Poverty Mapping

Partha Lahiri

Department of Mathematics & Joint Program In Survey Methodology
University of Maryland College Park

Based on joint research with Yuting Chen & Nicola Salvati

July 2, 2025

Complex parameters of interest

- Many measures of poverty and inequality are nonlinear functions of a quantitative welfare veraible for the population units.
- This makes many small area estimation methods, typically developed for estimating linear characteristics such as small area means, inapplicable.

FGT poverty measures; Foster et al. (1984)

- FGT poverty measures: $F_{\alpha i} = \frac{1}{N_i}\sum_{j=1}^{N_i} F_{\alpha ij}$,
- $F_{\alpha ij} = \left(\frac{z-E_{ij}}{z}\right)^{\alpha} I(E_{ij} < z),\ j = 1, \cdots, N_i;\ \alpha = 0, 1, 2,$ with $I(E_{ij} < z) = 1$ if $E_{ij} < z$, otherwise, it equals to 0;
- E_{ij}: a suitable quantitative measure of welfare for individual j in small area i,
- z: a fixed poverty line (threshold).

Empirical Best Prediction (EBP) for poverty indicators

- One-to-one transformation: $Y_{ij} = T(E_{ij})$
- $F_{\alpha ij} = \left(\frac{z - T^{-1}(Y_{ij})}{z}\right)^{\alpha} I(T^{-1}(Y_{ij}) < z) =: h_{\alpha}(Y_{ij})$,
- $\mathbf{y}_i = \{Y_{ij}; 1 \leqslant j \leqslant N_i\} = (\mathbf{y}'_{is}, \mathbf{y}'_{ir})'$: sub-vectors of sample units s and unsampled units r for ith area.

EBP for poverty indicators

The Best Predictor (BP) of $F_{\alpha i}$

$$\hat{F}^{B}_{\alpha i} = \mathrm{E}_{\mathbf{y}_r}(F_{\alpha i}|\mathbf{y}_s) = \frac{1}{N_i}\left\{\sum_{j\in s_i} F_{\alpha ij} + \sum_{j\notin s_i} \hat{F}^{B}_{\alpha ij}\right\},$$

where

- s_i: sampled individuals in area i
- $\hat{F}^{B}_{\alpha ij}$ is the BP of $F_{\alpha ij} = h_\alpha(Y_{ij})$ given by

$$\hat{F}^{B}_{\alpha ij} = \mathrm{E}_{\mathbf{y}_r}[h_\alpha(Y_{ij})|\mathbf{y}_s] = \int h_\alpha(t) f_{Y_{ij}}(t|\mathbf{y}_s)dt, \; j \notin s_i.$$

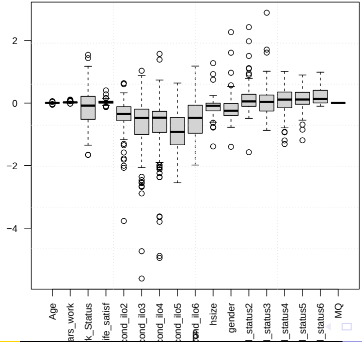
EBP for poverty indicators

- Molina and Rao (2010) [4] proposed a Monte Carlo approximation of EBP for $F_{\alpha i}$, where they used the traditional Nested Error Regression (NER) model to characterize the joint distribution of $\{Y_{ij}, i = 1, \cdots, m, j = 1, \cdots, N_i\}$.
- However, the homogeneity assumption of regression coefficients and sampling variances in NER may not hold in many real-world applications due to variations in socio-economic conditions, data quality, and sampling designs across areas.

Example

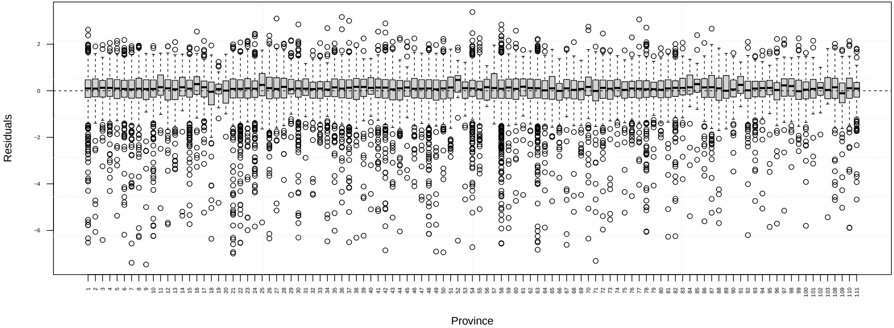
- 2021 Italy — European Survey on Income and Living Conditions (EUSILC)

Figure: Distributions of estimated regression coefficients across Italian provinces.



Example

Figure: Distributions of the residuals by province.



NER model with high-dimensional parameters

- An extension of the NER model:

$$Y_{ij} = \beta_0 + \mathbf{x}'_{ij}\boldsymbol{\beta}_{\boldsymbol{i}} + v_i + e_{ij}, i = 1, \cdots, m; j = 1, \cdots, N_i.$$

- β_0 is a common intercept;
- $\boldsymbol{\beta}_i$ is a $p \times 1$ vector of fixed regression coefficients for i^{th} area;
- v_i and e_{ij} are all independent with $v_i \sim N(0, \sigma_v^2)$ and $e_{ij} \sim N(0, \sigma_i^2)$.
- Nested error regression with high dimensional parameter (NERHDP).

Model parameter estimation

- Generalized estimating equations (GEE) with area-specific tuning parameters
- Allows to borrow strength across areas
- When the tuning parameters are known, the model parameters can be consistently estimated

Especially, for out-of-sample areas

- Less synthetic small area estimates compared to existing methods are produced for out-of-sample areas

Design-based simulations

- Fixed population: 2021 Italy European Survey on Income and Living Conditions (EUSILC)
- Parameters of interest: FGT measures ($\alpha = 0, 1$) for 107 Italian provinces
- Response variable: individual equalized annual net income (E_{ij}); shifted logarithm ($Y_{ij} = \log(E_{ij} + c)$).
- Simple random sampling (SRS) within province;
- Sample size: $0.1 \times N_i$

Design-based simulations

We compare the performances of:

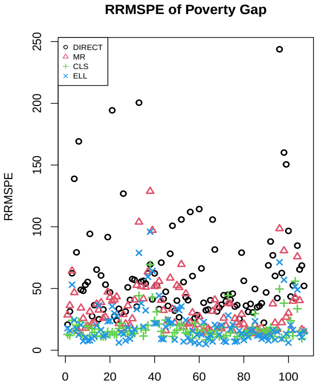
- DIRECT: Direct estimator
- MR: EBP introduced by Molina and Rao (2010) [4]
- CLS: EBP based on our proposed NERHDP model
- ELL: introduced by Elbers et al. (2003)[1]

Experiment 1: All 107 provinces are selected in the sample

Table: Performances of estimators/predictors of small area poverty indicators for 107 provinces in sample.

Predictor	Mean RRMSE (EFF) %		Mean absolute relative bias %	
	Head Count Ratio	Poverty Gap	Head Count Ratio	Poverty Gap
DIRECT	48.71 (1.00)	60.43 (1.00)	21.30	25.28
MR	28.73 (0.59)	33.39 (0.55)	26.9	23.75
CLS	22.94 (0.47)	17.81 (0.29)	18.31	12.76
ELL	22.48 (0.46)	20.25 (0.34)	21.41	17.85

Area-specific comparison



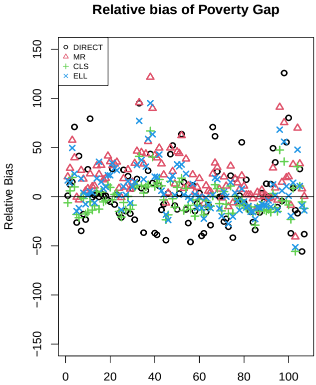


Figure: The RRMSE and Relative Bias of the four estimators of poverty gap for each area.

Area-specific comparison

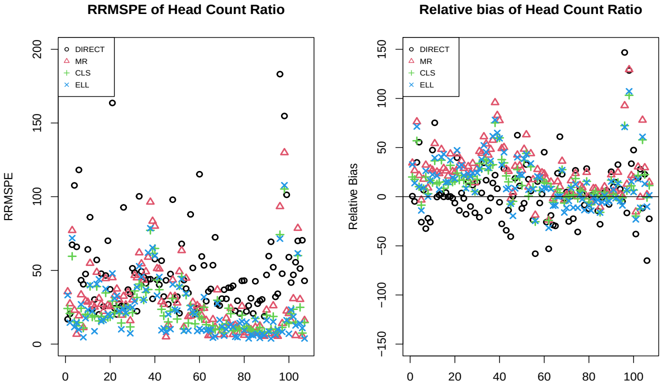


Figure: The RRMSE and Relative Bias of the four estimators of head count ratio for each area.

Experiment 2: 90 provinces are in sample & 17 are out of sample.

Table: Performances of estimators/predictors of small area poverty indicators for 17 out-of-sample provinces.

Predictor	Mean RRMSE (EFF) %		Mean absolute relative bias %	
	Head Count Ratio	Poverty Gap	Head Count Ratio	Poverty Gap
MR	35.36 (1.00)	41.87 (1.00)	34.52	31.79
CLS	27.49 (0.78)	16.69 (0.40)	25.22	15.27
ELL	26.04 (0.74)	20.55 (0.49)	24.92	17.44

Area-specific comparison

Figure: The RRMSE and Relative Bias of the EBP estimators of poverty gap for out of sample areas.

Area-specific comparison

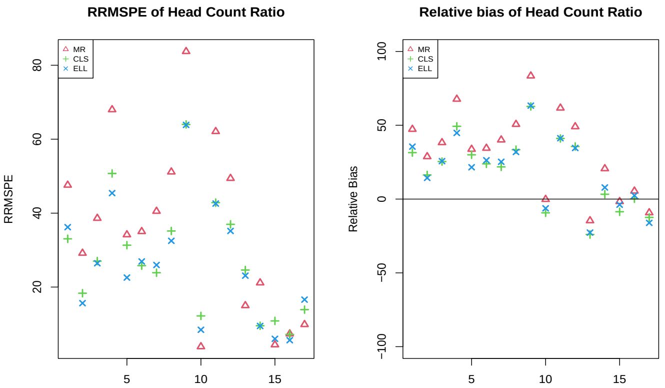


Figure: The RRMSE and Relative Bias of the EBP estimators of head count ratio for out of sample areas.

Application: poverty mapping for Albania

- 2002 Living Standards Measurement Survey (LSMS) data: $3,591$ households
- 374 municipalities: 213 sampled municipalities and 161 out-of-sample municipalities.
- Auxiliary data: 2001 Census, which covers 726,895 households across Albania.
- E_{ij}: household monthly consumption expenditure. Y_{ij}: shifted logarithm transformation of E_{ij}.
- The poverty line: 4,891 Leks per month.

Results

Table: Population size, sample size, direct and CLS estimates of headcount ratios (%) and poverty gaps (%), and the associated CVs (%) of direct and CLS estimators for the Albania municipalities with sample size closest to minimum, first quartile, median, third quartile, and maximum.

			Headcount Ratio				Poverty Gap			
Municipality	N_i	n_i	$\hat{F}_{0i}^{w}$	$\hat{F}_{0i}^{CLS}$	$CV(\hat{F}_{0i}^{w})$	$CV(\hat{F}_{0i}^{CLS})$	$\hat{F}_{1i}^{w}$	$\hat{F}_{1i}^{CLS}$	$CV(\hat{F}_{1i}^{w})$	$CV(\hat{F}_{1i}^{CLS})$
Hajmel	1111	6	33.33	23.82	56.96	27.01	1.52	5.92	70.56	31.31
Orenje	1419	16	12.50	20.69	66.19	15.06	2.00	4.97	66.03	17.44
B.Curri	1488	40	20.00	24.54	32.82	8.57	5.65	5.51	43.01	9.56
Gramsh	2503	64	23.55	17.97	24.48	0.69	4.75	3.64	38.93	21.90
Tirane	89764	600	12.43	11.72	10.91	4.46	2.49	2.31	14.71	4.72

- $\hat{F}_{0i}^{w}$: direct weighted estimator for headcount ratio;
- $\hat{F}_{1i}^{w}$: direct weighted estimator for poverty gap;

Albanian poverty maps

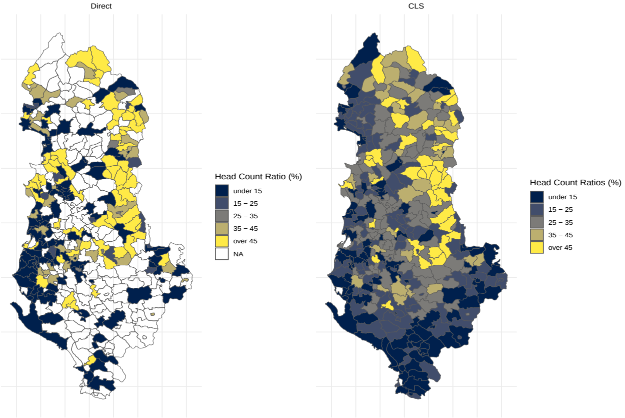


Figure: Municipality-level direct and CLS estimates of headcount ratios in Albania.

Albanian poverty maps

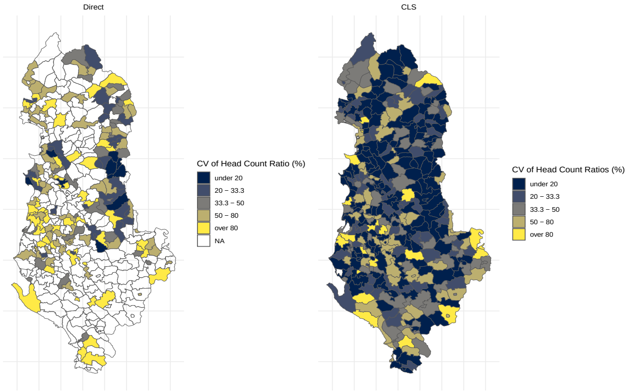


Figure: Municipality-level CV of headcount ratios for direct and CLS estimates in Albania.

Albanian poverty maps

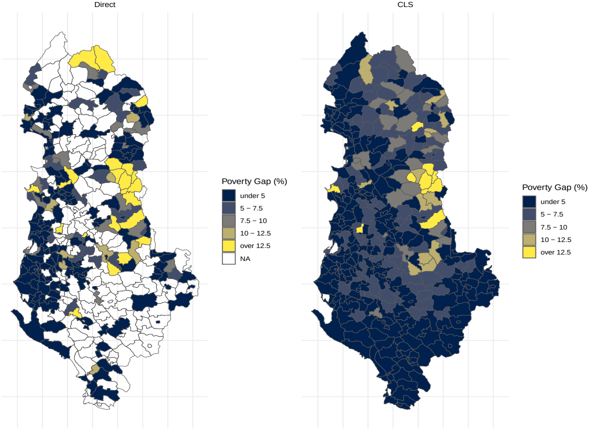


Figure: Municipality-level direct and CLS estimates of poverty gaps in Albania.

Albanian poverty maps

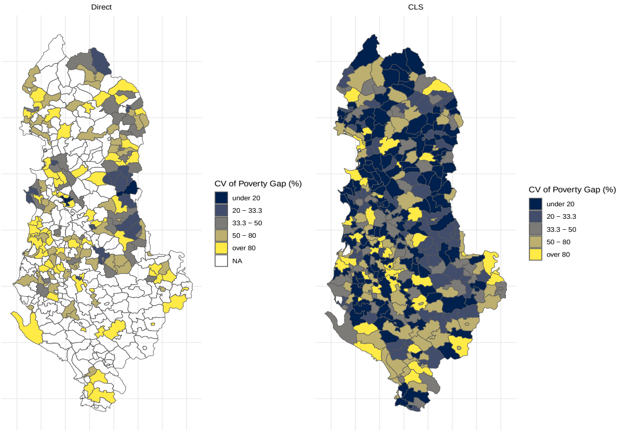


Figure: Municipality-level CV of poverty gaps for direct and CLS estimates in Albania.

A Quote from George Box

Statisticians, like artists, have the bad habit of falling in love with their models.

But as George Box said

Essentially, all models are wrong, but some are useful.

An interesting quote

...D.J. Finney once wrote about the statistician whose client comes in and says, 'Here is my mountain of trash. Find the gems that lie therein.' Finney's advice was to not throw him out of the office but to attempt to find out what he considers "gems". After all, if the trained statistician does not help, he will find some one who will...."

David Salsburg, ASA Connect Discussion

Paper & code

Figure: arXiv preprint: Chen et al. (2025)

Figure: GitHub: R function `hdp`

References

[1] C. Elbers, J.O. Lanjouw, and P. Lanjouw. Micro-level estimation of poverty and inequality. *Econometrica*, 71(1):355–364, 2003.

[2] James Foster, Joel Greer, and Erik Thorbecke. A class of decomposable poverty measures. *Econometrica: journal of the econometric society*, pages 761–766, 1984.

[3] Partha Lahiri and Nicola Salvati. A nested error regression model with high-dimensional parameter for small area estimation. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(2):212–239, 2023.

[4] Isabel Molina and Jon NK Rao. Small area estimation of poverty indicators. *Canadian Journal of statistics*, 38(3):369–385, 2010.

Thank you for attending!!