Small Area Estimation under time models.

Domingo Morales González

d.morales@umh.es

SAMPLE\textsuperscript{\textcopyright} project (FP7-SSH-2007-1)

\textsuperscript{a}In collaboration with M.D. Esteban, A. Pérez and L. Santamaría

\textsuperscript{b}http://www.sample-project.eu/
1. The small area estimation problem.
2. Prediction Theory.
4. Poverty measures for small areas.
5. Area-level time models in SAMPLE project.
7. Simulation experiment 2.
8. Application to Spanish Living Condition Survey data.
10. References.
The small area estimation problem.

**Problem:** Estimate the mean \( \overline{Y} \) of a finite population \( U = \{1, \ldots, N \} \).

**Solution:** Extract a sample of size \( n \) and use a direct estimator \( \overline{y} \), where

- \( n \) is selected such that \( \frac{\sqrt{MSE(\overline{y})}}{\overline{Y}} \times 100 < 10 \) (for example), i.e.

\[
\begin{array}{ccc}
N, \overline{Y} & U \\
n, \overline{y} \\
\end{array}
\]

**Consequence:** \( n_d < n \) is not a priori fixed and \( \frac{\sqrt{MSE(\overline{y}_d)}}{\overline{Y}_d} \times 100 > 10 \)

\[
\begin{array}{ccc}
N_1, \overline{Y}_1 & U_1 & n_1, \overline{y}_1 \\
N_2, \overline{Y}_2 & U_2 & n_2, \overline{y}_2 \\
N_3, \overline{Y}_3 & U_3 & n_3, \overline{y}_3 \\
N_4, \overline{Y}_4 & U_4 & n_4, \overline{y}_4 \\
N_5, \overline{Y}_5 & U_5 & n_5, \overline{y}_5 \\
N_6, \overline{Y}_6 & U_6 & n_6, \overline{y}_6 \\
N_7, \overline{Y}_7 & U_7 & n_7, \overline{y}_7 \\
N_8, \overline{Y}_8 & U_8 & n_8, \overline{y}_8 \\
N_9, \overline{Y}_9 & U_9 & n_9, \overline{y}_9 \\
\end{array}
\]
The small area estimation problem.

**Solutions:**

- $n \uparrow$
- Change the sampling design, so that $U_d$ become a planned domain.
- Use models to include auxiliary variables $x$ and borrow strength
  - *Cross-sectional data*: Data from all the domains in $U$
  - *Time correlation*: Use also data from the past.
  - *Spatial correlation*: Use the correlation structure of data.

**Models:** there are no restrictions on the type of models to be used.

- LM, LMM, GLM, GLMM
- nonparametric, semiparametric models
- robust approaches, and so on

**SAE** is thus a part of the Statistical Science that combines

- Survey sampling and finite population inference
- Statistical models
**Question:** How these two branches of Statistics combines?

There are basically two approaches:

- **Unit-level models - The Prediction Theory**
- **Area-level models**

**Prediction Theory:**

- The values of the target variable in the \( N \) population units
  \( y = (y_1, \ldots, y_N)^t \) is the realization of a random vector
  \( Y = (Y_1, \ldots, Y_N)^t \).
- The distribution of \( Y \) is the only source of randomness.
- The distribution of the sampling design is not considered at all.
- The distribution of \( Y \) is given by a model

**Parameter** of interest is the realization of \( h(Y) \), i.e. \( h(y) \), where

- \( h \) is a linear function (mean, total, proportion, etc.), or
- \( h \) is a nonlinear function (variance, rate, poverty indicator, etc.).
We consider 3 cases

<table>
<thead>
<tr>
<th>Case</th>
<th>Parameter</th>
<th>Example</th>
<th>Model</th>
<th>Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Linear</td>
<td>$Y_d$</td>
<td>Linear (LMM)</td>
<td>EBLUP</td>
</tr>
<tr>
<td>2</td>
<td>Linear</td>
<td>$p_d$</td>
<td>Non Linear (GLMM)</td>
<td>BP, G-EBLUP</td>
</tr>
<tr>
<td>3</td>
<td>Non linear</td>
<td>Poverty Gap</td>
<td>any</td>
<td>open</td>
</tr>
</tbody>
</table>

In what follows we briefly describe Case 1.

We are interested in estimating a linear combination of $y$’s,

$$a^t y,$$

where $a = (a_1, \ldots, a_N)^t$ is a vector with $N$ constants.

For example,

- $a_i = 1 \forall i$, then $a^t y = \sum_{i=1}^{N} y_i$ is the population total,
- $a_i = 1/N \forall i$, then $a^t y = \frac{1}{N} \sum_{i=1}^{N} y_i$ is the population mean.
Definition 1.1. A **linear estimator** of $\theta = a^t Y$ is $\hat{\theta} = g_s^t Y_s$, where $g_s = (g_1, \ldots, g_n)^t$ is a vector with $n$ coefficients.

Definition 1.2. The **estimation error** of an estimator $\hat{\theta} = g_s^t Y_s$ is $\hat{\theta} - \theta = g_s^t Y_s - a^t Y$.

- We are interested in the prediction problem under a linear model $M$:

$$E_M[Y] = X\beta, \quad var_M[Y] = V,$$

where $X_{N \times p}$ is the matriz of auxiliary variables, $\beta_{p \times 1}$ is the vector of unknown parameters and $V$ is a definite positive covariance matrix.

- We suppose that values of auxiliary variables, in all the units of the population, are known; i.e. $X_{N \times p}$ is known.
Definition 1.3. The estimator $\hat{\theta}$ is prediction unbiased for $\theta$, under the model $M$, if $E_M[\hat{\theta} - \theta] = 0$. Note that expression $E_M[\hat{\theta}] = \theta$ is not correct because $\theta$ is random.

Definition 1.4. The error variance of $\hat{\theta}$, under $M$, is $E_M[(\hat{\theta} - \theta)^2]$.

Rearranging the population elements one can write

$$X = \begin{bmatrix} X_s \\ X_r \end{bmatrix}, \quad V = \begin{bmatrix} V_{ss} & V_{sr} \\ V_{rs} & V_{rr} \end{bmatrix},$$

where $X_s$ is $n \times p$, $X_r$ es $(N - n) \times p$, $V_{ss}$ is $n \times n$, $V_{rr}$ es $(N - n) \times (N - n)$, $V_{sr}$ is $n \times (N - n)$ and $V_{rs} = V_{sr}^t$. We assume that $V_{ss}$ is positively definite.
Theorem 1.1. (General prediction Theorem).
Among linear, prediction-unbiased estimators $\hat{\theta}$ of $\theta$, the error variance is minimized by

$$
\hat{\theta}_{opt} = a_s^t Y_s + a_r^t \left[ X_r \hat{\beta} + V_{rs} V_{ss}^{-1} (Y_s - X_s \hat{\beta}) \right], 
$$

(1)

where $\hat{\beta} = (X_s^t V_{ss} X_s)^{-1} X_s^t V_{ss}^{-1} Y_s$. The error variance of $\hat{\theta}_{opt}$ is

$$
V_M[\hat{\theta}_{opt} - \theta] = a_r^t (V_{rr} - V_{rs} V_{ss}^{-1} V_{sr}) a_r 
+ a_r^t (X_r - V_{rs} V_{ss}^{-1} X_s)(X_s^t V_{ss}^{-1} X_s)^{-1} (X_r - V_{rs} V_{ss}^{-1} X_s)^t a_r
$$

$\hat{\theta}_{opt}$ is called BLU predictor, or BLUP (best linear unbiased predictor).
Let us consider the two-fold nested error regression model

\[ Y_{dtj} = x_{dtj} \beta + u_{1,d} + u_{2,dt} + w_{dtj}^{-1/2} e_{dtj}, \ d = 1, \ldots, D, \ t = 1, \ldots, T, \ j = 1, \ldots, N_{td}, \]

where

- \( Y_{dtj} \sim \) target variable in domain \( d \), time \( t \) and unit \( j \).
- \( x_{dtj} = (x_{1,dtj}, \ldots, x_{p,dtj}) \sim \) vector of auxiliary variables in \( (d, t, j) \).
- \( \beta = (\beta_1, \ldots, \beta_p)' \sim \) vector of regression parameters.
- The \( u_{1,d} \)'s are i.i.d. \( N(0, \sigma^2_1) \).
- The vectors \( (u_{2,d1}, \ldots, u_{2,dT}), \ d = 1, \ldots, D, \) are i.i.d. AR(1) with variance and auto-correlation parameters \( \sigma^2_2 \) and \( \rho \) respectively.
- The \( e_{dtj} \)'s are i.i.d. \( N(0, \sigma^2_0) \)
- The \( u_{1,d} \)'s, the \( u_{2,dt} \)'s and the \( e_{dtj} \)'s are independent.
Logistic model with correlated time effects

Let $u_{1,d}$ and $u_{2,dt}$ be the random effects for area $d$ and time instant $t$ (within area $d$).

Let $\mathbf{u}_1 = \text{col}_{1 \leq d \leq D} (u_{1,d})$, $\mathbf{u}_2 = \text{col}_{1 \leq d \leq D} (u_{2,d})$, $\mathbf{u}_{2,d} = \text{col}_{1 \leq t \leq T} (u_{2,dt})$ be such that

- $\mathbf{u}_1 \sim N(\mathbf{0}, \varphi_1 \mathbf{I}_D)$ and
- $\mathbf{u}_2 \sim N(\mathbf{0}, \varphi_2 \Omega(\rho))$ are independent with $\Omega(\rho) = \text{diag} \left( \Omega_d \right)$ and

$$
\Omega_d = \Omega_d(\rho) = \frac{1}{1 - \rho^2} \begin{pmatrix}
1 & \rho & \cdots & \rho^{T-2} & \rho^{T-1} \\
\rho & 1 & \ddots & \rho^{T-2} & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\rho^{T-2} & \ddots & 1 & \rho \\
\rho^{T-1} & \rho^{T-2} & \cdots & \rho & 1 \\
\end{pmatrix}_{T \times T}
$$
Logistic model with correlated time effects (continuation)

- \( y_{dtj} \), conditioned to \( u = (u'_1, u'_2)' \), are independent with distributions
  \[ y_{dtj} | u_{1,d}, u_{2,dt} \sim \text{Bin}(\nu_{dtj}, p_{dtj}), \quad d = 1, \ldots, D, \quad t = 1, \ldots, T, \quad j = 1, \ldots, N_{td}, \]

where \( \sum_{t=1}^{m_d} n_{dt} = n_d, \sum_{d=1}^{D} n_d = n \) and \( \sum_{d=1}^{D} m_d = M \).

- For the natural parameter \( \eta_{dtj} = \log \frac{p_{dtj}}{1-p_{dtj}} \) we assume the model
  \[ \eta_{dtj} = x_{dtj} \beta + u_{1,d} + u_{2,dt}, \quad d = 1, \ldots, D, \quad t = 1, \ldots, T, \quad j = 1, \ldots, N_{td}, \quad (2) \]
Let $E_{dj}$ be a quantitative measure of welfare (for example, income).

Let $z$ be the poverty line; that is, the threshold for $E_{dj}$ under which a person is considered as “under poverty”.

Foster, Greer and Thorbecke (1984) defined the poverty measures

$$F_{\alpha d} = \frac{1}{N_d} \sum_{j=1}^{N_d} \left( \frac{z - E_{dj}}{z} \right)^\alpha I(E_{dj} < z), \quad \alpha = 0, 1, 2, \quad d = 1, \ldots, D,$$

where

- $I(E_{dj} < z) = 1$ if $E_{dj} < z$ (person under poverty)
- $I(E_{dj} < z) = 0$ if $E_{dj} \geq z$ (person not under poverty).

$\alpha = 0$ gives the proportion of individuals under poverty.

$\alpha = 1$ gives the poverty gap,

$\alpha = 2$ gives the poverty severity.
Direct estimators of FGT poverty measures:

- Let us consider a sampling design with inclusion probabilities and weights

\[ \pi_{dj} = P(\text{Unit } j \text{ of domain } d \text{ is in sample } s), \quad w_{dj} = 1/\pi_{dj}. \]

- Direct estimators of \( F_{\alpha d} \) is

\[
 f_{\alpha d}^w = \frac{1}{\hat{N}_d} \sum_{j \in s_d} w_{dj} \left( \frac{z - E_{dj}}{z} \right)^{\alpha} I(E_{dj} < z), \quad \alpha = 0, 1, 2, \ d = 1, \ldots, D,
\]

where \( \hat{N}_d = \sum_{j \in s_d} w_{dj} \) is the direct estimator of the population size of small area \( d, N_d \).

\[ \Rightarrow \] As sample sizes \( n_d \) are usually too small, in what follows we will present an area-level model approach to estimate \( F_{\alpha d} \).
Let us consider the model

\[ y_{dt} = \mathbf{x}_{dt} \beta + u_{dt} + e_{dt}, \quad d = 1, \ldots, D, \quad t = 1, \ldots, m_d, \quad (3) \]

where

- \( y_{dt} \) is a direct estimator of the indicator of interest for area \( d \) and time instant \( t \),
- \( \mathbf{x}_{dt} \) is a vector containing the aggregated (population) values of \( p \) auxiliary variables,
- the random vectors \( (u_{d1}, \ldots, u_{dm_d}) \), \( d = 1, \ldots, D \), are i.i.d. AR(1), with variance and auto-correlation parameters \( \sigma_u^2 \) and \( \rho \) respectively,
- the errors \( e_{dtj} \)'s are independent \( N(0, \sigma_{dt}^2) \) with known \( \sigma_{dt}^2 \)'s, and
- the \( u_{dt} \)'s and the \( e_{dt} \)'s are independent.
The model (3) can be written alternatively in the form

\[ y = X\beta + Zu + e, \]  

(4)

where

- \( y = \operatorname{col}_{1 \leq d \leq D} (y_d), \quad y_d = \operatorname{col}_{1 \leq t \leq m_d} (y_{dt}), \)

- \( u = \operatorname{col}_{1 \leq d \leq D} (u_d), \quad u_d = \operatorname{col}_{1 \leq t \leq m_d} (u_{dt}), \)

- \( e = \operatorname{col}_{1 \leq d \leq D} (e_d), \quad e_d = \operatorname{col}_{1 \leq t \leq m_d} (e_{dt}), \)

- \( X = \operatorname{col}_{1 \leq d \leq D} (X_d), \quad X_d = \operatorname{col}_{1 \leq t \leq m_d} (x_{dt}), \quad x_{dt} = \operatorname{col}'_{1 \leq j \leq p} (x_{dtj}), \)

- \( \beta = \operatorname{col}_{1 \leq j \leq p} (\beta_j), \quad Z = I_{M \times M}, \quad M = \sum_{d=1}^{D} m_d. \)
We assume that

\( u \sim N(0, \Sigma_u) \) and \( e \sim N(0, \Sigma_e) \) are independent,

with covariance matrices

\[
\Sigma_u = \sigma^2_u \Omega(\rho), \quad \Omega(\rho) = \text{diag} \left( \Omega_d(\rho) \right), \quad 1 \leq d \leq D
\]

\[
\Sigma_e = \text{diag} \left( \Sigma_{ed} \right), \quad \Sigma_{ed} = \text{diag} \left( \sigma^2_{dt} \right), \quad 1 \leq t \leq m_d
\]

where the \( \sigma^2_{dt} \) are known, and

\[
\Omega_d = \Omega_d(\rho) = \frac{1}{1 - \rho^2} \left( \begin{array}{ccccccc}
1 & \rho & \cdots & \rho^{m_d-2} & \rho^{m_d-1} \\
\rho & 1 & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
\rho^{m_d-2} & \ddots & 1 & \rho \\
\rho^{m_d-1} & \rho^{m_d-2} & \cdots & \rho & 1
\end{array} \right)_{m_d \times m_d}
\]
The BLU estimator and predictor of $\beta$ and $u$ are

$$\hat{\beta} = \left( X^t V^{-1} X \right)^{-1} X^t V^{-1} y \quad \text{and} \quad \hat{u} = \sum_u Z^t V^{-1} (y - X\hat{\beta}),$$

where

$$\text{var}(y) = V = \sigma_u^2 \text{diag} \left( \Omega_d(\rho) \right) + \Sigma_e = \text{diag} \left( \sigma_u^2 \Omega_d(\rho) + \Sigma_{ed} \right) = \text{diag} \left( V_d \right).$$

To calculate $\hat{\beta}$ y $\hat{u}$ the following formulas are applied

$$\hat{\beta} = \left( \sum_{d=1}^{D} X_d^t V_d^{-1} X_d \right)^{-1} \left( \sum_{d=1}^{D} X_d^t V_d^{-1} y_d \right),$$

$$\hat{u} = \sigma_u^2 \text{col} \left( \Omega_d(\rho) V_d^{-1} (y_d - X_d\hat{\beta}) \right).$$
REML loglikelihood is

\[ l_{REML}(\sigma_u^2, \rho) = -\frac{M - p}{2} \log 2\pi + \frac{1}{2} \log |X^t X| - \frac{1}{2} \log |V| - \frac{1}{2} \log |X^t V^{-1} X| - \frac{1}{2} y^t P y, \]

where

\[ P = V^{-1} - V^{-1} X (X^t V^{-1} X)^{-1} X^t V^{-1}, \quad PVP = P, \quad PX = 0. \]

To maximize \( l_{REML}(\sigma_u^2, \rho) \), we first define

\[ \theta = (\theta_1, \theta_2) = (\sigma_u^2, \rho), \]

\[ V_1 = \frac{\partial V}{\partial \sigma_u^2} = \text{diag } (\Omega_d(\rho)), \quad V_2 = \frac{\partial V}{\partial \rho} = \sigma_u^2 \text{ diag } (\Omega'_d(\rho)), \]

Then

\[ P_a = \frac{\partial P}{\partial \theta_a} = -P \frac{\partial V}{\partial \theta_a} P = -PV_a P, \quad a = 1, 2. \]
By taking partial derivatives we get the scores

\[ S_a = \frac{\partial l_{REML}}{\partial \theta_a} = -\frac{1}{2} \text{tr}(PV_a) + \frac{1}{2} y^t PV_a Py, \quad a = 1, 2. \]

By taking partial derivatives and expectations and changing the sign, we get the Fisher information components

\[ F_{ab} = \frac{1}{2} \text{tr}(PV_a PV_b), \quad a, b = 1, 2. \]

To maximize \( l_{REML} \), Fisher-scoring updating formula is

\[ \theta^{k+1} = \theta^k + F^{-1}(\theta^k)(\theta^k). \]

As seeds we use \( \rho^{(0)} = 0 \), and \( \sigma^2_u^{(0)} = \hat{\sigma}^2_{uH} \),

where \( \hat{\sigma}^2_{uH} \) is the Henderson 3 estimator under model with \( \rho = 0 \).

The REML estimators of \( \beta \) is

\[ \hat{\beta}_{REML} = (X^t \hat{V}^{-1} X)^{-1} X^t \hat{V}^{-1} y. \]
EBLUP of $\mu_{dt} = x_{dt}\beta + u_{dt}$ is $\hat{\mu}_{dt} = x_{dt}\hat{\beta} + \hat{u}_{dt}$.

$\bar{Y}_{dt}$ is estimated by $\hat{Y}_{dt}^{eblup} = \hat{\mu}_{dt}$.

The MSE of $\hat{Y}_{dt}^{eblup}$ is $MSE(\hat{Y}_{dt}^{eblup}) = g_1(\theta) + g_2(\theta) + g_3(\theta)$, where $\theta = (\sigma_u^2, \rho)$,

$$g_1(\theta) = a^tZTZ^ta,$$

$$g_2(\theta) = [a^tX - a^tZTZ^t\Sigma^{-1}_eX]Q[X^ta - X^t\Sigma^{-1}_eZTZ^ta],$$

$$g_3(\theta) \approx tr \left\{ (\nabla b^t)V(\nabla b^t)^t E \left[ (\hat{\theta} - \theta)(\hat{\theta} - \theta)^t \right] \right\}$$

$$Q = (X^tV^{-1}X)^{-1}, \ T = \Sigma_u - \Sigma_uZ^tV^{-1}Z\Sigma_u$$

An estimator of $MSE(\hat{Y}_{dt}^{eblup})$ is

$$mse(\hat{Y}_{dt}^{eblup}) = g_1(\hat{\theta}) + g_2(\hat{\theta}) + 2g_3(\hat{\theta})$$. 
For $d = 1, \ldots, D$, $t = 1, \ldots, m_d$, the explicative and target variables are

$$x_{dt} = (b_{dt} - a_{dt})U_{dt} + a_{dt}, \quad U_{dt} = \frac{t}{m_d + 1}$$

$$b_{dt} = 1 + \frac{1}{D}(m_d(d - 1) + t), \quad a_{dt} = 1,$$

$$y_{dt} = \beta_1 + \beta_2 x_{dt} + u_{dt} + e_{dt}, \quad \beta_1 = 0, \beta_2 = 1,$$

where $e_{dt} \sim N(0, \sigma^2_{dt})$,

$$\sigma^2_{dt} = \frac{(\alpha_1 - \alpha_0)(m_d(d - 1) + t - 1)}{M - 1} + \alpha_0, \quad \alpha_0 = 0.8, \alpha_1 = 1.2.$$

The random effects $u_{dt}$ are generated from the AR(1) process

$$u_{d1} = (1 - \rho^2)^{-1/2} \varepsilon_{d1}, \quad u_{dt} = \rho u_{dt-1} + \varepsilon_{dt}, \quad t = 2, \ldots, m_d,$$

where $\varepsilon_{dt} \sim N(0, \sigma^2_u), \quad d = 1, \ldots, D, \quad t = 1, \ldots, m_d.$
Simulation experiment 1

The simulation experiment has the following steps:

1. Repeat $K = 10^4$ times ($k = 1, \ldots, K$)
   1.1. Generate a sample of size $m = \sum_{d=1}^{D} m_d$. Calculate
   
   $$\mu_{dt}^{(k)} = \beta_1 + \beta_2 x_{dt} + u_{dt}^{(k)}.$$ 
   
   1.2. Calculate $\hat{\beta}_1^{(k,0)}$, $\hat{\beta}_2^{(k,0)}$, $\hat{\sigma}_u^{2(k,0)}$ and $\hat{\mu}_{dt}^{(k,0)}$ by using REML method under model (3) restricted to $\rho = 0$.
   
   1.3. Calculate $\hat{\beta}_1^{(k,1)}$, $\hat{\beta}_2^{(k,1)}$, $\hat{\sigma}_u^{2(k,1)}$, $\hat{\rho}^{(k,1)}$ and $\hat{\mu}_{dt}^{(k,1)}$ by using REML method under model (3).
2. For \( d = 1, \ldots, D \), \( t = 1, \ldots, m_d \), calculate

\[
BIAS_{dt}^{(0)} = \frac{1}{K} \sum_{k=1}^{K} \left( \hat{\mu}_{dt}^{(k,0)} - \mu_{dt}^{(k)} \right), \quad MSE_{dt}^{(0)} = \frac{1}{K} \sum_{k=1}^{K} (\hat{\mu}_{dt}^{(k,0)} - \mu_{dt}^{(k)})^2,
\]

\[
BIAS_{dt}^{(1)} = \frac{1}{K} \sum_{k=1}^{K} \left( \hat{\mu}_{dt}^{(k,1)} - \mu_{dt}^{(k)} \right), \quad MSE_{dt}^{(1)} = \frac{1}{K} \sum_{k=1}^{K} (\hat{\mu}_{dt}^{(k,1)} - \mu_{dt}^{(k)})^2,
\]

\[
BIAS^{(0)} = \frac{1}{D} \sum_{d=1}^{D} \sum_{i=1}^{m_d} BIAS_{dt}^{(0)}, \quad MSE^{(0)} = \frac{1}{D} \sum_{d=1}^{D} \sum_{i=1}^{m_d} MSE_{dt}^{(0)}.
\]

\[
BIAS^{(1)} = \frac{1}{D} \sum_{d=1}^{D} \sum_{i=1}^{m_d} BIAS_{dt}^{(1)}, \quad MSE^{(1)} = \frac{1}{D} \sum_{d=1}^{D} \sum_{i=1}^{m_d} MSE_{dt}^{(1)}.
\]
### Table 1. MSE’s. for $D = 100$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>REML</th>
<th>$m = 200$</th>
<th>$m = 500$</th>
<th>$m = 1000$</th>
<th>$m = 2000$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$m_d = 2$</td>
<td>$m_d = 5$</td>
<td>$m_d = 10$</td>
<td>$m_d = 20$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.5086</td>
<td>0.5026</td>
<td>0.5003</td>
<td>0.4996</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.5138</td>
<td>0.5046</td>
<td>0.5014</td>
<td>0.5001</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>0.5263</td>
<td>0.5204</td>
<td>0.5185</td>
<td>0.5176</td>
</tr>
<tr>
<td>0.25</td>
<td>1</td>
<td>0.5214</td>
<td>0.5074</td>
<td>0.5026</td>
<td>0.5007</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0.6263</td>
<td>0.6189</td>
<td>0.6183</td>
<td>0.6193</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.5457</td>
<td>0.5133</td>
<td>0.5052</td>
<td>0.5015</td>
</tr>
<tr>
<td>0.75</td>
<td>0</td>
<td>1.2021</td>
<td>1.1903</td>
<td>1.1930</td>
<td>1.1971</td>
</tr>
<tr>
<td>0.75</td>
<td>1</td>
<td>0.5953</td>
<td>0.5230</td>
<td>0.5029</td>
<td>0.4935</td>
</tr>
</tbody>
</table>
### Simulation experiment 1

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>REML</th>
<th>$m = 200$</th>
<th>$m = 500$</th>
<th>$m = 1000$</th>
<th>$m = 2000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.000782</td>
<td>-0.000108</td>
<td>0.000529</td>
<td>-0.000014</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.000780</td>
<td>-0.000108</td>
<td>0.000529</td>
<td>-0.000014</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>0.000787</td>
<td>-0.000114</td>
<td>0.000534</td>
<td>-0.000016</td>
</tr>
<tr>
<td>0.25</td>
<td>1</td>
<td>0.000785</td>
<td>-0.000107</td>
<td>0.000535</td>
<td>-0.000012</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>-0.000199</td>
<td>-0.001330</td>
<td>0.001962</td>
<td>0.001029</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>-0.000218</td>
<td>-0.001317</td>
<td>0.001927</td>
<td>0.001043</td>
</tr>
<tr>
<td>0.75</td>
<td>0</td>
<td>-0.000304</td>
<td>-0.001301</td>
<td>0.001973</td>
<td>0.001060</td>
</tr>
<tr>
<td>0.75</td>
<td>1</td>
<td>-0.000325</td>
<td>-0.001289</td>
<td>0.001916</td>
<td>0.001057</td>
</tr>
</tbody>
</table>

**Table 2.** BIAS’s. for $D = 100$
Figure 1. $MSE_{d_{m_d}}$'s for $D = 100$, $m_d = 2$ and $\rho = 0$. 

\[ MSE_d^{(0)}, MSE_d^{(1)}, \text{ for } \rho = 0, \ m_d = 2 \]
Figure 2. $MSE_{dm_d}$’s for $D = 100, m_d = 2$ and $\rho = 0.25$. 

$MSE_d^{(0)}, MSE_d^{(1)}$, for $\rho=0.25, m_d=2$
Figure 3. $MSE_{d_{m_{d}}}$'s for $D = 100$, $m_{d} = 2$ and $\rho = 0.5$. 
Figure 4. $MSE_{d_{m_d}}$'s for $D = 100$, $m_d = 2$ and $\rho = 0.75$. 
BIAS_d^{(0)}, BIAS_d^{(1)}, in absolute value for $\rho=0$, $m_d=5$

**Figure 5.** $BIAS_{d_{m_d}}$'s for $D = 100$, $m_d = 2$ and $\rho = 0$. 
Figure 6. $BIA S_{dm_d}$’s for $D = 100$, $m_d = 2$ and $\rho = 0.75$. 
The second simulation experiment takes the MSEs obtained in the first experiment and includes the following additional steps:

1.4 Calculate $mse(\hat{\mu}_{di}^{(k,0)})$ and $mse(\hat{\mu}_{di}^{(k,1)})$.

3. For $d = 1, \ldots, D$, $i = 1, \ldots, m_d$, calculate

$$B_{di}^{(a)} = \frac{1}{K} \sum_{k=1}^{K} (mse(\hat{\mu}_{di}^{(k,a)}) - MSE_{di}^{(a)}) , \ a = 0, 1,$$

$$E_{di}^{(a)} = \frac{1}{K} \sum_{k=1}^{K} (mse(\hat{\mu}_{di}^{(k,a)}) - MSE_{di}^{(a)})^2 , \ a = 0, 1,$$

$$B^{(a)} = \frac{1}{D} \sum_{d=1}^{D} \sum_{i=1}^{m_d} B_{di}^{(a)} , \ E^{(a)} = \frac{1}{D} \sum_{d=1}^{D} \sum_{i=1}^{m_d} E_{di}^{(a)} , \ a = 0, 1.$$
<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\alpha$</th>
<th>$m_d$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.00347</td>
<td>0.00194</td>
<td>0.00140</td>
<td>0.00112</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.00350</td>
<td>0.00194</td>
<td>0.00140</td>
<td>0.00112</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>0.00350</td>
<td>0.00202</td>
<td>0.00150</td>
<td>0.00122</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>1</td>
<td>0.00352</td>
<td>0.00203</td>
<td>0.00146</td>
<td>0.00118</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0.00365</td>
<td>0.00242</td>
<td>0.00195</td>
<td>0.00168</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>0.00398</td>
<td>0.00222</td>
<td>0.00161</td>
<td>0.00132</td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>0</td>
<td>0.00465</td>
<td>0.00395</td>
<td>0.00361</td>
<td>0.00336</td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>1</td>
<td>0.00513</td>
<td>0.00243</td>
<td>0.00173</td>
<td>0.00141</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. $E$’s of EBLUP0 and EBLUP1 for $D = 100$
Table 4. $B$’s of EBLUP0 and EBLUP1 for $D = 100$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\alpha$</th>
<th>$m_2$</th>
<th>$m_5$</th>
<th>$m_{10}$</th>
<th>$m_{20}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>-0.00118</td>
<td>-0.00015</td>
<td>0.00014</td>
<td>-0.00038</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>-0.00086</td>
<td>-0.00018</td>
<td>0.00013</td>
<td>-0.00038</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>-0.00118</td>
<td>-0.00006</td>
<td>-0.00007</td>
<td>-0.00023</td>
</tr>
<tr>
<td>0.25</td>
<td>1</td>
<td>-0.00116</td>
<td>-0.00047</td>
<td>-0.00007</td>
<td>-0.00059</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>-0.00139</td>
<td>-0.00028</td>
<td>-0.00052</td>
<td>-0.00030</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>-0.00198</td>
<td>-0.00109</td>
<td>-0.00073</td>
<td>-0.00113</td>
</tr>
<tr>
<td>0.75</td>
<td>0</td>
<td>-0.00307</td>
<td>-0.00209</td>
<td>-0.00232</td>
<td>-0.00190</td>
</tr>
<tr>
<td>0.75</td>
<td>1</td>
<td>1 -0.00405</td>
<td>-0.00225</td>
<td>-0.00165</td>
<td>-0.00162</td>
</tr>
</tbody>
</table>
$E_d^{(0)}$, $E_d^{(1)}$, for $\rho=0$, $m_d=5$

**Figure 7.** $E_{d_{m_d}}$'s for $D = 100$, $m_d = 2$ and $\rho = 0$. 
Figure 8. $E_{d m_d}$'s for $D = 100$, $m_d = 2$ and $\rho = 0.75$. 
Simulation experiment 2

$B_d^{(0)}$, $B_d^{(1)}$, for $\rho=0$, $m_d=5$

Figure 9. $B_{d,m_d}$’s for $D = 100$, $m_d = 2$ and $\rho = 0$. 
Figure 10. $B_{dm_d}$’s for $D = 100$, $m_d = 2$ and $\rho = 0.75$. 
The goal is to estimate the poverty incidence (proportion of individuals under poverty) and the poverty gap in Spanish domains.

We use data from the Spanish Living Conditions Survey (SLCS) of years 2004-2006 with sample sizes 44648, 37491, 34694 respectively.

The SLCS is the Spanish version of the “European Statistics on Income and Living Conditions” (EU-SILC).

The SLCS started in 2004 with an annual periodicity.

The EU-SILC provides comparative statistics on the distribution of income and social exclusion in the European environment.

The SLCS is an annual survey with a rotating panel design with a sample formed by four independent subsamples, each of which is a four-year panel.

Each year the sample is renewed in one of the panels.
In order to select each subsample, a two-stage design is implemented independently in each Autonomous Community with first stage unit stratification.

The first stage is formed by census sections grouped into strata in agreement with the size of the municipality to which they belong.

The second stage is formed by main family dwellings.

Within these no sub-sampling is carried out, investigating all dwellings that are their usual residence.

The sample includes 16000 dwellings distributed in 2000 census sections.

We consider $D = 104$ domains obtained by crossing 52 provinces with 2 sexes.

The quantiles of the distribution of the domain sample sizes are $q_0 = 17$, $q_1 = 170$, $q_2 = 293$, $q_3 = 640$ and $q_4 = 2113$ in 2004, 13, 149, 251, 530, 1494 in 2005 and 18, 129, 233, 481, 1494 in 2006, so they are too small to employ direct estimators to estimate the parameters of interest in all the domains.
The SLCS does not produce official estimates at the domain level (provinces × sex), but the analogous direct estimator is

\[ \hat{Y}_{d}^{\text{dir}} = \sum_{j \in S_{d}} w_{j} y_{j}. \]

where the \( w_{j} \)'s are the official calibrated sampling weights which take into account for non response.

- In the particular case \( y_{j} = 1 \), for all \( j \in P_{d} \), we get the estimated domain size

\[ \hat{N}_{d}^{\text{dir}} = \sum_{j \in S_{d}} w_{j}. \]

- Using this quantity, a direct estimator of the domain mean \( \bar{Y}_{d} \) is

\[ \bar{y}_{d} = \frac{\hat{Y}_{d}^{\text{dir}}}{\hat{N}_{d}^{\text{dir}}}. \]
The direct estimates of the domain means are used as responses in the area-level time model.

The design-based variances of these estimators can be approximated as

$$\hat{V}_\pi(\hat{Y}^{dir}_d) = \sum_{j \in S_d} w_j (w_j - 1) (y_j - \bar{y}_d)^2 \quad \text{and} \quad \hat{V}_\pi(\bar{y}_d) = \hat{V} (\hat{Y}^{dir}_d) / \hat{N}_d^2.$$  \hfill (5)

The last formulas are obtained from Särndal (1992), pp. 43, 185 and 391 with the simplifications $w_j = 1/\pi_j$, $\pi_{jj} = \pi_j$ and $\pi_{ij} = \pi_i \pi_j$, $i \neq j$ in the second order inclusion probabilities.

The target variables are the direct estimates of the poverty incidence and poverty gap at domain level.
The considered auxiliary variables are:

- **AGE**: Age groups are 1-5 for the intervals $\leq 15$, $16 - 24$, $25 - 49$, $50 - 64$ and $\geq 65$.

- **EDUCATION**: Highest level of education completed, with 4 categories taking the values 1 for *Less than primary education level*, 2 for *Primary education level*, 3 for *Secondary education level* and 4 for *University level*.

- **CITIZENSHIP**: with 2 categories taking the values 1 for *Spanish* and 2 for *Not Spanish*.

- **LABOR**: Labor situation with 4 categories taking the values 0 for *Below 16 years*, 2 for *employed*, 3 for *unemployed* and 4 for *inactive*.
The Poverty Threshold is fixed as the 60% of the median of the normalized incomes in Spanish households.

The aim of normalizing the household income is to adjust for the varying size and composition of households.

The total number of *normalized household members* is

$$H_{dti} = 1 + 0.5(H_{dti\geq14} - 1) + 0.3H_{dti<14}$$

where $H_{dti\geq14}$ is the number of people aged 14 and over and $H_{dti<14}$ is the number of children aged under 14.

The normalized net annual income of a household is obtained by dividing its net annual income by its normalized size.

The Spanish poverty thresholds (in euros) in 2004-06 are $z_{2004} = 6098.57$, $z_{2005} = 6160.00$ and $z_{2005} = 6556.60$ respectively.

These are the $z$-values used in the calculation of the direct estimates of the poverty incidence and gap.
We consider the linear model

\[ \overline{y}_{dt} = \overline{X}_{dt}\beta + u_{dt} + e_{dt}, \quad d = 1, \ldots, D \]

where \( \overline{X}_d \) is the \( 1 \times p \) vector containing the population (aggregated) mean values of all the categories (except the last one) of the explanatory variables.

The first position of \( \overline{X}_d \) contains a “1”, so that

\[ p = 1 + 4 + 3 + 1 + 3 = 12. \]

Random effects errors are assumed to follow the distributional assumption model CORRELATION and/or of model INDEPENDENT.
Final selected models contain the following auxiliary variables:

<table>
<thead>
<tr>
<th>$\alpha = 0$</th>
<th>constant</th>
<th>age3</th>
<th>age4</th>
<th>edu1</th>
<th>edu2</th>
<th>cit1</th>
<th>lab2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.544</td>
<td>-0.582</td>
<td>-2.607</td>
<td>0.089</td>
<td>-0.056</td>
<td>0.331</td>
<td>0.204</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha = 1$</th>
<th>constant</th>
<th>edu0</th>
<th>edu1</th>
<th>edu2</th>
<th>cit1</th>
<th>lab1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.2613</td>
<td>0.7767</td>
<td>0.2184</td>
<td>0.1286</td>
<td>0.1165</td>
<td>-0.0538</td>
</tr>
</tbody>
</table>

Table 5. Auxiliary variables for selected models

By observing the signs of the regression parameters in model 1 for $\alpha = 0$, we interpret that poverty proportion tends to be smaller in those domains with larger proportion of population in the subset defined by:

- age in the interval 25-64 (age interval with greater incomes),
- education in the category of secondary studies completed, and
- non Spanish citizenship (may be because immigrants tends to go to regions with greater richness where it is easier to find job),

and with lower proportion of population with

- unemployed people.
By doing the same exercise with the signs of the regression parameter in model 1 for $\alpha = 1$, we interpret that poverty gap tends to be smaller in those domains with larger proportion of population characterized by

- university education completed,
- non Spanish citizenship, and
- being employed.

Model 1 is preferred because

- Correlation estimates are $\hat{\rho} = 0.846$ and $\hat{\rho} = 0.683$ for the poverty proportion and gap respectively.

In addition

- Differences $\bar{y}_{dt} - \hat{\mu}_{dt}$ are plotted against $\bar{y}_{dt}$ in Figures 11 and 12.
- Basic results are presented in Table 6.
Figure 11. $\bar{y}_{dt} - \hat{\mu}_{dt}$ versus $\bar{y}_{dt}$ for the poverty proportion.
Figure 12. $\bar{y}_{dt} - \hat{\mu}_{dt}$ versus $\bar{y}_{dt}$ for the poverty gap.
Table 6. Direct and EBLUP estimates of poverty proportions (0) and gaps (1).

<table>
<thead>
<tr>
<th></th>
<th>dir₀</th>
<th>eblup₀</th>
<th>dir₁</th>
<th>eblup₁</th>
<th>dir₀</th>
<th>eblup₀</th>
<th>dir₁</th>
<th>eblup₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.083</td>
<td>0.064</td>
<td>0.025</td>
<td>0.022</td>
<td>0.079</td>
<td>0.082</td>
<td>0.015</td>
<td>0.016</td>
</tr>
<tr>
<td>2</td>
<td>0.237</td>
<td>0.251</td>
<td>0.096</td>
<td>0.091</td>
<td>0.285</td>
<td>0.286</td>
<td>0.117</td>
<td>0.106</td>
</tr>
<tr>
<td>3</td>
<td>0.160</td>
<td>0.163</td>
<td>0.050</td>
<td>0.051</td>
<td>0.189</td>
<td>0.185</td>
<td>0.059</td>
<td>0.058</td>
</tr>
<tr>
<td>4</td>
<td>0.318</td>
<td>0.272</td>
<td>0.108</td>
<td>0.092</td>
<td>0.354</td>
<td>0.322</td>
<td>0.112</td>
<td>0.100</td>
</tr>
<tr>
<td>5</td>
<td>0.335</td>
<td>0.257</td>
<td>0.108</td>
<td>0.083</td>
<td>0.453</td>
<td>0.318</td>
<td>0.123</td>
<td>0.094</td>
</tr>
<tr>
<td>6</td>
<td>0.366</td>
<td>0.329</td>
<td>0.126</td>
<td>0.109</td>
<td>0.393</td>
<td>0.360</td>
<td>0.121</td>
<td>0.110</td>
</tr>
<tr>
<td>7</td>
<td>0.094</td>
<td>0.099</td>
<td>0.029</td>
<td>0.032</td>
<td>0.115</td>
<td>0.122</td>
<td>0.029</td>
<td>0.031</td>
</tr>
<tr>
<td>8</td>
<td>0.083</td>
<td>0.084</td>
<td>0.031</td>
<td>0.031</td>
<td>0.108</td>
<td>0.108</td>
<td>0.036</td>
<td>0.036</td>
</tr>
<tr>
<td>9</td>
<td>0.127</td>
<td>0.122</td>
<td>0.042</td>
<td>0.044</td>
<td>0.124</td>
<td>0.118</td>
<td>0.035</td>
<td>0.033</td>
</tr>
<tr>
<td>10</td>
<td>0.252</td>
<td>0.271</td>
<td>0.075</td>
<td>0.081</td>
<td>0.332</td>
<td>0.345</td>
<td>0.093</td>
<td>0.097</td>
</tr>
<tr>
<td>11</td>
<td>0.267</td>
<td>0.269</td>
<td>0.072</td>
<td>0.078</td>
<td>0.303</td>
<td>0.302</td>
<td>0.109</td>
<td>0.108</td>
</tr>
<tr>
<td>12</td>
<td>0.122</td>
<td>0.159</td>
<td>0.040</td>
<td>0.049</td>
<td>0.122</td>
<td>0.167</td>
<td>0.039</td>
<td>0.049</td>
</tr>
<tr>
<td>13</td>
<td>0.269</td>
<td>0.268</td>
<td>0.073</td>
<td>0.076</td>
<td>0.324</td>
<td>0.317</td>
<td>0.072</td>
<td>0.075</td>
</tr>
<tr>
<td>14</td>
<td>0.312</td>
<td>0.315</td>
<td>0.082</td>
<td>0.088</td>
<td>0.307</td>
<td>0.306</td>
<td>0.080</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>$p &lt; 10$</td>
<td>$10 &lt; p &lt; 20$</td>
<td>$20 &lt; p &lt; 30$</td>
<td>$p &gt; 30$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------------------------</td>
<td>-------------------------------------------------------------</td>
<td>-------------------------------------------------------------</td>
<td>---------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>men</td>
<td>1 Álava, 17 Gerona, 20 Guipúzcoa, 8 Barcelona, 22 Huesca, 31 Navarra, 48 Vizcaya, 39 Cantabria, 7 Baleares</td>
<td>19 Guadalajara, 50 Zaragoza, 33 Asturias, 28 Madrid, 44 Teruel, 9 Burgos, 46 Valencia, 43 Tarragona, 12 Castellón, 26 La Rioja, 3 Alicante, 47 Valladolid, 36 Pontevedra, 25 Lérida, 24 León, 27 Lugo, 34 Palencia, 21 Huelva, 32 Orense</td>
<td>29 Málaga, 41 Sevilla, 15 Coruña La, 30 Murcia, 35 Palmas Las, 42 Soria, 52 Melilla, 45 Toledo, 40 Segovia, 38 S.C. de Tenerife, 5 Ávila, 2 Albacete, 23 Jaén, 11 Cádiz, 13 Ciudad Real, 10 Cáceres, 4 Almería, 18 Granada, 37 Salamanca, 49 Zamora</td>
<td>14 Córdoba, 16 Cuenca, 51 Ceuta, 6 Badajoz,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>women</td>
<td>$p &lt; 10$</td>
<td>$10 &lt; p &lt; 20$</td>
<td>$20 &lt; p &lt; 30$</td>
<td>$p &gt; 30$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17, 1, 31, 22, 48, 20</td>
<td>8, 9, 7, 28, 33, 39, 50, 46, 19, 12, 43, 44, 3, 32, 36, 24</td>
<td>26, 47, 25, 21, 41, 45, 27, 15, 29, 35, 34, 30, 52, 49, 38, 2, 37</td>
<td>11, 14, 23, 13, 5, 18, 4, 42, 10, 6, 40, 51, 16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Spanish provinces classified by poverty proportion in %.
### Table 8. Spanish provinces classified by poverty gap in %.

<table>
<thead>
<tr>
<th>Gender</th>
<th>$g &lt; 3$</th>
<th>$3 &lt; g &lt; 6$</th>
<th>$6 &lt; g &lt; 10$</th>
<th>$g &gt; 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>men</td>
<td></td>
<td>19 Guadalajara, 17 Gerona, 43 Tarragona, 1 Álava, 48 Vizcaya, 39 Cantabria, 33 Asturias, 20 Guipúzcoa, 22 Huesca</td>
<td>31 Navarra, 8 Barcelona, 36 Pontevedra, 7 Baleares, 28 Madrid, 41 Sevilla, 44 Teruel, 26 La Rioja, 50 Zaragoza, 9 Burgos, 12 Castellón, 3 Alicante, 46 Valencia, 32 Orense, 34 Palencia, 47 Valladolid, 27 Lugo</td>
<td>42 Soria, 24 León, 15 Coruña La, 35 Palmas Las, 40 Segovia, 30 Murcia, 13 Ciudad Real, 45 Toledo, 11 Cádiz, 29 Málaga, 10 Cáceres, 25 Lérida, 21 Huelva, 38 S.C. de Tenerife, 37 Salamanca, 2 Albacete, 5 Ávila, 23 Jaén, 14 Córdoba, 4 Almería, 16 Cuenca</td>
</tr>
<tr>
<td>women</td>
<td></td>
<td>1, 17, 48, 31, 19, 22</td>
<td>39, 7, 43, 33, 9, 28, 8, 20, 32, 46, 26, 36, 41, 12, 50, 27, 3, 34, 44</td>
<td>45, 25, 21, 24, 47, 42, 13, 15, 35, 37, 30, 14, 29, 5, 38, 40, 49, 10, 4</td>
</tr>
</tbody>
</table>
Figure 13. Estimates of Spanish poverty proportions (men).
Figure 14. Estimates of Spanish poverty proportions (women).
Figure 15. Estimates of Spanish poverty gaps (men).
Figure 16. Estimates of Spanish poverty gaps (women).
In this talk

1. an introduction to SAE and
2. a brief description to SAMPLE project

have been given.

- SAE still has a lot of open problems
- SAMPLE project
  - deals with estimation of nonlinear parameters
  - uses temporal and spatial models
  - develops software in the R programming language
  - applies SAE methodology to real data from the SLCS
Thank you
for your attention
Let us consider the model

\[ y_{dt} = x_{dt}\beta + v_d + u_{dt} + e_{dt}, \quad d = 1, \ldots, D, \quad t = 1, \ldots, m_d, \quad (6) \]

where

- \( y_{dt} \) is a direct estimator of the indicator of interest for area \( d \) and time instant \( t \),
- \( x_{dt} \) is a vector containing the aggregated (population) values of \( p \) auxiliary variables,
- the random effects \( v_1, \ldots, v_D \) are SAR(1) with parameters \( \sigma_v^2 \) and \( \rho_1 \)
- the random vectors \( (u_{d1}, \ldots, u_{dm_d}) \), \( d = 1, \ldots, D \), are i.i.d. AR(1), with variance and auto-correlation parameters \( \sigma_u^2 \) and \( \rho_2 \) respectively,
- the errors \( e_{dt} \)’s are independent \( N(0, \sigma_{dt}^2) \) with known \( \sigma_{dt}^2 \)’s, and
- the \( v_d \)’s, the \( u_{dt} \)’s and the \( e_{dt} \)’s are independent.
The model (6) can be written alternatively in the form

\[ y = X\beta + Z_1 v + Z_2 u + e, \]  

(7)

where

- \[ y = \text{col}_{1 \leq d \leq D} (y_d),\]
  \[ y_d = \text{col}_{1 \leq t \leq m_d} (y_{dt}),\]

- \[ v = \text{col}_{1 \leq d \leq D} (v_d),\]

- \[ u = \text{col}_{1 \leq d \leq D} (u_d),\]
  \[ u_d = \text{col}_{1 \leq t \leq m_d} (u_{dt}),\]

- \[ e = \text{col}_{1 \leq d \leq D} (e_d),\]
  \[ e_d = \text{col}_{1 \leq t \leq m_d} (e_{dt}),\]

- \[ X = \text{col}_{1 \leq d \leq D} (X_d),\]
  \[ X_d = \text{col}_{1 \leq t \leq m_d} (x_{dt}),\]
  \[ x_{dt} = \text{col}'_{1 \leq j \leq p} (x_{dtj}),\]

- \[ \beta = \text{col}_{1 \leq j \leq p} (\beta_j),\]
  \[ Z_1 = \text{col}_{1 \leq d \leq D} (1_{m_d}),\]
  \[ Z_2 = I_{M \times M},\]
  \[ M = \sum_{d=1}^D m_d.\]
We assume that

\[ \mathbf{v} = \rho_1 \mathbf{Wv} + \varepsilon, \quad \text{where} \quad \varepsilon \sim N(0, \sigma_v^2 \mathbf{I}_D). \]

The diagonal elements of the proximity matrix \( \mathbf{W} \) are zero and we consider that the rows of this matrix are standardized in the sense that they sum up to one.

Under this setup, \( \rho_1 \in (-1, 1) \) is called spatial autocorrelation parameter.

We also assume that the matrix \( (\mathbf{I}_D - \rho_1 \mathbf{W}) \) is non-singular.

Then \( \mathbf{v} \) can be expressed as

\[ \mathbf{v} = (\mathbf{I}_D - \rho_1 \mathbf{W})^{-1} \varepsilon \sim N(0, \mathbf{G}), \]

where

\[ \mathbf{G} = \sigma_v^2 [(\mathbf{I}_D - \rho \mathbf{W})'(\mathbf{I}_D - \rho \mathbf{W})]^{-1}. \]
We assume that

\[ u \sim N(0, \Sigma_u) \text{ and } e \sim N(0, \Sigma_e) \] are independent,

with covariance matrices

\[ \Sigma_u = \sigma_u^2 \Omega(\rho), \quad \Omega(\rho) = \text{diag } (\Omega_d(\rho)), \]

\[ \Sigma_e = \text{diag } (\Sigma_{ed}), \quad \Sigma_{ed} = \text{diag } (\sigma_{dt}^2), \]

where the \( \sigma_{dt}^2 \) are known, and

\[ \Omega_d = \Omega_d(\rho) = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \ldots & \rho^{m_d - 2} & \rho^{m_d - 1} \\ \rho & 1 & \ddots & & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ \rho^{m_d - 2} & \ddots & 1 & \rho \\ \rho^{m_d - 1} & \rho^{m_d - 2} & \ldots & \rho & 1 \end{pmatrix}_{m_d \times m_d} \]