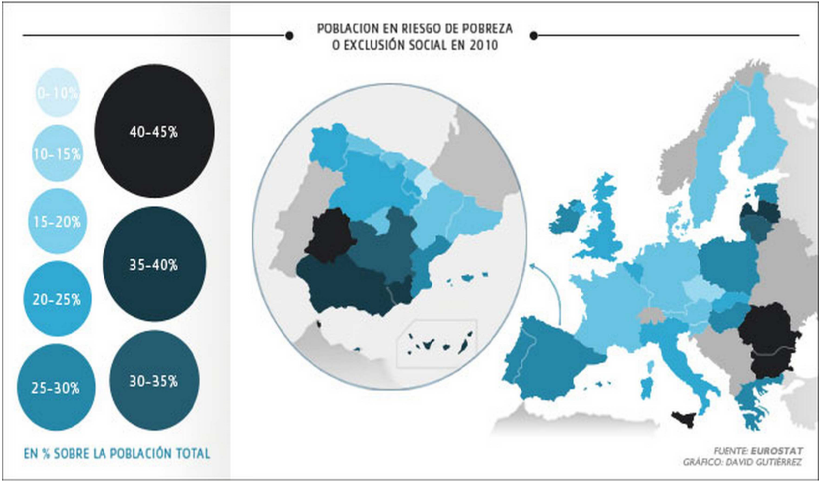
A review of small area estimation methods for poverty mapping

Isabel Molina
Dept. of Statistics, Univ. Carlos III de Madrid

Coauthors: María Guadarrama, Balgobin Nandram, J.N.K. Rao

POPULATION AT RISK OF POVERTY



EXAMPLE

Survey on Income and Living Conditions 2006

Total sample size: $n = 34,389$ persons.
Province×gender sample sizes:

(Barcelona,F)	(Córdoba,F)	(Tarragona,M)	(Soria,F)
1483	230	129	**17**

NOTATION

- U **finite** population of size N.
- U partitioned into D subsets $U_1, \ldots, U_D$ of sizes $N_1, \ldots, N_D$, called **domains** or **areas**.
- s sample of size n drawn from the population U.
- $s_d = s \cap U_d$ sub-sample from domain d of size n_d.
- $r_d = U_d - s_d$ out-of-sample elements from domain d.

DOMAIN PARAMETERS

- y_{dj} **outcome** for unit j in area d.
- $\mathbf{y}_d = (y_{d1}, \ldots, y_{dN_d})'$ vector of outcomes for **area** d.
- **Target quantities:** Possibly **non-linear** function of $\mathbf{y}_d$,

$$\delta_d = h_d(\mathbf{y}_d), \quad d = 1, \ldots, D.$$

- **Example:** mean of d-th area,

$$\bar{Y}_d = \frac{1}{N_d} \sum_{j=1}^{N_d} y_{dj}.$$

DIRECT ESTIMATION

- Based essentially on the **area-specific** sample data.
- $\pi_{dj} = P(j \in s_d)$ inclusion prob., $w_{dj} = 1/\pi_{dj}$ sampling weight.
- Sampling weights w_{dj} protect against **informative sampling** (probability of selection depending on outcomes).
- **Example:** Horvitz-Thompson direct estimator,

$$\hat{\bar{Y}}_d^{DIR} = \frac{1}{N_d} \sum_{j \in s_d} w_{dj} y_{dj}.$$

- Sampling variance $V_\pi(\hat{\bar{Y}}_d^{DIR})$ can be estimated easily with the **area-specific** data.

POVERTY AND INEQ. INDICATORS

- E_{dj} **welfare** measure for indiv. j in domain d.
- $z =$ poverty line.
- **FGT poverty indicator of order α for domain d:**

$$F_{\alpha d} = \frac{1}{N_d} \sum_{j=1}^{N_d} \left(\frac{z - E_{dj}}{z} \right)^{\alpha} I(E_{dj} < z), \quad \alpha \geq 0.$$

- When $\alpha = 0 \Rightarrow$ **Poverty incidence** (or at-risk-of-poverty rate)
- When $\alpha = 1 \Rightarrow$ **Poverty gap**
- **Other:** Quintile share ratio, Gini coef., Sen index, Theil index, Generalized entropy, Fuzzy monetary/supplementary index.

✓ *Foster, Greer & Thornbecke (1984), Econom.*
✓ *Neri, Ballini & Betti (2005), Stat. in Transition*

DIRECT ESTIMATORS

- FGT pov. indicator as a mean:

$$F_{\alpha d} = \frac{1}{N_d} \sum_{j=1}^{N_d} F_{\alpha dj}, \quad F_{\alpha dj} = \left(\frac{z - E_{dj}}{z} \right)^{\alpha} I(E_{dj} < z)$$

- HT estimator:

$$\hat{F}_{\alpha d}^{DIR} = \frac{1}{N_d} \sum_{j \in s_d} w_{dj} F_{\alpha dj}.$$

DIRECT ESTIMATORS

ADVANTAGES:

- **No model** assumptions.
- Sampling weights can be used $\Rightarrow$ Approx. **design-unbiased** even under informative sampling.
- Additivity (**Benchmarking** property):

$$\sum_{d=1}^{D} \hat{Y}_d^{DIR} = \hat{Y}^{DIR}.$$

DISADVANTAGES:

- $V_\pi(\hat{\hat{Y}}_d^{DIR}) \uparrow$ as $n_d \downarrow$. Very **inefficient** for small domains.
- Estimator of sampling error very **inefficient** for small domains.
- Cannot be calculated for out-of-sample areas.

INDIRECT ESTIMATORS

- **Indirect estimator:** It **borrows strength** from other areas by making some kind of **homogeneity** assumption across areas (model with **common** parameters) that uses **auxiliary information**.

FAY-HERRIOT (FH) MODEL

(i) Linking model:

$$\delta_d = \mathbf{x}_d'\boldsymbol{\beta} + u_d,\ \ u_d \stackrel{iid}{\sim} (0, \sigma_u^2), \quad d = 1, \ldots, D$$

σ_u^2 **unknown**

(ii) Sampling model:

$$\hat{\delta}_d^{DIR} = \delta_d + e_d,\ \ e_d \stackrel{ind}{\sim} (0, \psi_d), \quad d = 1, \ldots, D$$

u_d and e_d indep., $\psi_d = V_\pi(\hat{\delta}_d^{DIR}|\delta_d)$ **known** $\forall d$

(iii) Combined model: Linear mixed model

$$\hat{\delta}_d^{DIR} = \mathbf{x}_d'\boldsymbol{\beta} + u_d + e_d, \quad d = 1, \ldots, D.$$

✓ *Fay & Herriot (1979), JASA*

BEST LINEAR UNBIASED PREDICTOR

- Minimizes the MSE among **linear** and **unbiased** estimators.
- Easily obtained by fitting the mixed model:

$$\tilde{\delta}_d^{BLUP} = \mathbf{x}_d' \tilde{\boldsymbol{\beta}} + \tilde{u}_d,$$

where

$$\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\sigma_u^2) = \left(\sum_{d=1}^{D} \gamma_d \mathbf{x}_d \mathbf{x}_d' \right)^{-1} \sum_{d=1}^{D} \gamma_d \mathbf{x}_d \hat{\delta}_d^{DIR},$$

$$\tilde{u}_d = \tilde{u}_d(\sigma_u^2) = \gamma_d(\hat{\delta}_d^{DIR} - \mathbf{x}_d' \tilde{\boldsymbol{\beta}}), \quad \gamma_d = \frac{\sigma_u^2}{\sigma_u^2 + \psi_d}$$

GOOD PROPERTY OF THE BLUP

- Weighted combination of direct and "regression synthetic" estimator:

$$\tilde{\delta}_d^{BLUP} = \gamma_d \, \hat{\delta}_d^{DIR} + (1-\gamma_d)\mathbf{x}_d'\tilde{\boldsymbol{\beta}}, \quad \gamma_d = \frac{\sigma_u^2}{\sigma_u^2 + \psi_d}.$$

- When $\hat{\delta}_d^{DIR}$ is **reliable** ($\downarrow \psi_d$) or when area heterogeneity is not well explained by $\mathbf{x}_d'\tilde{\boldsymbol{\beta}}$ ($\uparrow \sigma_u^2$), $\tilde{\delta}_d^{BLUP} \longrightarrow \hat{\delta}_d^{DIR}$.
- Otherwise, if $\hat{\delta}_d^{DIR}$ **unreliable** or $\mathbf{x}_d'\tilde{\boldsymbol{\beta}}$ **reliable**, $\tilde{\delta}_d^{BLUP} \longrightarrow \mathbf{x}_d'\tilde{\boldsymbol{\beta}}$.

EMPIRICAL BLUP (EBLUP)

- $\tilde{\delta}_d^{BLUP}$ depends on unknown σ_u^2 through $\tilde{\beta}$ and γ_d:

$$\tilde{\delta}_d^{BLUP} = \tilde{\delta}_d^{BLUP}(\boldsymbol{\sigma_u^2})$$

- **Empirical** BLUP (EBLUP) of δ_d: $\hat{\sigma}_u^2$ estimator of σ_u^2,

$$\hat{\delta}_d^{EBLUP} = \tilde{\delta}_d^{BLUP}(\hat{\boldsymbol{\sigma}}_{\mathbf{u}}^{\mathbf{2}}), \quad d = 1, \ldots, D$$

- The EBLUP remains **model-unbiased** under certain conditions (typically satisfied).
- MSE of EBLUP under FH model can be approximated with $o(1/D)$ bias **under normality**.

NESTED ERROR MODEL

- Nested error **unit level** model:

$$y_{dj} = \mathbf{x}'_{dj}\boldsymbol{\beta} + u_d + e_{dj}, \quad j = 1, \ldots, N_d, \; d = 1, \ldots, D$$
$$u_d \overset{iid}{\sim} N(0, \sigma_u^2), \quad e_{dj} \overset{iid}{\sim} N(0, \sigma_e^2)$$

✓ *Battese, Harter & Fuller (1988), JASA*

- The distribution of incomes E_{dj} is highly right skewed.
- Select a transformation $T()$ such that the distribution of $y_{dj} = T(E_{dj})$ is approximately Normal.
- **Assumption:** $y_{dj} = T(E_{dj})$ satisfies the nested error model.

EB METHOD FOR POVERTY ESTIMATION

- Poverty indicators in terms of $\mathbf{y}_d = (y_{d1}, \ldots, y_{dN_d})'$:

$$F_{\alpha d} = \frac{1}{N_d} \sum_{j=1}^{N_d} \left\{ \frac{z - T^{-1}(y_{dj})}{z} \right\}^{\alpha} I\left\{ T^{-1}(y_{dj}) < z \right\} = h_{\alpha}(\mathbf{y}_d).$$

- Partition $\mathbf{y}_d$ into sample and out-of-sample: $\mathbf{y}_d = (\mathbf{y}_{ds}', \mathbf{y}_{dr}')'$
- **Best predictor:** Minimizes the MSE

$$\tilde{F}_{\alpha d}^{B} = E_{\mathbf{y}_{dr}}\left[F_{\alpha d} | \mathbf{y}_{ds}; \boldsymbol{\beta}, \sigma_u^2, \sigma_e^2 \right].$$

- **Empirical best (EB) predictor:** $\hat{F}_{\alpha d}^{EB} = \tilde{F}_{\alpha d}^{B}(\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2, \hat{\sigma}_e^2)$.

✓ *Molina and Rao (2010), CJS*

MODEL-BASED EXPERIMENT

- Simulate $I = 10^4$ populations from the **nested-error model**.
- For each population $i = 1, \ldots, 10^4$, compute true domain FGT poverty indicators $F_{\alpha d}^{(i)}$ for $\alpha = 0, 1$ and $d = 1, \ldots, D$.
- Take the sample part of each population (assuming SRS within each domain) and compute EB, direct and ELL estimates (*Elbers, Lanjouw and Lanjouw (2003), Econometrica*).
- Approximate **true MSEs** of EB estimators as

$$\mathrm{MSE}(\hat{F}_{\alpha d}^{EB}) = \frac{1}{I}\sum_{i=1}^{I}\left(\hat{F}_{\alpha d}^{EB(i)} - F_{\alpha d}^{(i)}\right)^2, \ \alpha = 0, 1, \ d = 1, \ldots, D.$$

- Similarly for direct and ELL estimators.

MODEL-BASED EXPERIMENT

- **Population and sample sizes:**

$$N = 20000, \quad \mathbf{D = 80}$$
$$N_d = 250, \quad \mathbf{n_d = 50}, \; d = 1, \ldots, D$$

- **Variance components:**

$$\sigma_e^2 = (0.5)^2, \quad \sigma_u^2 = (0.15)^2$$

- **Explanatory variables:** 2 dummies:

$$X_1 \in \{0, 1\}, \quad p_{1d} = 0.3 + 0.5d/80, \quad d = 1, \ldots, D.$$
$$X_2 \in \{0, 1\}, \quad p_{2d} = 0.2, \quad d = 1, \ldots, D.$$

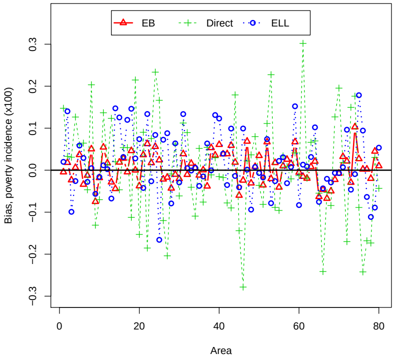
- **Coefficients:**

$$\boldsymbol{\beta} = (3, 0.03, -0.04)'$$

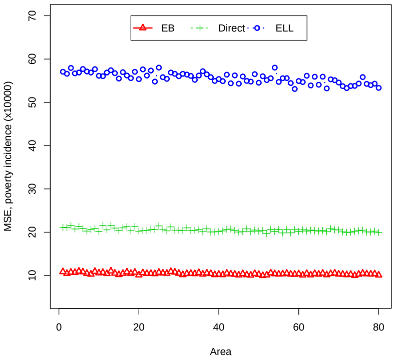
POVERTY INCIDENCE

- EB **much more efficient** than ELL and direct estimators.
- ELL even **less efficient** than direct estimators!

Bias (%)



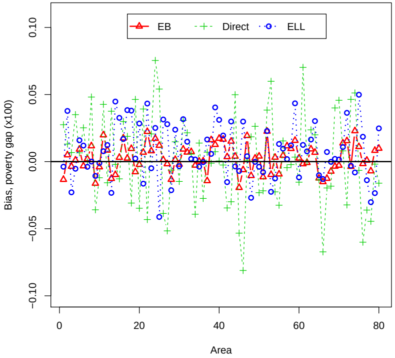
MSE ($\times 10^4$)



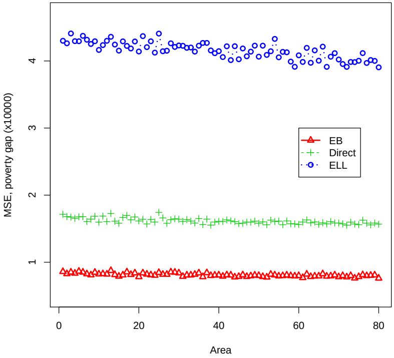
POVERTY GAP

- Same conclusions as for poverty incidence.

Bias (%)



MSE ($\times 10^4$)



HIERARCHICAL BAYES METHOD

- Reparameterization:

$$\rho = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$$

- Reparameterized nested-error model:

$$y_{dj}|u_d, \boldsymbol{\beta}, \sigma_e^2 \overset{ind}{\sim} N(\mathbf{x}'_{dj}\boldsymbol{\beta} + u_d, \sigma_e^2),$$
$$u_d|\rho, \sigma_e^2 \overset{ind}{\sim} N\left(0, \frac{\rho}{1-\rho}\sigma_e^2\right)$$

- **Noninformative** prior:

$$\pi(\boldsymbol{\beta}, \sigma_e^2, \rho) \propto 1/\sigma_e^2$$

✓ *Rao, Nandram & Molina (2014), Annals of Applied Statistics*

HIERARCHICAL BAYES METHOD

- **Proper** posterior density (provided $\mathbf{X}$ full column rank and ρ is in a closed interval from $(0,1)$):

$$\pi(\mathbf{u},\boldsymbol{\beta},\sigma_e^2,\rho|\mathbf{y}_s)=\pi_1(\mathbf{u}|\boldsymbol{\beta},\sigma_e^2,\rho,\mathbf{y}_s)\,\pi_2(\boldsymbol{\beta}|\sigma_e^2,\rho,\mathbf{y}_s)\,\pi_3(\sigma_e^2|\rho,\mathbf{y}_s)\,\pi_4(\rho|\mathbf{y}_s)$$

- Conditional distributions:

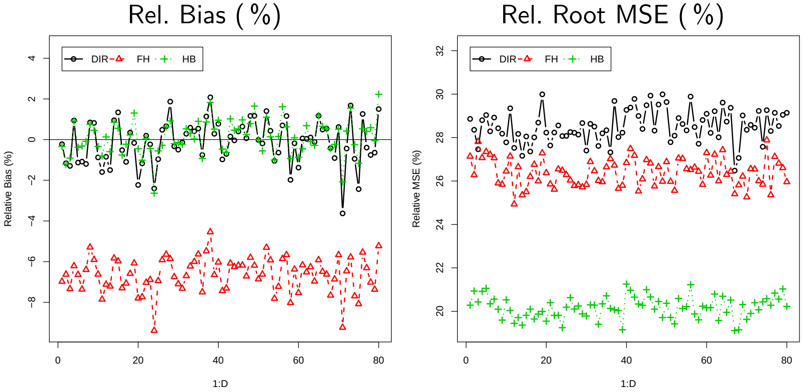
$$\begin{aligned} &u_d|\boldsymbol{\beta},\sigma_e^2,\rho,\mathbf{y}_s \overset{ind}{\sim} \text{Normal},\\ &\boldsymbol{\beta}|\sigma_e^2,\rho,\mathbf{y}_s \sim \text{Normal},\\ &\sigma_e^{-2}|\rho,\mathbf{y}_s \sim \text{Gamma} \end{aligned}$$

- $\pi_4(\rho|\mathbf{y}_s)$ not simple but ρ-values can be generated using a grid method.

✓ *Rao, Nandram & Molina (2014), Annals of Applied Statistics*

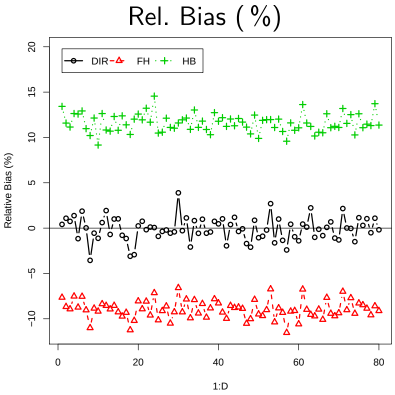
COMPARISON WITH FH ESTIMATES

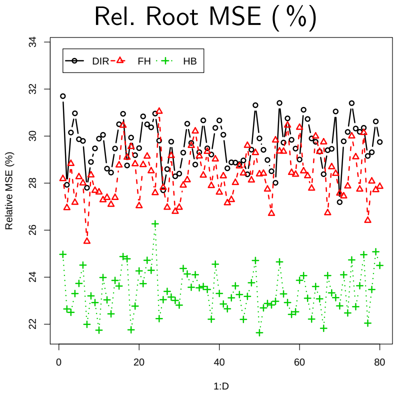
- FH estimates **biased** because of **linearity** problems.
- HB≈EB estimates nearly unbiased and **much more efficient**.



INFORMATIVE SAMPLING

- HB and EB estimates **biased** under **informative sampling**.
- FH estimates with known dependency of inclusion probabilities on true responses **less biased**.





PSEUDO EB

- Best predictor for additive area parameters:

$$\tilde{F}_{\alpha d}^{B} = E_{\mathbf{y}_{dr}}\left[F_{\alpha d} | \mathbf{y}_{ds}\right] = \frac{1}{N_d}\left[\sum_{j \in s_d} F_{\alpha dj} + \sum_{j \in r_d} \underbrace{E(F_{\alpha dj} | \mathbf{y}_{ds})}\right],$$

- Under the nested-error model:

$$E(F_{\alpha dj} | \mathbf{y}_{ds}) = E(F_{\alpha dj} | \bar{y}_d) \longrightarrow E(F_{\alpha dj} | \bar{y}_{dw}).$$

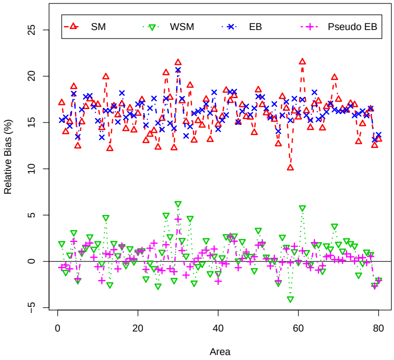
- **Pseudo Best predictor** for additive parameters:

$$\tilde{F}_{\alpha d}^{PB} = \frac{1}{N_d}\left[\sum_{j \in s_d} F_{\alpha dj} + \sum_{j \in r_d} \underbrace{E(F_{\alpha dj} | \bar{y}_{dw})}\right].$$

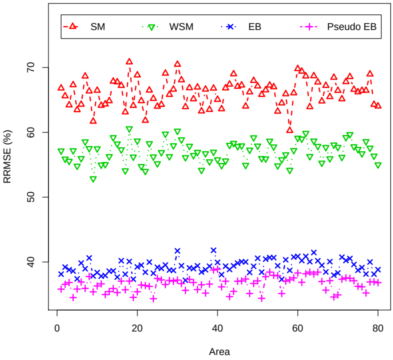
PSEUDO EB

- Including **sampling weights** reduces the design bias!
- Pseudo EB estimators do not lose much efficiency.

Rel. Bias (%)



Rel. Root MSE (%)



OTHER EXTENSIONS

- Existence of **two grouping levels** $\longrightarrow$ EB method under a **two-fold** nested-error model.
  ✓ *Marhuenda, Molina, Morales and Rao (2017), JRSSA*
- Particular non linear parameter: **Area mean** under a model for the **log-transformation** of the target variable $\longrightarrow$ **Explicit exact EB** estimator and **asymptotic MSE**.
  ✓ *Molina and Martín (2018), AOS*
- EB method for poverty estimation assumes **normality** for some transformation of the variable of interest $\longrightarrow$ Extension to **skewed** distributions.
  ✓ *Graf, Marín and Molina (2018), Test*

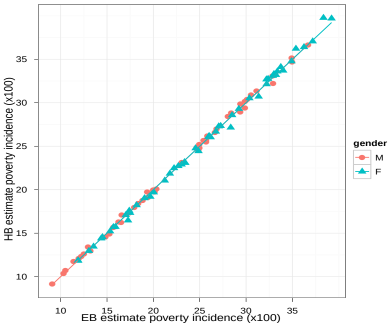
POVERTY MAPPING IN SPAIN

- **Data source:** Spanish Survey on Income and Living Conditions (EU-SILC) of 2006.
- **Target:** Calculate EB and HB estimates of poverty incidences and gaps for Spanish provinces by gender.
- **Areas:** $D = 52$ **provinces** for each **gender**. We fit a separate model for each gender.
- **Transformation:** We consider the nested-error model for the log-equivalized disposable income: $y_{dj} = T(E_{dj}) = \log(E_{dj} + k)$.
- **Explanatory variables:** indicators of 5 **age** groups, of having Spanish **nationality**, of 3 **education** levels and of **labor** force status (unemployed, employed or inactive).

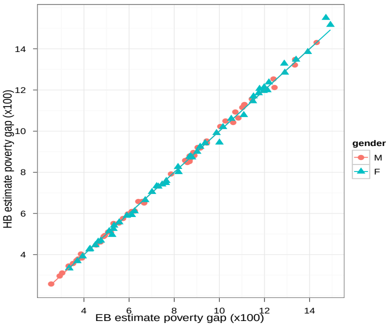
HIERARCHICAL BAYES METHOD

- HB estimates practically the same as EB ones. The same in simulations under the frequentist setup (**frequentist** validity).

Poverty incidence (%)



Poverty gap (%)



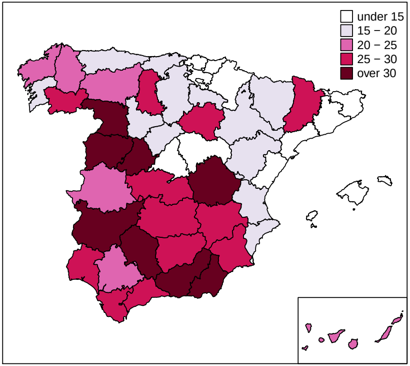
POVERTY MAPPING IN SPAIN

- Estimated CVs of direct, EB and HB estimators of poverty incidences for selected provinces crossed with gender:

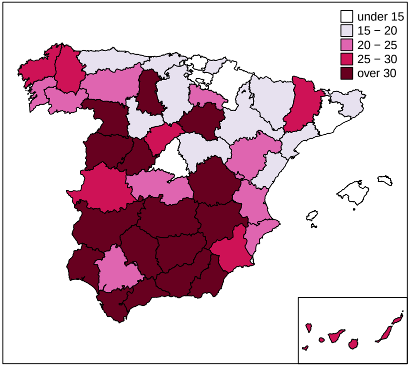
Province	Gender	n_d	Obs. Poor	$\hat{CV}$ Dir.	$\hat{CV}$ EB	$\hat{CV}$ HB
Soria	F	**17**	**6**	**51.87**	**16.56**	**19.82**
Tarragona	M	129	**18**	**24.44**	**14.88**	**12.35**
Córdoba	F	230	73	13.05	6.24	6.93
Badajoz	M	472	175	8.38	3.48	4.24
Barcelona	F	1483	191	9.38	6.51	4.52

RESULTS

Poverty incidence (%): Men



Poverty incidence (%): Women

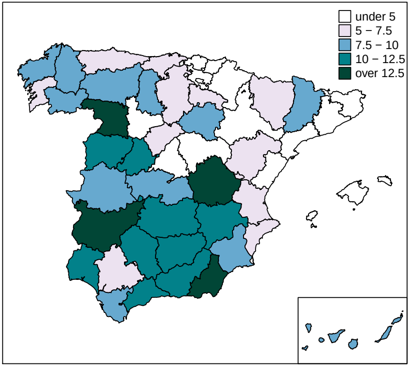


Pov.inc.$\geq$ 30 %, **Men:** Almería, Granada, Córdoba, Badajoz, Ávila, Salamanca, Zamora, Cuenca.

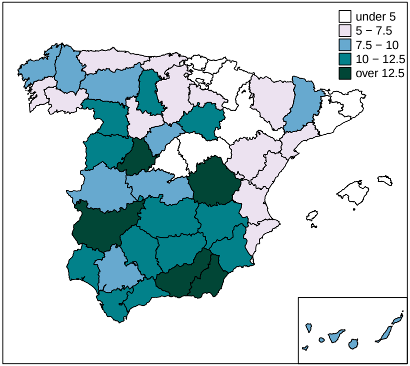
Women: also Jaén, Albacete, Ciudad Real, Palencia, Soria.

RESULTS

Poverty gap (%): Men



Poverty gap (%): Women



Pov.gap $\geq$ 12.5 %, **Men:** Almería, Badajoz, Zamora, Cuenca.
Women: Granada, Amería, Badajoz, Ávila, Cuenca.

COMPARISON WITH DIRECT ESTIMATORS

DISADVANTAGES:

- Require **model** assumptions (model checking important!).
- **Not design-unbiased** in general $\Longrightarrow$ Sampling weights can be incorporated to reduce design bias.
- Require **adjustment** to satisfy the benchmarking property:

$$\sum_{d=1}^{D} \hat{Y}_d = \hat{Y}^{DIR}.$$

ADVANTAGES:

- Very **efficient** for small domains.
- Estimator of model MSE **efficient** for small domains as well.
- Can be calculated for **out-of-sample** areas.

SOFTWARE

The R package `sae` contains functions:

- FH model: `eblupFH`, `mseFH`.
- Spatial FH model: `eblupSFH`, `mseSFH`, `pbmseSFH`, `npbmseSFH`.
- Spatio-temporal FH model: `eblupSTFH`, `pbmseSTFH`.
- Nested-error model: `eblupBHF`, `pbmseBHF`.
- EB method: `ebBHF`, `pbmseebBHF`.
- Other: `direct`, `pssynt`, `ssd`.
- Data sets and examples.

MERCI BEAUCOUP!!