardheere	0.532	0.5414604
Galgaduud	Dhuusamarreeb	0.457	0.4380751
Galgaduud	Cabudwaaq	0.547	0.5593798
Galgaduud	Cadaado	0.620	0.7002371
Galgaduud	Ceel Buur	0.470	0.6315186
Galgaduud	Ceel Dheer	0.512	0.7459147
Hiraan	Belet Weyne	0.804	0.8117509
Hiraan	Bulo Burto	0.872	0.8241094
Hiraan	Jalalaqsi	0.859	0.7751405
Middle Shabelle	Jowhar	0.813	0.7609432
Middle Shabelle	Adan Yabaal	0.472	0.7364308

Region	District	Final FH Model Estimate	Initial FH Model Estimate
Middle Shabelle	Balcad	0.755	0.7819597
Middle Shabelle	Cadale	0.588	0.6920718
Banadir	Banadir	0.458	0.4506796
Lower Shabelle	Marka	0.560	0.4829774
Lower Shabelle	Afgooye	0.562	0.5583717
Lower Shabelle	Baraawe	0.567	0.5506913
Lower Shabelle	Kurtunwaarey	0.685	0.4852593
Lower Shabelle	Qoryooley	0.707	0.5583049
Lower Shabelle	Sablaale	0.368	0.5000973
Lower Shabelle	Wanla Weyn	0.574	0.8096928
Bay	Baydhaba	0.539	0.5460101
Bay	Buur Hakaba	0.482	0.5932554
Bay	Diinsoor	0.513	0.5439663
Bay	Qansax Dheere	0.540	0.5202485
Bakool	Xudur	0.658	0.6562309
Bakool	Ceel Barde	0.762	0.7753613
Bakool	Tayeeglow	0.716	0.7678550
Bakool	Waajid	0.779	0.7515893
Bakool	Rab Dhuure	0.795	0.7842077
Gedo	Garbahaarey	0.487	0.1040450
Gedo	Baardheere	0.558	0.4276045
Gedo	Belet Xaawo	0.367	0.3398158
Gedo	Ceel Waaq	0.456	0.3102931
Gedo	Doolow	0.538	0.6036457
Gedo	Luuq	0.632	0.5611038
Gedo	Bu'aale	0.473	0.5185329
Gedo	Jilib	0.712	0.4229273
Gedo	Saakow	0.527	0.6041703
Lower Juba	Kismaayo	0.539	0.5286834
Lower Juba	Afmadow	0.595	0.5650695
Lower Juba	Badhaadhe	0.510	0.3299765
Lower Juba	Jamaame	0.468	0.4167271

REFERENCES

Battese, George E, Rachel M Harter, and Wayne A Fuller. 1988. "An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data." *Journal of the American Statistical Association* 83 (401): 28–36.

Carrão, Hugo, Gustavo Naumann, and Paulo Barbosa. 2016. "Mapping Global Patterns of Drought Risk: An Empirical Framework Based on Sub-National Estimates of Hazard, Exposure and Vulnerability." *Global Environmental Change* 39: 108–24. https://doi.org/https://doi.org/10.1016/j.gloenvcha.2016.04.012.

Chi, Guanghua, Han Fang, Sourav Chatterjee, and Joshua E Blumenstock. 2022. "Microestimates of Wealth for All Low-and Middle-Income Countries." *Proceedings of the National Academy of Sciences* 119 (3): e2113658119.

Corral, Paul, Heath Henderson, and Sandra Segovia. 2025. "Poverty Mapping in the Age of Machine Learning." *Journal of Development Economics* 172: 103377. https://doi.org/https://doi.org/10.1016/j.jdeveco.2024.103377.

Corral, Paul, Isabel Molina, Alexandru Cojocaru, and Sandra Segovia. 2022. *Guidelines to Small Area Estimation for Poverty Mapping*. World Bank Washington.

Corral Rodas, Paul, Heath Henderson, and Sandra Segovia. 2023. "Poverty Mapping in the Age of Machine Learning." *Available at SSRN 4587156*.

Dick, Peter. 1995. "Modelling Net Undercoverage in the 1991 Canadian Census." *Survey Methodology* 21 (1): 45–54.

Fay III, Robert E, and Roger A Herriot. 1979. "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data." *Journal of the American Statistical Association* 74 (366a): 269–77.

Georganos, Stefanos, Assane Niang Gadiaga, Catherine Linard, Tais Grippa, Sabine Vanhuysse, Nicholus Mboga, Eléonore Wolff, Sébastien Dujardin, and Moritz Lennert. 2019. "Modelling the Wealth Index of Demographic and Health Surveys Within Cities Using Very High-Resolution Remotely Sensed Information." *Remote Sensing* 11 (21): 2543.

Meza, I., S. Siebert, P. Döll, J. Kusche, C. Herbert, E. Eyshi Rezaei, H. Nouri, et al. 2020. "Global-Scale Drought Risk Assessment for Agricultural Systems." *Natural Hazards and Earth System Sciences* 20 (2): 695–712. https://doi.org/10.5194/nhess-20-695-2020.

Molina, Isabel, and J. N. K. Rao. 2010. "Small Area Estimation of Poverty Indicators." *Canadian Journal of Statistics* 38 (3): 369–85. https://doi.org/https://doi.org/10.1002/cjs.10051.

Nelson, Andy, Daniel J Weiss, Jacob van Etten, Andrea Cattaneo, Theresa S McMenomy, and Jawoo Koo. 2019. "A Suite of Global Accessibility Indicators." *Scientific Data* 6 (1): 266.

Rivest, LOUIS-PAUL, and Eve Belmonte. 2000. "A Conditional Mean Squared Error of Small Area Estimators." *Survey Methodology* 26 (1): 67–78.

Seitz, William Hutchins. 2019. "Where They Live: District-Level Measures of Poverty, Average Consumption, and the Middle Class in Central Asia." *World Bank Policy Research Working Paper*, no. 8940.

Van Der Weide, Roy, Brian Blankespoor, Chris Elbers, and Peter Lanjouw. 2024. "How Accurate Is a Poverty Map Based on Remote Sensing Data? An Application to Malawi." *An Application to Malawi*.

You, Yong, and Michel A. Hidiroglou. 2012. "Sampling Variance Smoothing Methods for Small Area Proportion Estimators." SRID-2012-08E. Methodology Branch Working Paper, SRID.

You, Yong, and Mike Hidiroglou. 2023. "Application of Sampling Variance Smoothing Methods for Small Area Proportion Estimation." *Journal of Official Statistics* 39 (4): 571–90. https://doi.org/10.2478/jos-2023-0026.

WORLD BANK GROUP