Generalized linear mixed models for small area estimation

Tomáš Hobza

Department of Mathematics

Czech Technical University in Prague, Czech Republic

Based on joined work with

Domingo Morales and **Yolanda Marhuenda**

University of Miguel Hernández, Elche, Spain

SPMS 2022, 26.6.2022, Rumburk

Outline

1. Introduction
2. Description of the real data set
3. Unit level gamma mixed model
4. Application to real data
5. Conclusions

Introduction

- U **finite** population of size N.
- The population is partitioned into D subsets $U_1, \ldots, U_D$ of sizes $N_1, \ldots, N_D$, called **domains** or **areas**.
- Variable of interest Y.
- **Target:** to estimate the means of Y in the D domains/areas.

  Y_{dj} value of Y in unit j from domain d.

$$\bar{Y}_d = \frac{1}{N_d} \sum_{j=1}^{N_d} Y_{dj}, \quad d = 1, \ldots, D.$$

- We want to use data from a sample $S \subset U$ of size n drawn from the whole population.
- $S_d = S \cap U_d$ sub-sample from domain d of size n_d.

Introduction

- **Direct estimator:** Estimator that uses only the sample data from the corresponding domain (usually design-based),

$$\hat{\bar{Y}}_d^{DIR} = \sum_{j \in S_d} w_{dj} Y_{dj} / \sum_{j \in S_d} w_{dj}, \quad d = 1, \ldots, D.$$

  w_{dj} sampling weight of unit j within domain/area d.
  Under SRS without replacement within each area,

$$w_{dj} = \frac{N_d}{n_d}, \ \forall j \in S_d \Rightarrow \hat{\bar{Y}}_d^{DIR} = \frac{1}{n_d} \sum_{j \in S_d} Y_{dj}.$$

- **Problem:** n_d **small** for some d.
- **Small area/domain:** subset of the population that is target of inference and for which the direct estimator does not have enough precision.
- What does "enough precision" means? Some National Statistical Offices (GB, Spain) allow a maximum CV of 20%.

Introduction

- **Direct estimator:** Estimator that uses only the sample data from the corresponding domain (usually design-based),

$$\hat{\bar{Y}}_d^{DIR} = \sum_{j \in S_d} w_{dj} Y_{dj} / \sum_{j \in S_d} w_{dj}, \quad d = 1, \ldots, D.$$

  w_{dj} sampling weight of unit j within domain/area d.
  Under SRS without replacement within each area,

$$w_{dj} = \frac{N_d}{n_d},\ \forall j \in S_d \Rightarrow \hat{\bar{Y}}_d^{DIR} = \frac{1}{n_d} \sum_{j \in S_d} Y_{dj}.$$

- **Problem:** n_d **small** for some d.
- **Small area/domain:** subset of the population that is target of inference and for which the direct estimator does not have enough precision.
- What does "enough precision" means? Some National Statistical Offices (GB, Spain) allow a maximum CV of 20%.

Introduction

Small area estimation: field of statistics dealing with the problem of obtaining reliable estimates for domains for which only small samples or no samples are available

Idea: to use statistical models that "borrow strength"

- by using variables from related or similar areas
- through auxiliary data obtained from external sources (large surveys, census, administrative records)

SAE methods can be divided into

- "design-based" methods
- "model-based" methods

Description of the real data set

Data from 2013 Spanish Living Conditions Survey (SLCS) in the Autonomous Community of Valencia

We are interested in estimating the domain mean income and domain poverty proportions in 2013

We consider $D = 26$ domains, comarcas (counties) appearing in the sample

Total sample size: $n = 2\,492$ *(SLCS 2013)*

Smallest area: 10 records

Largest area: 405 records

Population size: $N = 4\,877\,512$

Auxiliary agregated data (totals of covariate patterns) are taken from SLFS 2013

Description of the real data set

Data from 2013 Spanish Living Conditions Survey (SLCS) in the Autonomous Community of Valencia

We are interested in estimating the domain mean income and domain poverty proportions in 2013

We consider $D = 26$ domains, comarcas (counties) appearing in the sample

Total sample size: $n = 2\,492$ *(SLCS 2013)*

Smallest area: 10 records

Largest area: 405 records

Population size: $N = 4\,877\,512$

Auxiliary agregated data (totals of covariate patterns) are taken from SLFS 2013

Description of the real data set

- SLCS provides information regarding the **household income** received during the last year

- **Equivalent personal income**
  - is calculated in order to take into account scale economies in households
  - it is assigned to each member of the household (denoted as y_{dj}).

- **The poverty risk** is the proportion of people with equivalent personal income below the poverty threshold.

  E.g. the 2013 Valencia poverty threshold is $z = 6999.6$ (in EUR).

Description of the real data set

- Our parameter of interest is

$$\delta_d = \frac{1}{N_d} \sum_{j=1}^{N_d} h(y_{dj}),$$

where h is a known measurable function.

- For $h(y) = y$ we obtain the area mean income

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

- For $h(y) = I(y < z)$ we obtain the area poverty proportions

$$p_d = \frac{1}{N_d} \sum_{j=1}^{N_d} I\left(y_{dj} < z\right).$$

Unit level gamma mixed model - Model 2

- D - domains, N_d - population size, $d = 1, \ldots, D$
- The distribution of the target variable y_{dj}, conditioned to the random effect v_d is

$$y_{dj}|_{v_d} \sim \text{Gamma}\Big(\nu_{dj}, \frac{\nu_{dj}}{\mu_{dj}}\Big), \quad \nu_{dj} = a_{dj}\varphi, \quad j = 1, \ldots, N_d.$$

- For the inverse of the mean parameter, we assume

$$g(\mu_{dj}) = \frac{1}{\mu_{dj}} = \boldsymbol{x}_{dj}^T \boldsymbol{\beta} + \phi v_d,$$

where
  - $\{v_d : d = 1, \ldots, D\}$ are i.i.d. $N(0, 1)$
  - y_{dj}'s are independent conditioned to $\boldsymbol{v}$.

- The vector of unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi, \varphi)$ is estimated by maximizing the Laplace approximation of the log-likelihood.

Unit level gamma mixed model - Model 2

- D - domains, N_d - population size, $d = 1, \dots, D$
- The distribution of the target variable y_{dj}, conditioned to the random effect v_d is

$$y_{dj}|_{v_d} \sim \text{Gamma}\big(\nu_{dj}, \frac{\nu_{dj}}{\mu_{dj}}\big), \quad \nu_{dj} = a_{dj}\varphi, \quad j = 1, \dots, N_d.$$

- For the inverse of the mean parameter, we assume

$$g(\mu_{dj}) = \frac{1}{\mu_{dj}} = \boldsymbol{x}_{dj}^T \boldsymbol{\beta} + \phi v_d,$$

where

- $\{v_d : d = 1, \dots, D\}$ are i.i.d. $N(0,1)$
- y_{dj}'s are independent conditioned to $\boldsymbol{v}$.

- The vector of unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi, \varphi)$ is estimated by maximizing the Laplace approximation of the log-likelihood.

Unit level gamma mixed model - Model 2

- D - domains, N_d - population size, $d = 1, \ldots, D$
- The distribution of the target variable y_{dj}, conditioned to the random effect v_d is

$$y_{dj}|_{v_d} \sim \text{Gamma}\big(\nu_{dj}, \frac{\nu_{dj}}{\mu_{dj}}\big), \quad \nu_{dj} = a_{dj}\varphi, \quad j = 1, \ldots, N_d.$$

- For the inverse of the mean parameter, we assume

$$g(\mu_{dj}) = \frac{1}{\mu_{dj}} = \boldsymbol{x}_{dj}^T \boldsymbol{\beta} + \phi v_d,$$

where

- $\{v_d : d = 1, \ldots, D\}$ are i.i.d. $N(0, 1)$
- y_{dj}'s are independent conditioned to $\boldsymbol{v}$.

- The vector of unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi, \varphi)$ is estimated by maximizing the Laplace approximation of the log-likelihood.

Empirical best predictor

- Our parameter of interest is

$$\delta_d = \frac{1}{N_d} \sum_{j=1}^{N_d} h(y_{dj}).$$

- Let us denote by S_d and R_d the sets of sampled and non-sampled individuals in domain d
- Best predictor (BP) of δ_d is

$$\hat{\delta}_d = \hat{\delta}_d(\boldsymbol{\theta}) = \frac{1}{N_d} \Big[\sum_{j \in S_d} h(y_{dj}) + \sum_{j \in R_d} E_{\boldsymbol{\theta}}[h(y_{dj}) | \boldsymbol{y}_s] \Big].$$

- We would need a census file with all the $\boldsymbol{x}$ variables
- Might be overcome if all the $\boldsymbol{x}$ variables are categorical

Empirical best predictor

- Our parameter of interest is

$$\delta_d = \frac{1}{N_d} \sum_{j=1}^{N_d} h(y_{dj}).$$

- Let us denote by S_d and R_d the sets of sampled and non-sampled individuals in domain d
- Best predictor (BP) of δ_d is

$$\hat{\delta}_d = \hat{\delta}_d(\boldsymbol{\theta}) = \frac{1}{N_d} \Big[\sum_{j \in S_d} h(y_{dj}) + \sum_{j \in R_d} E_{\boldsymbol{\theta}}[h(y_{dj}) | \boldsymbol{y}_s] \Big].$$

- We would need a census file with all the $\boldsymbol{x}$ variables
- Might be overcome if all the $\boldsymbol{x}$ variables are categorical

Empirical best predictor

- Our parameter of interest is

$$\delta_d = \frac{1}{N_d} \sum_{j=1}^{N_d} h(y_{dj}).$$

- Let us denote by S_d and R_d the sets of sampled and non-sampled individuals in domain d
- Best predictor (BP) of δ_d is

$$\hat{\delta}_d = \hat{\delta}_d(\boldsymbol{\theta}) = \frac{1}{N_d}\Big[\sum_{j \in S_d} h(y_{dj}) + \sum_{j \in R_d} E_{\boldsymbol{\theta}}[h(y_{dj})|\boldsymbol{y}_s]\Big].$$

- We would need a census file with all the $\boldsymbol{x}$ variables
- Might be overcome if all the $\boldsymbol{x}$ variables are categorical

Empirical best predictor

- Suppose that the covariates are categorical such that

$$\boldsymbol{x}_{dj} \in \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K\}.$$

- Then

$$\sum_{j \in R_d} E_{\boldsymbol{\theta}}[h(y_{dj})|\boldsymbol{y}_s] = \sum_{k=1}^{K} w_{dk} E_{\boldsymbol{\theta}}[h(y_{dk})|\boldsymbol{y}_s],$$

where $y_{dk} \sim Gamma\left(\nu_{dk}, \frac{\nu_{dk}}{\mu_{dk}}\right)$,

$$\mu_{dk} = \mu_{dk}(\boldsymbol{\theta}) = \left(\boldsymbol{z}_k^T \beta + \phi v_d\right)^{-1}$$

and

$$w_{dk} = \#\{j \in R_d : \boldsymbol{x}_{dj} = \boldsymbol{z}_k\}$$

is the size of the covariate class $\boldsymbol{z}_k$ at R_d (available from external data sources).

Empirical best predictor

- Suppose that the covariates are categorical such that

$$\boldsymbol{x}_{dj} \in \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K\}.$$

- Then

$$\sum_{j \in R_d} E_{\boldsymbol{\theta}}[h(y_{dj})|\boldsymbol{y}_s] = \sum_{k=1}^{K} w_{dk} E_{\boldsymbol{\theta}}[h(y_{dk})|\boldsymbol{y}_s],$$

where $y_{dk} \sim Gamma\left(\nu_{dk}, \frac{\nu_{dk}}{\mu_{dk}}\right)$,

$$\mu_{dk} = \mu_{dk}(\boldsymbol{\theta}) = \left(\boldsymbol{z}_k^T \boldsymbol{\beta} + \phi v_d\right)^{-1}$$

and

$$w_{dk} = \#\{j \in R_d : \boldsymbol{x}_{dj} = \boldsymbol{z}_k\}$$

is the size of the covariate class $\boldsymbol{z}_k$ at R_d (available from external data sources).

Empirical best predictor

- Under this categorical setup the BP of δ_d is

$$\hat{\delta}_d^{BP}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}[\delta_d|\boldsymbol{y}_s] = \frac{1}{N_d}\Big[\sum_{j\in S_d} h(y_{dj}) + \sum_{k=1}^{K} w_{dk} E_{\boldsymbol{\theta}}[h(y_{dk})|\boldsymbol{y}_s]\Big],$$

  where

$$E_{\boldsymbol{\theta}}[h(y_{dk})|\boldsymbol{y}_s]$$

  must be approximated numerically.

- The EBP of δ_d is then obtained as

$$\hat{\delta}_d^{EBP} = \hat{\delta}_d^{BP}(\hat{\boldsymbol{\theta}}).$$

Empirical best predictor

- Under this categorical setup the BP of δ_d is

$$\hat{\delta}_d^{BP}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}[\delta_d|\boldsymbol{y}_s] = \frac{1}{N_d}\Big[\sum_{j\in S_d} h(y_{dj}) + \sum_{k=1}^{K} w_{dk} E_{\boldsymbol{\theta}}[h(y_{dk})|\boldsymbol{y}_s]\Big],$$

where

$$E_{\boldsymbol{\theta}}[h(y_{dk})|\boldsymbol{y}_s]$$

must be approximated numerically.

- The EBP of δ_d is then obtained as

$$\hat{\delta}_d^{EBP} = \hat{\delta}_d^{BP}(\hat{\boldsymbol{\theta}})\,.$$

PLUG-IN estimator

The plug-in estimator of δ_d is

$$\tilde{\delta}_d = \tilde{\delta}_d(\hat{\boldsymbol{\theta}}) = \frac{1}{N_d}\Big[\sum_{j\in S_d} h(y_{dj}) + \sum_{k=1}^{K} w_{dk} h(\tilde{\mu}_{dk})\Big],$$

where

$$\tilde{\mu}_{dk} = \left(\boldsymbol{z}_k^T\hat{\boldsymbol{\beta}} + \hat{\phi}\hat{v}_d\right)^{-1}.$$

Marginal predictor

Let us consider the predicted marginal distribution of y_{dk}, i.e. the p.d.f. and d.f. of

$$Gamma\left(\widehat{\nu}_{dk}, \frac{\widehat{\nu}_{dk}}{\tilde{\mu}_{dk}}\right), \quad \widehat{\nu}_{dk} = a_{dk}\widehat{\varphi}.$$

The marginal predictor of δ_d is

$$\hat{\delta}_d^{MAR} = \frac{1}{N_d}\Big[\sum_{j\in S_d} h(y_{dj}) + \sum_{k=1}^{K} w_{dk} E[h(y_{dk})|\hat{\nu}_{dk}, \tilde{\mu}_{dk}]\Big].$$

- For $h(y) = y$ we get

$$E[h(y_{dk})|\hat{\nu}_{dk}, \tilde{\mu}_{dk}] = \int_0^{\infty} yf(y|\hat{\nu}_{dk}, \tilde{\mu}_{dk})\, dy = \tilde{\mu}_{dk}.$$

- For the function $h(y) = I(y < z)$

$$E[h(y_{dk})|\hat{\nu}_{dk}, \tilde{\mu}_{dk}] = \int_0^{z} f(y|\hat{\nu}_{dk}, \tilde{\mu}_{dk})\, dy = F_{\hat{\nu}_{dk},\tilde{\mu}_{dk}}(z).$$

Marginal predictor

Let us consider the predicted marginal distribution of y_{dk}, i.e. the p.d.f. and d.f. of

$$Gamma\left(\widehat{\nu}_{dk}, \frac{\widehat{\nu}_{dk}}{\tilde{\mu}_{dk}}\right), \quad \widehat{\nu}_{dk} = a_{dk}\widehat{\varphi}.$$

The marginal predictor of δ_d is

$$\hat{\delta}_d^{MAR} = \frac{1}{N_d}\Big[\sum_{j\in S_d} h(y_{dj}) + \sum_{k=1}^{K} w_{dk} E[h(y_{dk})|\hat{\nu}_{dk}, \tilde{\mu}_{dk}]\Big].$$

- For $h(y) = y$ we get

$$E[h(y_{dk})|\hat{\nu}_{dk}, \tilde{\mu}_{dk}] = \int_0^\infty y f(y|\hat{\nu}_{dk}, \tilde{\mu}_{dk})\, dy = \tilde{\mu}_{dk}.$$

- For the function $h(y) = I(y < z)$

$$E[h(y_{dk})|\hat{\nu}_{dk}, \tilde{\mu}_{dk}] = \int_0^z f(y|\hat{\nu}_{dk}, \tilde{\mu}_{dk})\, dy = F_{\hat{\nu}_{dk},\tilde{\mu}_{dk}}(z).$$

Marginal predictor

Let us consider the predicted marginal distribution of y_{dk}, i.e. the p.d.f. and d.f. of

$$Gamma\left(\widehat{\nu}_{dk}, \frac{\widehat{\nu}_{dk}}{\tilde{\mu}_{dk}}\right), \quad \widehat{\nu}_{dk} = a_{dk}\widehat{\varphi}.$$

The marginal predictor of δ_d is

$$\hat{\delta}_d^{MAR} = \frac{1}{N_d}\Big[\sum_{j\in S_d} h(y_{dj}) + \sum_{k=1}^{K} w_{dk} E[h(y_{dk})|\hat{\nu}_{dk}, \tilde{\mu}_{dk}]\Big].$$

- For $h(y) = y$ we get

$$E[h(y_{dk})|\hat{\nu}_{dk}, \tilde{\mu}_{dk}] = \int_0^{\infty} yf(y|\hat{\nu}_{dk}, \tilde{\mu}_{dk})\, dy = \tilde{\mu}_{dk}.$$

- For the function $h(y) = I(y < z)$

$$E[h(y_{dk})|\hat{\nu}_{dk}, \tilde{\mu}_{dk}] = \int_0^{z} f(y|\hat{\nu}_{dk}, \tilde{\mu}_{dk})\, dy = F_{\hat{\nu}_{dk},\tilde{\mu}_{dk}}(z).$$

Bootstrap estimator of MSE

1) Fit the model to the sample and calculate $\hat{\boldsymbol{\theta}}$.

2) Repeat B times $(b = 1, \dots, B)$:
   a) Generate bootstrap population from the assumed model with the estimated $\hat{\boldsymbol{\theta}}$
   b) Calculate the true quantity $\delta_d^{*(b)}$
   c) Extract bootstrap sample, calculate $\hat{\boldsymbol{\theta}}^{*(b)}$ and the predictor $\hat{\delta}_d^{*(b)}$.

3) Output:

$$mse^*(\hat{\mu}_d) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\delta}_d^{*(b)} - \delta_d^{*(b)}\right)^2 .$$

Simulation experiment - bootstrap

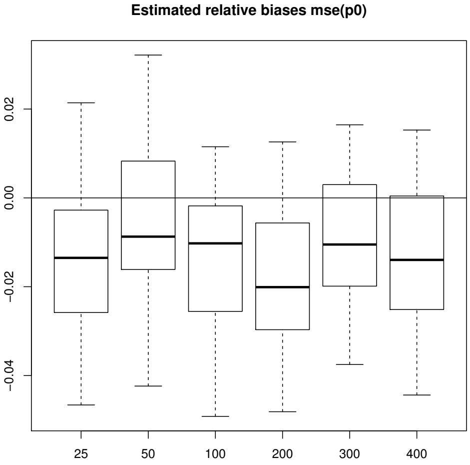


Figure 1. Relative biases of MSE estimators of MAR predictors for poverty proportions. Case $D = 30$, $n_d = 50$.

Simulation experiment - bootstrap

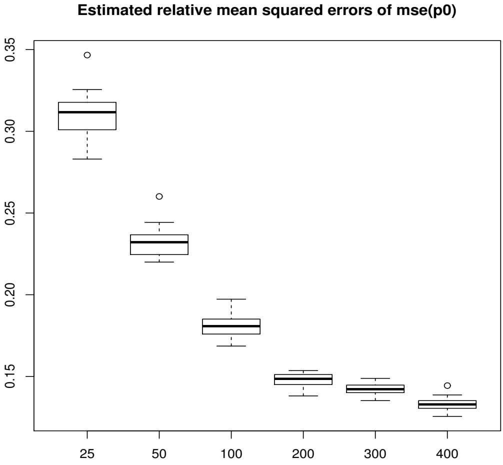


Figure 2. Relative root-MSEs of MSE estimators of MAR predictors for poverty proportions. Case $D = 30$, $n_d = 50$.

Model 2 for personal income (in 10 000 EUR):

We assume that

$$y_{dj}|_{v_d} \sim \text{Gamma}\big(\nu_{dj}, \frac{\nu_{dj}}{\mu_{dj}}\big), \quad d = 1, \dots, D, \; j = 1, \dots, N_d.$$

where v_d are i.i.d. $N(0,1)$, $\nu_{dj} = a_{dj}\varphi$ and

$$g(\mu_{dj}) = \frac{1}{\mu_{dj}} = \beta_0 + \beta_1 \text{Employed}_{dj} + \beta_2 \text{Unemployed}_{dj} + \phi v_d.$$

To fit the Model 2, we need the constants a_{dj}.

Application to real data

Algorithmic procedure:

1. Fit Model 1 to data and calculate the plug-in $\tilde{\mu}_{dj}$.
2. Fit the Model 2 to the data, assuming that

$$a_{dj} = \tilde{\mu}_{dj}^{t}, \qquad \text{for} \quad t \in (0.25, 3)$$

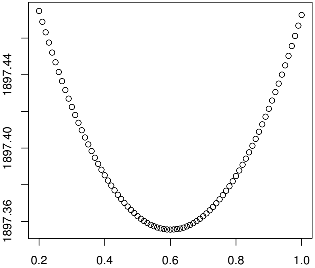
   is known.
3. For each considered t, calculate the plug-in $\hat{\mu}_{dj}^{(t)}$ and the sum of the squared residuals

$$r^2(t) = \sum_{d=1}^{D} \sum_{j=1}^{n_d} (y_{dj} - \hat{\mu}_{dj}^{(t)})^2.$$

4. Select t_* minimizing $r^2(t)$.
5. Do the inferences with Model 2 and $a_{dj} = \tilde{\mu}_{dj}^{t_*}$ known.

Application to real data

For the considered data set, the optimal choice is $t_* = 0.60$.



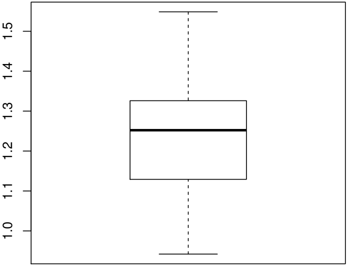


Figure 3: Function $r^2(t)$ (left) and boxplot of a_{dj} (right).

Application to real data

	estimate	standard error	p-value
$\hat{\beta}_0$	0.775	0.0132	$<$ 2E-16
$\hat{\beta}_1$	-0.141	0.0157	$<$ 2E-16
$\hat{\beta}_2$	0.140	0.0300	3.09E-06
$\hat{\phi}$	0.1113	0.0112	$<$ 2E-16
$\hat{\varphi}$	2.4646	0.0675	$<$ 2E-16

Table 2: Parameter estimates under Model 2.

Application to real data

Log-linear normal mixed model (MODEL 3):

- Let us consider the *log* transformation of data

$$z_{dj} = \log(y_{dj} + c)$$

and the nested error regression model

$$z_{dj} = \boldsymbol{x}_{dj}^T \mathbf{b} + u_d + e_{dj},$$

where $u_d \sim N(0, \sigma_u^2)$ and $e_{dj} \sim N(0, \sigma_e^2)$.

	estimate	standard error	p-value
$\hat{b}_0$	0.803	0.0201	$<$ 2E-16
$\hat{b}_1$	0.137	0.0135	$<$ 2E-16
$\hat{b}_2$	-0.112	0.0180	5.41E-10

Table 3: Parameter estimates under Model 3.

Application to real data

Log-linear normal mixed model (MODEL 3):

- Let us consider the *log* transformation of data

$$z_{dj} = \log(y_{dj} + c)$$

and the nested error regression model

$$z_{dj} = \boldsymbol{x}_{dj}^T \mathbf{b} + u_d + e_{dj},$$

where $u_d \sim N(0, \sigma_u^2)$ and $e_{dj} \sim N(0, \sigma_e^2)$.

	estimate	standard error	p-value
$\hat{b}_0$	0.803	0.0201	$<$ 2E-16
$\hat{b}_1$	0.137	0.0135	$<$ 2E-16
$\hat{b}_2$	-0.112	0.0180	5.41E-10

Table 3: Parameter estimates under Model 3.

Application to real data

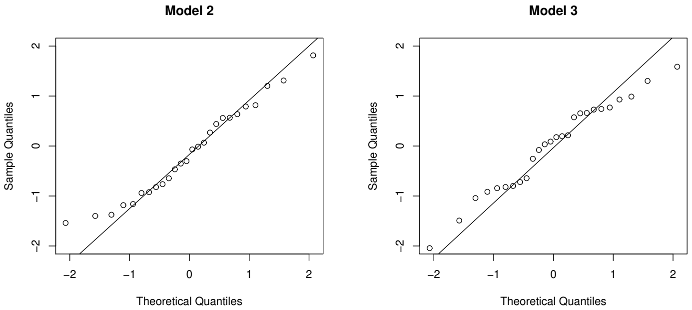


Figure 4: Q-Q plots of random effects for models 2 (left) and 3 (right).

Application to real data

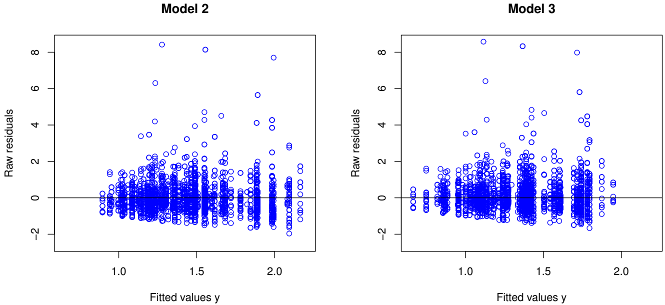


Figure 5: Dispersion graphs of raw residuals for Model 2 (left) and Model 3 (right).

The sum of squares of raw residuals for models 2 and 3 are

$$r_2^2 = 1897.35, \quad r_3^2 = 1938.30.$$

Application to real data - MODEL 2

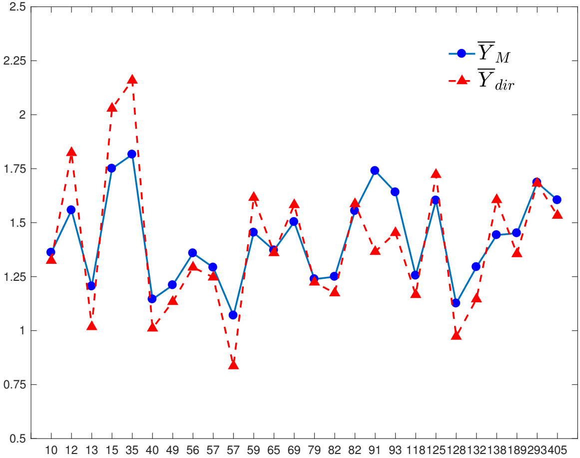


Figure 6: Predictions of average income in 10^4 euros .

Application to real data - MODEL 2

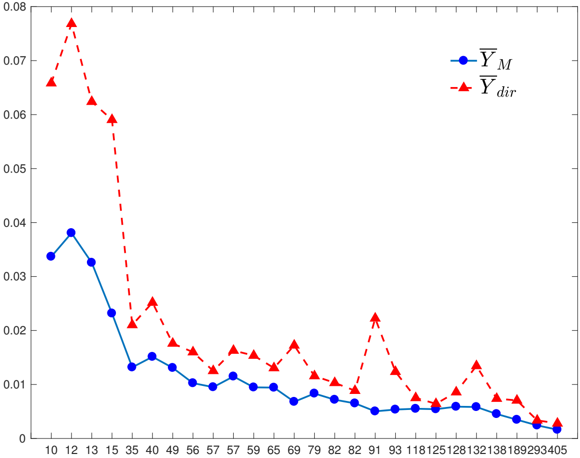


Figure 7: Estimated MSEs of average income estimates.
(based on $B = 500$ bootstrap samples)

Application to real data - MODEL 2

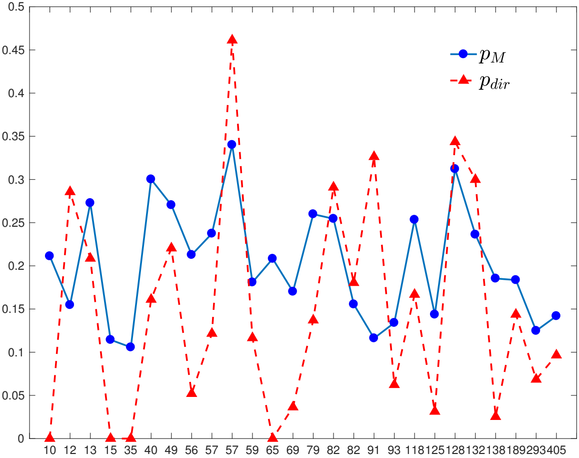


Figure 8: Marginal and Direct poverty proportions estimates.

Application to real data - MODEL 2

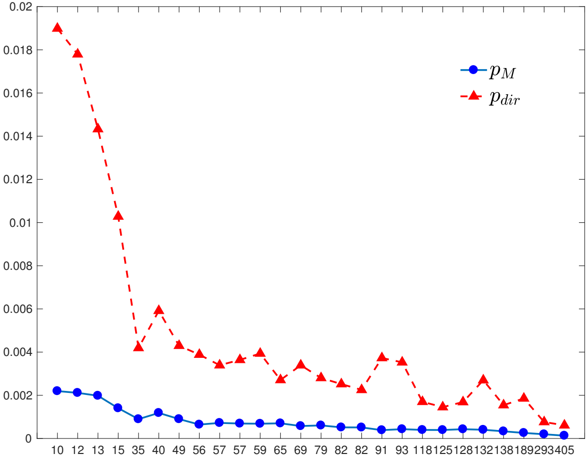


Figure 9: Estimated MSEs of poverty proportions estimates. (based on $B = 500$ bootstrap samples)

Conclusions:

- Model 2 has a high flexibility for fitting real data because a_{dj}'s may vary within and between domains.
- The EBP and marginal predictor have a similarly good behaviour. From computational reasons, the marginal predictor can be recommended.
- Marginal predictors can increase precision of the direct estimators.
- For the studied data sets, the Model 2 is a good alternative to the log-normal nested error model considered by Molina and Rao (2010).

Thank you for your attention!!!

References

Rao, J.N.K. and Molina, I. (2015). *Small Area Estimation*, 2nd Edition, New York: Wiley.

Molina, I. and Rao, J.N.K. (2010). Small area estimation of poverty indicators. *The Canadian Journal of Statistics*, 38, 369-385.

Hobza, T., Marhuenda, Y., Morales, D. (2020). Small area estimation of additive parameters under unit-level generalized linear mixed models. *SORT*, 44, 3-38.

Morales, D., Esteban, M.D., Pérez, A., Hobza, T. (2021). *A Course on Small Area Estimation and Mixed Models, Methods, Theory and Applications in R*, Springer, Cham.