SAMPLE DELIVERABLES 17

PILOT APPLICATIONS

Grant agreement No:	SSH - CT - 2007 - 217565
Project Acronym:	SAMPLE
Project Full title:	Small Area Methods for Poverty and Living Conditions Estimates
Funding Scheme:	Collaborative Project - Small or medium scale focused research project
Deliverable n.:	D17
Deliverable name:	Pilot applications
WP n.:	2
Lead Beneficiary:	1, 5
Nature:	Report
Dissemination level:	PU
Status:	Final
Due delivery date:	September 30^{th}, 2010
Actual delivery date:	November 16^{th}, 2010
Project co-ordinator name:	Monica Pratesi
Title:	Associate Professor of Statistics - University of Pisa
Organization:	Department of Statistics and Mathematics Applied to Economics of the University of Pisa (UNIPI-DSMAE)
Tel.:	+39-050-2216252, +39-050-2216492
Fax:	+39-050-2216375
E-mail:	coordinator@sample-project.eu
Project website address:	www.sample-project.eu

PILOT APPLICATIONS

Project Acronym:	SAMPLE
Project Full title:	Small Area Methods for Poverty and Living Conditions Estimates
Project/Contract No:	EU-FP7-SSH-2007-1
Grant agreement No:	217565
Work Package 2:	Small area estimation of poverty and inequality indicators
Document title:	Pilot applications
Date:	October 30^{th}, 2010
Type of document:	Deliverable D17
Status:	Final
Editors:	I. Molina, D. Morales, M. Pratesi and N. Tzavidis
Authors:	Monica Pratesi (m.pratesi@ec.unipi.it), UNIPI-DSMAE
	Caterina Giusti (caterina.giusti@ec.unipi.it), UNIPI-DSMAE
	Stefano Marchetti (s.marchetti@ds.unifi.it), UNIPI-DSMAE
	Nicola Salvati (salvati@ec.unipi.it), UNIPI-DSMAE
	Nikos Tzavidis (n.tzavidis@soton.ac.uk), Southampton
	Isabel Molina (imolina@est-econ.uc3m.es), UC3M
	María Durbán (mdurban@est-econ.uc3m.es), UC3M
	Aurea Grané (agrane@est-econ.uc3m.es), UC3M
	Juan Miguel Marín (jmmarin@est-econ.uc3m.es), UC3M
	Maria Helena Veiga (mhveiga@est-econ.uc3m.es), UC3M
	Domingo Morales (d.morales@umh.es), UMH
	María Dolores Esteban (md.esteban@umh.es), UMH
	Ángel Sánchez (a.sanchez@umh.es), UMH
	Laureano Santamaría (l.santamaria@umh.es), UMH
	Yolanda Marhuenda (y.marhuenda@umh.es), UMH
	Agustín Pérez (agustin.perez@umh.es), UMH
	Maria Chiara Pagliarella (mc.pagliarella@unicas.it), UMH
	J.N.K. Rao (jrao@math.carleton.ca), Carleton University
	Caterina Ferretti (ferretti@ds.unifi.it), University of Florence

Contents

Prologue

This report contains the pilot applications based on the final small area developments of the WP2 partners in the SAMPLE project. The target of the report is to illustrate the use of the developed statistical methodology by means of applications to real data. The manuscript is organized in ten chapters.

Chapters 1 and 2 introduce the Spanish and Italian data. They present a descriptions of the variables and data sources.

Chapters 3 and 4 describe the applications to real data of the introduced area level time models with correlated or independent time effects and with different random effect variances for men and women.

Chapters 5 and 6 deal with the prediction of poverty measures by using EB prediction or fast EB prediction under unit level models.

Chapters 7 and 8 present EBLUP estimates of poverty measures based on Spatial Fay-Herriot models and Spatio-temporal Fay-Herriot models.

Chapter 9 proposes to estimate poverty measures by aggregating data at a subdomain level and by applying unit level model to at that level.

Chapter 10 reports a review of the M-quantile methodologies for small area estimation.

Chapter 11 treats the problem of estimating Small Area Averages. It uses M-quantile models, nonparametric M-quantile models, M-quantile GWR models and semiparametric Fay and Herriot models.

Chapter 12 presents some procedures to estimate the small area income distribution functions. They are based on M-quantile models and nonparametric M-quantile models.

Chapter 13 gives an application to the estimation of poverty indicators by using M-quantile models.

The manuscript has also a section of references and two appendixes with Tables of numerical results.

This report has been coordinated by Domingo Morales (UMH). He has also been in charge of writing Chapters 1, 3, 4 and 9. Isabel Molina (UC3M) has been responsible for the elaboration of Chapters 5, 6 and 7. Chapter 8 has been produced in collaboration by the UMH and UC3M teams. Finally, Nikos Tzavidis (Southampton) and Monica Pratesi (UNIPI-DSMAE) have coordinated the production of the contents of Chapters 2, 10, 11, 12 and 13.

Chapter 1

The Spanish data

1.1 The Spanish Living Conditions Survey

The Living Conditions Survey (LCS) 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 used with first stage unit stratification. The first stage is formed by census sections and the second stage by main family dwellings. Within these no sub-sampling is carried out, investigating all dwellings that are their usual residence.

The framework used for the sample selection is an area framework formed by the relation of existing census sections used in the 2003 Municipal Register of Inhabitants. Second stage units use the list of main family dwellings in each of the sections selected for the sample. The first stage units are grouped into strata in agreement with the size of the municipality to which the section belongs. For each Autonomous Community an independent sample that represents it is designed, due to one of the objectives of the survey being to facilitate data. In order to achieve the goals set out for the survey to provide estimates with an acceptable level of reliability on a national and Autonomous Community level, the sample selected includes 16000 dwellings distributed in 2000 census sections.

The sample is distributed between Autonomous Communities assigning a portion uniformly and another in proportion to the size of the Community. The uniform part corresponds to approximately 40% of the sections. The sections are selected within each stratum with a probability proportional to their size. The dwellings, in each section, with the same probability via random start systematic sampling. This procedure leads to self-weighted samples in each stratum.

We use unit-level data from the Spanish Living Conditions Survey (SLCS) corresponding to 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), which is one of the statistical operations that have been harmonized for EU countries. The SLCS started in 2004 with an annual periodicity. Its main goal is to provide a reference source on comparative statistics on the distribution of income and social exclusion in the European environment.

Description of variables within the SLCS data files is given in Tables 1.1a-1.1c.

Position	Description	Observations
1-3	Autonomous Community	1 - Andalucía 2 - Aragón 3 - Principado de Asturias 4 - Islas Baleares 5 - Canarias 6 - Cantabria 7 - Castilla León 8 - Castilla La Mancha 9 - Cataluña 10 - Comunidad Valencia 11 - Extremadura 12 - Galicia 13 - Comunidad de Madrid 14 - Región de Murcia 15 - Comunidad Foral de Navarra 16 - País Vasco 17 - La Rioja 18 - Ceuta y Melilla
4-6	Province	1 - Álava 2 - Albacete 3 - Alicante 4 - Almería 5 - Ávila 6 - Badajoz 7 - Baleares 8 - Barcelona 9 - Burgos 10 - Cáceres 11 - Cádiz 12 - Castellón 13 - Ciudad Real 14 - Córdoba 15 - Coruña, La 16 - Cuenca 17 - Gerona 18 - Granada 19 - Guadalajara 20 - Guipúzcoa 21 - Huelva 22 - Huesca 23 - Jaén 24 - León 25 - Lérida 26 - La Rioja

Table 1.1a. Description of variables in the SLCS data files

Position	Description	Observations
4-6	Province	27 - Lugo 28 - Madrid 29 - Málaga 30 - Murcia 31 - Navarra 32 - Orense 33 - Asturias (Oviedo) 34 - Palencia 35 - Palmas Las 36 - Pontevedra 37 - Salamanca 38 - Santa Cruz de Tenerife 39 - Cantabria (Santander) 40 - Segovia 41 - Sevilla 42 - Soria 43 - Tarragona 44 - Teruel 45 - Toledo 46 - Valencia 47 - Valladolid 48 - Vizcaya 49 - Zamora 50 - Zaragoza 51 - Ceuta 52 - Melilla
7-11	Size of municipality	0 - Municipalities with more than 500.000 people 1 - Capitals of provinces (except previous ones - epo) 2 - Municipalities with more than 100.000 people (epo) 3 - Municipalities between 50.000 and 100.000 people (epo) 4 - Municipalities between 20.000 and 50.000 people (epo) 5 - Municipalities between 10.000 and 20.000 people 6 - Municipalities with less than 10.000 people
12-21	Census section	
22-36	Dwelling	Numbered within Census section
37-39	Person	Numbered within dwelling
40-42	Sex	1 - Man 2 - Woman
43-45	Age	0 - $\leq$ 13 years 1 - 14-15 2 - 16-24 3 - 25-49 4 - 50-64 5 - $\geq$ 65

Table 1.1b. Description of variables in the SLCS data files.

Position	Description	Observations
46-47	Citizenship	1 - Spanish 2 - Non Spanish
48-49	Education	0 - $\leq$ 16 years 1 - Primary education level 2 - Secondary education level 3 - University level
50-51	Labor	0 - $\leq$ 16 years 1 - Employed 2 - Unemployed 3 – Inactive
52-66	Net annual income of a household	
67-86	Elevation factor	calibrated survey weight
87-101	Longitudinal household identifier	
102-116	longitudinal individual identifier	
117-120	Province-Sex (domain)	
121	Age2	"1" if Age=2, "0" otherwise
122	Age3	"1" if Age=3 "0" otherwise
123	Age4	"1" if Age=4, "0" otherwise
124	Age5	"1" if Age=5, "0" otherwise
125	Edu1	"1" if Education=1, "0" otherwise
126	Edu2	"1" if Education=2, "0" otherwise
127	Edu3	"1" if Education=3, "0" otherwise
128	Cit1	"1" if Citizenship=1, "0" otherwise
129	Lab1	"1" if Labor=1, "0" otherwise
130	Lab2	"1" if Labor=2, "0" otherwise
131	Lab3	"1" if Labor=3, "0" otherwise
132-137	H-value	Normalized household size
138-152	NNAIH	Normalized net annual income of a household
153-154	Poverty (PI0)	1 if NNAIH $<$ poverty threshold, 0 otherwise
155-165	Gap (PI1)	
166-176	PI2	

Table 1.1c. Description of variables in the SLCS data files

Variables appearing in positions 1-131 of Tables 1.1 are unit-level auxiliary x-variables. Variables H-value and NNAIH are needed to construct the target variables PI0, PI1 and PI2. These two auxiliary variables are used to calculate the poverty thresholds. Following the standards of the Spanish Statistical Office, the poverty threshold is fixed as the 60% of the median of the normalized incomes in Spanish households.

The aim of normalizing the household net annual income (see position 52-66 in Table 1.1c) is to adjust for the varying size and composition of households. The definition of the total number of normalized household members is the modified OECD scale used by EUROSTAT. This scale gives a weight of 1.0 to the first adult, 0.5 to the second and each subsequent person aged 14 and over and 0.3 to each child aged under 14 in the household. The *normalized size* of a household is the sum of the weights assigned to each person. So the total number of normalized household members is

$$H_{dth} = 1 + 0.5(H_{dth\geq 14} - 1) + 0.3H_{dth<14}$$

where $H_{dth\geq 14}$ is the number of people aged 14 and over and $H_{dth<14}$ 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. This variable (labeled NNAIH in Table 1.1c) is denoted by z_{dtj} when it is measured at individual j, time instant t and domain (province -sex) d. The Spanish poverty thresholds (in euros) in 2004-06 are $z_{2004} = 6098.57$, $z_{2005} = 6160.00$ and $z_{2006} = 6556.60$ respectively. These are the z_t-values used in the calculation of the direct estimates of the poverty incidence and gap.

At the unit-level, variables PI0, PI1 and PI2 ($y_{\alpha,dtj}$, $\alpha = 0,1,2$, in mathematical notation) are

$$y_{\alpha;dtj} = \left(\frac{z_t - z_{dtj}}{z_t}\right)^{\alpha} I(z_{dtj} < z_t), \tag{1.1}$$

$I(z_{dtj} < z_t) = 1$ if $z_{dtj} < z_t$ and $I(z_{dtj} < z_t) = 0$ otherwise. At the domain-level these variables are the target population parameters, i.e.

$$Y_{\alpha;dt} = \frac{1}{N_{dt}} \sum_{j=1}^{N_{dt}} y_{\alpha;dtj},$$

where N_{dt} is the total number of individuals in domain d at time instant t. The proportion of units under poverty in the domain d and period t is $Y_{0;dt}$ and the poverty gap is $Y_{1;dt}$.

1.2 The Economically Active Population Survey

The main objective of the Economically Active Population Survey (EAPS) is to reveal information on economic activities as regards their human component. It focuses on providing data on the main population categories related to the labour market (employed, unemployed, active population, inactive population). The survey has been designed to provide detailed results on a national level. As regards Autonomous Communities and provinces, information is featured on the main characteristics, using the breakdown level that can be achieved using the estimators' variation coefficients.

The survey uses a two-stage sampling with first stage unit stratification. First stage units are composed by census sections. Second stage units are composed by main family dwellings (permanently inhabited) and permanent accommodations. Sub-sampling is not carried out in second stage units, information is collected on all persons who regularly live in the same.

Since the first quarter of 2005, a sample size of 3,588 sections and 18 dwellings per section is established, except in the provinces of Madrid, Barcelona, Sevilla, Valencia and Zaragoza, in which the number of interviews per section rises to 22. Sample sizes are considered large enough to provide reliable direct estimates at the province level. Further, sampling weights are calibrated to adjust the survey estimates to the information from external sources. Because of this reason, EAPS sampling weights have been used to construct the Spanish aggregated auxiliary variables for the sex-province domains. More concretely, let x_{dtj} be the EAPS value of auxiliary variable x at individual j, time instant t and domain d. Let w_{dtj} be the corresponding EAPS calibrated sampling weight which take into account for non response. The direct estimator of the total $X_{dt} = \sum_{j=1}^{N_{dt}} x_{dtj}$ is

$$\hat{X}_{dt}^{dir} = \sum_{j \in S_{dt}} w_{dtj} x_{dtj}.$$

where S_{dt} is the domain sample at time period t. In the particular case $x_{dtj} = 1$, for all $j \in S_{dt}$, we get the estimated domain size

$$\hat{N}_{dt}^{dir} = \sum_{j \in S_{dt}} w_{dtj}.$$

Using this quantity, a direct estimator of the domain mean $\bar{X}_{dt}$ is $\bar{X}_{dt} = \hat{X}_{dt}^{dir}/\hat{N}_{dt}^{dir}$. The EAPS direct estimates of the domain means are used as auxiliary aggregated counterparts of unit-level auxiliary variables appearing in the SLCS files. The considered auxiliary aggregated variables are:

- INTERCEPT: $\bar{X}_{dt}^{(0)} = 1$.
- AGE: Age groups are *age1-age5* for the intervals ≤ 15, $16-24$, $25-49$, $50-64$ and ≥ 65. The corresponding auxiliary variables (domain proportions) are denoted by $\bar{X}_{dt,1}^{(1)}, \ldots, \bar{X}_{dt,5}^{(1)}$.
- EDUCATION: Highest level of education completed, with 4 categories denoted by *edu0* for Less than primary education level, *edu1* for Primary education level, *edu2* for Secondary education level and *edu3* for University level. Auxiliary variables are $\bar{X}_{dt,0}^{(2)}, \bar{X}_{dt,1}^{(2)}, \bar{X}_{dt,2}^{(2)}, \bar{X}_{dt,3}^{(2)}$.
- CITIZENSHIP: with 2 categories denoted by *cit1* for Spanish and *cit2* for Not Spanish. Auxiliary variables are $\bar{X}_{dt,1}^{(3)}, \bar{X}_{dt,2}^{(3)}$.
- LABOR: Labor situation with 4 categories taking the values *lab0* for Below 16 years, *lab1* for Employed, *lab2* for Unemployed and *lab3* for Inactive. Auxiliary variables are $\bar{X}_{dt,0}^{(4)}, \bar{X}_{dt,1}^{(4)}, \bar{X}_{dt,2}^{(4)}, \bar{X}_{dt,3}^{(4)}$.

The values of the considered auxiliary variables are stored in the following files:

- VauxN-EAPS-Province contains the population totals (N_d) per sex-province. They have been estimated by using the survey data from the EAPS 2004-06.
- VauxTot-EAPS-Age contains the population totals of the age categories per sex-province. They have been estimated by using the survey data from the EAPS 2004-06.
- VauxTot-EAPS-Education contains the population totals of the Education categories per sex-province. They have been estimated by using the survey data from the EAPS 2004-06.
- VauxTot-EAPS-Nac contains the population totals of the citizenship categories per sex-province. They have been estimated by using the survey data from the EAPS 2004-06.
- VauxTot-EAPS-Labour contains the population totals of the Labor categories per sex-province. They have been estimated by using the survey data from the EAPS 2004-06.

Chapter 2

The Italian data

2.1 EU-SILC Data and Census Microdata

In Italy, the European Survey on Income and Living Conditions (EU-SILC) is conducted yearly by ISTAT to produce estimates on the living conditions of the population at national and regional (NUTS-2) levels.

Regions are planned domains for which EU-SILC estimates are published, while the provinces are unplanned domains. These are the administrative areas (LAU-1 level) constituted by a different number of municipalities (LAU-2 level) whose boundaries do not cut across the municipalities themselves. The regional samples are based on a stratified two stage sample design: in each province the municipalities are the Primary Sampling Units (PSUs), while the households are the Secondary Sampling Units (SSUs). The PSUs are divided into strata according to their dimension in terms of population size; the SSUs are selected by means of systematic sampling in each PSU. All the members of each sampled household are interviewed through an individual questionnaire, and one individual in each household (usually, the head of the household) is interviewed through a household questionnaire. It is useful to note that some provinces, generally the smaller ones, may have very few sampled municipalities; furthermore, many municipalities are not even included in the sample at all. Direct estimates may therefore have large errors at provincial level or they may not even be computable at municipality level, thereby requiring resort to small area estimation techniques.

In this Deliverable we will use data from the survey EU-SILC 2007, focusing on three Italian regions: Lombardia, in the North of the country, Toscana, in Central Italy, and Campania, in Southern Italy. The choice of these three regions, out of the 20 existing regions in Italy, is motived by the geographical differences characterizing the Italian territory. In particular, we believe that using this data it is possible to investigate the so-called ”north-south” divide characterizing the Italian territory, since each of the three regions can be considered as representative of the corresponding geographical area of Italy (Northern, Central and Southern/Insular Italy).

Table 2.1 reports the total number of households and the number of sampled households for each province in the three regions of interest.

As concerns the municipalities, Table 2.2 reports the total number of municipalities and the number of sampled municipalities in each province of the three regions. As we can see, the number of sampled

Table 2.1: Number of total households and number of sampled households in the provinces of Lombardia, Toscana and Campania.

Province / Region	Total households	Sampled households
Varese / Lombardia	320899	253
Como / Lombardia	205963	153
Sondrio / Lombardia	69817	41
Milano Province / Lombardia	1545502	798
Bergamo / Lombardia	375778	219
Brescia / Lombardia	437706	216
Pavia / Lombardia	211786	60
Cremona / Lombardia	135321	75
Mantova/ Lombardia	146249	234
Lecco / Lombardia	121321	103
Lodi / Lombardia	77978	62
Massa-Carrara / Toscana	80810	96
Lucca / Toscana	146117	131
Pistoia / Toscana	104466	129
Firenze Province / Toscana	376255	445
Livorno / Toscana	133729	115
Pisa / Toscana	150259	136
Arezzo / Toscana	123880	145
Siena / Toscana	101399	116
Grosseto / Toscana	87720	59
Prato / Toscana	83617	117
Caserta / Campania	279684	155
Benevento / Campania	102441	70
Napoli Province / Campania	969310	817
Avellino / Campania	152340	84
Salerno/ Campania	359080	191

municipalities is rather low: only 109 out of 1546 municipalities in Lombardia, 65 out of 287 in Toscana and 55 out of 551 in Campania. Therefore, producing estimates at municipality level requires the use of advanced small area models, as those presented in Deliverables 12 and 6 that are applied to EU-SILC data in this Deliverable.

Applying the small area methodologies presented in Deliverable 12 and 16 to data from the EU-SILC survey requires covariate information that is also known for not sampled (out of sample) households. This information is available from the 2001 Population Census of Italy. The Population Census of Italy has a very comprehensive questionnaire, collecting information on each household and on each individual living in the Italian territory. For the purpose of obtaining estimates on poverty and living conditions in the three regions of interest we selected census variables that are also available from the EU-SILC

Table 2.2: Number of total municipalities and number of sampled municipalities in the provinces of Lombardia, Toscana and Campania.

Province / Region	Total municipalities	Sampled municipalities
Varese / Lombardia	141	16
Como / Lombardia	162	8
Sondrio / Lombardia	78	2
Milano / Lombardia	189	25
Bergamo / Lombardia	244	14
Brescia / Lombardia	206	14
Pavia / Lombardia	190	5
Cremona / Lombardia	115	3
Mantova/ Lombardia	70	12
Lecco / Lombardia	90	6
Lodi / Lombardia	61	4
Massa-Carrara / Toscana	17	5
Lucca / Toscana	35	5
Pistoia / Toscana	22	6
Firenze / Toscana	44	18
Livorno / Toscana	20	3
Pisa / Toscana	39	7
Arezzo / Toscana	39	7
Siena / Toscana	36	7
Grosseto / Toscana	28	3
Prato / Toscana	7	4
Caserta / Campania	104	7
Benevento / Campania	78	5
Napoli / Campania	92	26
Avellino / Campania	119	9
Salerno/ Campania	158	8

survey. These variables include information referring to the head of the household, namely gender, age, occupational status, civil status and years in education, and information referring to the household, that is ownership of the house, squared meters of the house and number of household members. All these variables can be used as covariates in the working small area models used for estimating income distributions and poverty indicators. Thus, we can say that the EU-SILC datasets, together with data coming from the Population Census of Italy, represent a complete and valuable source of information that can be used for applying advanced small area estimation techniques for producing poverty and living condition estimates in Italy. As a final remark, it is important to underline that EU-SILC and census data are confidential. These data were provided by ISTAT, the Italian National Institute of Statistics, to the researchers of the SAMPLE project and were analyzed by respecting all confidentiality restrictions.

Chapter 3

Area level time models

3.1 Area-level model with correlated time effects

3.1.1 The model

Let us consider the model

$$y_{dt} = \mathbf{x}_{dt}\beta + u_{dt} + e_{dt}, \quad d = 1,\dots,D, \quad t = 1,\dots,m_d, \tag{3.1}$$

where y_{dt} is a direct estimator of the indicator of interest for area d and time instant t, and $\mathbf{x}_{dt}$ is a vector containing the aggregated (population) values of p auxiliary variables. The index d is used for domains and the index t for time instants. We further assume that the random vectors $(u_{d1},\dots,u_{dm_d})$, $d = 1,\dots,D$, follow i.i.d. AR(1) processes with variance and auto-correlation parameters σ_u^2 and ρ respectively, the errors e_{dtj}'s are independent $N(0,\sigma_{dt}^2)$ with known σ_{dt}^2's, and the u_{dt}'s are independent of the e_{dt}'s.

In matrix notation the model is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{3.2}$$

where $\mathbf{y} = \underset{1\le d\le D}{\text{col}}(\mathbf{y}_d)$, $\mathbf{y}_d = \underset{1\le t\le m_d}{\text{col}}(y_{dt})$, $\mathbf{u} = \underset{1\le d\le D}{\text{col}}(\mathbf{u}_d)$, $\mathbf{u}_d = \underset{1\le t\le m_d}{\text{col}}(u_{dt})$, $\mathbf{e} = \underset{1\le d\le D}{\text{col}}(\mathbf{e}_d)$, $\mathbf{e}_d = \underset{1\le t\le m_d}{\text{col}}(e_{dt})$, $\mathbf{X} = \underset{1\le d\le D}{\text{col}}(\mathbf{X}_d)$, $\mathbf{X}_d = \underset{1\le t\le m_d}{\text{col}}(\mathbf{x}_{dt})$, $\mathbf{x}_{dt} = \underset{1\le i\le p}{\text{col}'}(x_{dti})$, $\beta = \underset{1\le i\le p}{\text{col}}(\beta_i)$, $\mathbf{Z} = \mathbf{I}_{M\times M}$ and $M = \sum_{d=1}^{D} m_d$. In this notation, $\mathbf{u} \sim N(\mathbf{0},\mathbf{V}_u)$ and $\mathbf{e} \sim N(\mathbf{0},\mathbf{V}_e)$ are independent with covariance matrices

$$\mathbf{V}_u = \sigma_u^2\Omega(\rho), \quad \Omega(\rho) = \underset{1\le d\le D}{\text{diag}}(\Omega_d(\rho)), \quad \mathbf{V}_e = \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_{ed}), \quad \mathbf{V}_{ed} = \underset{1\le t\le m_d}{\text{diag}}(\sigma_{dt}^2),$$

where the σ_{dt}^2 are known and

$$\Omega_d = \Omega_d(\rho) = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \dots & \rho^{m_d-2} & \rho^{m_d-1} \\ \rho & 1 & \ddots & & \rho^{m_d-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{m_d-2} & & \ddots & 1 & \rho \\ \rho^{m_d-1} & \rho^{m_d-2} & \dots & \rho & 1 \end{pmatrix}_{m_d\times m_d}.$$

If the variance components are known, then the BLUE of β and the BLUP of $\mathbf{u}$ are

$$\widehat{\beta}=(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \qquad \text{and} \qquad \widehat{\mathbf{u}}=\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\widehat{\beta}),$$

where

$$var(\mathbf{y})=\mathbf{V}=\sigma_u^2 \underset{1\le d\le D}{\text{diag}}\,(\Omega_d(\rho))+\mathbf{V}_e=\underset{1\le d\le D}{\text{diag}}\,(\sigma_u^2\Omega_d(\rho)+\mathbf{V}_{ed})=\underset{1\le d\le D}{\text{diag}}\,(\mathbf{V}_d).$$

To calculate $\widehat{\beta}$ and $\widehat{\mathbf{u}}$ we apply the formulas

$$\widehat{\beta}=\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)^{-1}\left(\sum_{d=1}^{D}\mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{y}_d\right),\quad \widehat{\mathbf{u}}=\sigma_u^2\underset{1\le d\le D}{\text{col}}\Big(\Omega_d(\rho)\mathbf{V}_d^{-1}(\mathbf{y}_d-\mathbf{X}_d\widehat{\beta})\Big).$$

3.1.2 The REML estimators

The REML log-likelihood is

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

where

$$\mathbf{P}=\mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1},\quad \mathbf{PVP}=\mathbf{P},\quad \mathbf{PX}=\mathbf{0}.$$

Let us define $\theta=(\theta_1,\theta_2)=(\sigma_u^2,\rho)$, $\mathbf{V}_1=\frac{\partial\mathbf{V}}{\partial\sigma_u^2}=\underset{1\le d\le D}{\text{diag}}\,(\Omega_d(\rho))$ and $\mathbf{V}_2=\frac{\partial\mathbf{V}}{\partial\rho}=\sigma_u^2\underset{1\le d\le D}{\text{diag}}\,(\dot{\Omega}_d(\rho))$. Then

$$\mathbf{P}_a=\frac{\partial\mathbf{P}}{\partial\theta_a}=-\mathbf{P}\frac{\partial\mathbf{V}}{\partial\theta_a}\mathbf{P}=-\mathbf{P}\mathbf{V}_a\mathbf{P},\quad a=1,2.$$

By taking partial derivatives of l_{REML} with respect to θ_a, we get

$$S_a=\frac{\partial l_{REML}}{\partial\theta_a}=-\frac{1}{2}\text{tr}(\mathbf{P}\mathbf{V}_a)+\frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{y},\quad a=1,2.$$

If we take again partial derivatives with respect to θ_a and θ_b, we take expectations and we change the sign, we obtain the elements of the REML Fisher information matrix. These elements are

$$F_{ab}=\frac{1}{2}\text{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_b),\quad a,b=1,2.$$

We use the Fisher-scoring algorithm to calculate the REML estimates of θ. The updating formula is

$$\theta^{k+1}=\theta^k+\mathbf{F}^{-1}(\theta^k)\mathbf{S}(\theta^k).$$

As seeds we use $\rho=0$ and $\sigma_u^{2(0)}=\widehat{\sigma}_{uH}^2$, where $\widehat{\sigma}_{uH}^2$ is the Henderson 3 estimator of σ_u^2 under the model restricted to $\rho=0$. The REML estimator of β is calculated by applying the formula

$$\widehat{\beta}=(\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{y}.$$

The asymptotic distributions of the REML estimators of θ and β are

$$\hat{\theta} \sim N_2(\theta, \mathbf{F}^{-1}(\theta)), \quad \hat{\beta} \sim N_p(\beta, (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}).$$

Asymptotic confidence intervals at the level $1-\alpha$ for θ_a and β_i are

$$\hat{\theta}_a \pm z_{\alpha/2}\nu_{aa}^{1/2},\ a=1,2, \quad \hat{\beta}_i \pm z_{\alpha/2}q_{ii}^{1/2},\ i=1,\dots,p,$$

where $\hat{\theta}=\theta^{\kappa}$, $\mathbf{F}^{-1}(\theta^{\kappa})=(\nu_{ab})_{a,b=1,2}$, $(\mathbf{X}'\mathbf{V}^{-1}(\theta^{\kappa})\mathbf{X})^{-1}=(q_{ij})_{i,j=1,\dots,p}$, κ is the final iteration of the Fisher-scoring algorithm and z_α is the α-quantile of the standard normal distribution $N(0,1)$). Observed $\hat{\beta}_i=\beta_0$, the p-value for testing the hypothesis $H_0:\beta_i=0$ is

$$p = 2P_{H_0}(\hat{\beta}_i > |\beta_0|) = 2P(N(0,1) > |\beta_0|/\sqrt{q_{ii}}).$$

3.1.3 The EBLUP and its mean squared error

We are interested in predicting the value of $\mu_{dt}=\mathbf{x}_{dt}\beta+u_{dt}$ by using the EBLUP $\widehat{\mu}_{dt}=\mathbf{x}_{dt}\widehat{\beta}+\widehat{u}_{dt}$. If we do not take into account the error, e_{dt}, this is equivalent to predict $y_{dt}=\mathbf{a}'\mathbf{y}$, where $\mathbf{a}=\underset{1\le\ell\le D}{\text{col}}(\underset{1\le k\le m_\ell}{\text{col}}(\delta_{d\ell}\delta_{tk}))$ is a vector having one 1 in the position $t+\sum_{\ell=1}^{d-1}m_\ell$ and 0's in the remaining cells. To estimate $\overline{Y}_{dt}$ we use $\widehat{\overline{Y}}_{dt}^{eblup}=\widehat{\mu}_{dt}$. The mean squared error of $\widehat{\overline{Y}}_{dt}^{eblup}$ is

$$MSE(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\theta)+g_2(\theta)+g_3(\theta),$$

where $\theta=(\sigma_u^2,\rho)$,

$$\begin{aligned} g_1(\theta) &= \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}, \\ g_2(\theta) &= [\mathbf{a}'\mathbf{X}-\mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a}-\mathbf{X}'\mathbf{V}_e^{-1}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}], \\ g_3(\theta) &\approx \text{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\} \end{aligned}$$

The estimator of $MSE(\widehat{\overline{Y}}_{dt}^{eblup})$ is

$$mse(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\hat{\theta})+g_2(\hat{\theta})+2g_3(\hat{\theta}).$$

Calculation of $g_1(\theta)$

In the formula of $g_1(\theta)=\mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}$, we have that $\mathbf{Z}=\mathbf{I}_{M\times M}$, and

$$\mathbf{T}=\mathbf{V}_u-\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{V}_u=\sigma_u^2\underset{1\le d\le D}{\text{diag}}(\Omega_d(\rho))-\sigma_u^4\underset{1\le d\le D}{\text{diag}}(\Omega_d(\rho))\underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d^{-1})\underset{1\le d\le D}{\text{diag}}(\Omega_d(\rho)).$$

Let us write $\Omega_d=\Omega_d(\rho)$ and $\mathbf{a}_d=\underset{1\le k\le m_d}{\text{col}}(\delta_{tk})$. Then, $g_1(\theta)$ can be expressed in the form

$$g_1(\theta)=\sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{a}_d-\sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d.$$

Calculation of $g_2(\theta)$

We have that $g_2(\theta) = [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{ZTZ}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{ZTZ}'\mathbf{a}]$, where

$$\begin{aligned}\mathbf{ZTZ}'\mathbf{V}_e^{-1}\mathbf{X} &= \left[\sigma_u^2 \underset{1\le d\le D}{\text{diag}}(\Omega_d) - \sigma_u^4 \underset{1\le d\le D}{\text{diag}}(\Omega_d) \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d^{-1}) \underset{1\le d\le D}{\text{diag}}(\Omega_d)\right] \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_{ed}^{-1}) \underset{1\le d\le D}{\text{col}}(\mathbf{X}_d) \\ &= \sigma_u^2 \underset{1\le d\le D}{\text{col}}(\Omega_d \mathbf{V}_{ed}^{-1}\mathbf{X}_d) - \sigma_u^4 \underset{1\le d\le D}{\text{col}}(\Omega_d \mathbf{V}_d^{-1}\Omega_d \mathbf{V}_{ed}^{-1}\mathbf{X}_d).\end{aligned}$$

Therefore

$$\begin{aligned} g_2(\theta) &= \left[\mathbf{a}_d'\mathbf{X}_d - \sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d + \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right]\mathbf{Q} \\ &\cdot \left[\mathbf{X}_d'\mathbf{a}_d - \sigma_u^2\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{a}_d + \sigma_u^4\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d\right].\end{aligned}$$

Calculation of $g_3(\theta)$

We have that

$$g_3(\theta) \approx \text{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\},$$

where

$$\mathbf{b}' = \mathbf{a}'\mathbf{Z}\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1} = \sigma_u^2\mathbf{a}' \underset{1\le \ell\le D}{\text{diag}}(\Omega_\ell) \underset{1\le \ell\le D}{\text{diag}}(\mathbf{V}_\ell^{-1}) = \sigma_u^2 \underset{1\le \ell\le D}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell\Omega_\ell\mathbf{V}_\ell^{-1}).$$

It holds that

$$\begin{aligned}\frac{\partial\mathbf{b}'}{\partial\sigma_u^2} &= \underset{1\le \ell\le D}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}) - \sigma_u^2 \underset{1\le \ell\le D}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}\mathbf{V}_{\ell u}\mathbf{V}_\ell^{-1}), \quad \mathbf{V}_{\ell u} = \frac{\partial\mathbf{V}_\ell}{\partial\sigma_u^2} = \Omega_\ell, \\ \frac{\partial\mathbf{b}'}{\partial\rho} &= \sigma_u^2 \underset{1\le \ell\le D}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\dot{\Omega}_\ell\mathbf{V}_\ell^{-1}) - \sigma_u^2 \underset{1\le \ell\le D}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\Omega_\ell\mathbf{V}_\ell^{-1}\mathbf{V}_{\ell\rho}\mathbf{V}_\ell^{-1}), \quad \mathbf{V}_{\ell\rho} = \frac{\partial\mathbf{V}_\ell}{\partial\rho} = \sigma_u^2\dot{\Omega}_\ell.\end{aligned}$$

We define

$$\begin{aligned} q_{11} = \frac{\partial\mathbf{b}'}{\partial\sigma_u^2} \underset{1\le \ell\le D}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}'}{\partial\sigma_u^2}\right)' &= \mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d - 2\sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d \\ &+ \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d, \\ q_{12} = \frac{\partial\mathbf{b}'}{\partial\sigma_u^2} \underset{1\le \ell\le D}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}'}{\partial\rho}\right)' &= \sigma_u^2\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d - \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d \\ &- \sigma_u^4\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d + \sigma_u^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d, \\ q_{22} = \frac{\partial\mathbf{b}'}{\partial\rho} \underset{1\le \ell\le D}{\text{diag}}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}'}{\partial\rho}\right)' &= \sigma_u^4\mathbf{a}_d'\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d - 2\sigma_u^6\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{a}_d \\ &+ \sigma_u^8\mathbf{a}_d'\Omega_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\dot{\Omega}_d\mathbf{V}_d^{-1}\Omega_d\mathbf{a}_d.\end{aligned}$$

Finally

$$g_3(\theta) = \text{tr}\left\{\begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22}\end{pmatrix}\begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22}\end{pmatrix}^{-1}\right\},$$

where F_{ab} is the element of the REML Fisher information matrix.

3.2 Area-level model with independent time effects

3.2.1 The model

Let us assume that

$$y_{dt} = \mathbf{x}_{dt}\beta + u_{dt} + e_{dt}, \quad d = 1,\dots,D, \quad t = 1,\dots,m_d, \tag{3.3}$$

where y_{dt} is a direct estimator of the indicator of interest for area d and time instant t, and $\mathbf{x}_{dt}$ is a vector containing the aggregated (population) values of p auxiliary variables. The index d is used for domains and the index t for time instants. We assume that the vectors u_{dt}'s are $N(0,\sigma_u^2)$, the errors e_{dt}'s are independent $N(0,\sigma_{dt}^2)$, and the u_{dt}'s are independent of the e_{dt}'s.

Model (3.3) can be alternatively written in the form

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}, \tag{3.4}$$

where $\mathbf{y} = \underset{1\le d\le D}{\text{col}}(\mathbf{y}_d)$, $\mathbf{y}_d = \underset{1\le t\le m_d}{\text{col}}(y_{dt})$, $\mathbf{u} = \underset{1\le d\le D}{\text{col}}(\mathbf{u}_d)$, $\mathbf{u}_d = \underset{1\le t\le m_d}{\text{col}}(u_{dt})$, $\mathbf{e} = \underset{1\le d\le D}{\text{col}}(\mathbf{e}_d)$, $\mathbf{e}_d = \underset{1\le t\le m_d}{\text{col}}(e_{dt})$, $\mathbf{X} = \underset{1\le d\le D}{\text{col}}(\mathbf{X}_d)$, $\mathbf{X}_d = \underset{1\le t\le m_d}{\text{col}}(\mathbf{x}_{dt})$, $\mathbf{x}_{dt} = \underset{1\le i\le p}{\text{col}'}(x_{dti})$, $\beta = \underset{1\le i\le p}{\text{col}}(\beta_i)$, $\mathbf{Z} = \mathbf{I}_M$, $M = \sum_{d=1}^D m_d$. We assume that $\mathbf{u} \sim N(\mathbf{0},\mathbf{V}_u)$ and $\mathbf{e} \sim N(\mathbf{0},\mathbf{V}_e)$ are independent with covariance matrices

$$\mathbf{V}_u = \sigma_u^2\mathbf{I}_M, \quad \mathbf{I}_M = \underset{1\le d\le D}{\text{diag}}(\mathbf{I}_{m_d}), \quad \mathbf{V}_e = \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_{ed}), \quad \mathbf{V}_{ed} = \underset{1\le t\le m_d}{\text{col}}(\sigma_{dt}^2),$$

and known variances σ_{dt}^2.

The BLUE of β and the BLUP of $\mathbf{u}$ are

$$\widehat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \qquad \text{and} \qquad \widehat{\mathbf{u}} = \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\beta}),$$

where

$$var(\mathbf{y}) = \mathbf{V} = \sigma_u^2 \underset{1\le d\le D}{\text{diag}}(\mathbf{I}_{m_d}) + \mathbf{V}_e = \underset{1\le d\le D}{\text{diag}}(\sigma_u^2\mathbf{I}_{m_d} + \mathbf{V}_{ed}) = \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d).$$

To calculate $\widehat{\beta}$ and $\widehat{\mathbf{u}}$ we apply the formulas

$$\widehat{\beta} = \left(\sum_{d=1}^D \mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{X}_d\right)^{-1}\left(\sum_{d=1}^D \mathbf{X}_d'\mathbf{V}_d^{-1}\mathbf{y}_d\right), \quad \widehat{\mathbf{u}} = \sigma_u^2 \underset{1\le d\le D}{\text{col}}\left(\mathbf{V}_d^{-1}(\mathbf{y}_d - \mathbf{X}_d\widehat{\beta})\right).$$

3.2.2 The Henderson 3 estimator

The Henderson 3 estimator of σ_u^2 is

$$\widehat{\sigma}_{uH}^2 = \frac{\mathbf{y}'\mathbf{P}_2\mathbf{y} - (M-p)}{\text{tr}\{\mathbf{P}_2\}},$$

where

$$\begin{aligned}
\mathbf{P}_2 &= \mathbf{V}_e^{-1} - \mathbf{V}_e^{-1}\mathbf{X}\mathbf{Q}_2\mathbf{X}'\mathbf{V}_e^{-1} = \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_{ed}^{-1}) - \underset{1\le d\le D}{\text{col}}(\mathbf{V}_{ed}^{-1}\mathbf{X}_d)\mathbf{Q}_2 \underset{1\le d\le D}{\text{col}'}(\mathbf{X}_d'\mathbf{V}_{ed}^{-1}),\\
\mathbf{Q}_2 &= (\mathbf{X}'\mathbf{V}_e^{-1}\mathbf{X})^{-1} = \left(\sum_{d=1}^{D}(\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right)^{-1},\\
\text{tr}\{\mathbf{P}_2\} &= \sum_{d=1}^{D}\sum_{t=1}^{m_d}\sigma_{dt}^{-2} - \sum_{d=1}^{D}\text{tr}\{\mathbf{X}_d'\mathbf{V}_{ed}^{-2}\mathbf{X}_d\mathbf{Q}_2\},\\
\mathbf{y}'\mathbf{P}_2\mathbf{y} &= \underset{1\le d\le D}{\text{col}'}(\mathbf{y}_d)\left[\underset{1\le d\le D}{\text{diag}}(\mathbf{V}_{ed}^{-1}) - \underset{1\le d\le D}{\text{col}}(\mathbf{V}_{ed}^{-1}\mathbf{X}_d)\mathbf{Q}_2 \underset{1\le d\le D}{\text{col}'}(\mathbf{X}_d'\mathbf{V}_{ed}^{-1})\right]\underset{1\le d\le D}{\text{col}}(\mathbf{y}_d)\\
&= \sum_{d=1}^{D}\sum_{t=1}^{m_d}\sigma_{dt}^{-2}y_{dt}^2 - \left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right)\mathbf{Q}_2\left(\sum_{d=1}^{D}\mathbf{y}_d'\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right)'.
\end{aligned}$$

3.2.3 The REML estimator

The REML log-likelihood is

$$l_{REML}(\sigma_u^2) = -\frac{M-p}{2}\log 2\pi + \frac{1}{2}\log|\mathbf{X}'\mathbf{X}| - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y},$$

where $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$, $\mathbf{PVP} = \mathbf{P}$ and $\mathbf{PX} = \mathbf{0}$. Let us define $\mathbf{V}_u = \frac{\partial \mathbf{V}}{\partial \sigma_u^2} = \mathbf{I}_M$, $\mathbf{P}_u = \frac{\partial \mathbf{P}}{\partial \sigma_u^2} = -\mathbf{P}\frac{\partial \mathbf{V}}{\partial \sigma_u^2}\mathbf{P} = -\mathbf{P}\mathbf{V}_u\mathbf{P} = -\mathbf{P}^2$. The derivative of l_{REML} with respect to $\theta = \sigma_u^2$ is

$$S = S(\theta) = \frac{\partial l_{REML}}{\partial \theta} = -\frac{1}{2}\text{tr}(\mathbf{P}\mathbf{V}_u) + \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{V}_u\mathbf{P}\mathbf{y} = -\frac{1}{2}\text{tr}(\mathbf{P}) + \frac{1}{2}\mathbf{y}'\mathbf{P}^2\mathbf{y}.$$

The minus expectation of the second order derivative of l_{REML} with respect to $\theta = \sigma_u^2$ is

$$F = F(\theta) = \frac{1}{2}\text{tr}(\mathbf{P}\mathbf{V}_u\mathbf{P}\mathbf{V}_u) = \frac{1}{2}\text{tr}(\mathbf{P}^2). \tag{3.5}$$

The updating formula of the Fisher-scoring algorithm is

$$\theta^{k+1} = \theta^k + F^{-1}(\theta^k)S(\theta^k).$$

The Henderson 3 estimator $\widehat{\sigma}_{uH}^2$ can be used as seed of the Fisher-scoring algorithm. The REML estimator of β is

$$\widehat{\beta}_{REML} = (\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{y}.$$

The asymptotic distributions of the REML estimators of σ_u^2 and β are

$$\hat{\sigma}_u^2 \sim N_2(\theta, F^{-1}(\sigma_u^2)), \quad \hat{\beta} \sim N_p(\beta, (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}).$$

Asymptotic confidence intervals at the level $1-\alpha$ for σ_u^2 and β_i are

$$\hat{\sigma}_u^2 \pm z_{\alpha/2}\nu^{1/2}, \quad \hat{\beta}_i \pm z_{\alpha/2}q_{ii}^{1/2},\ i = 1,\ldots,p,$$

where $\hat{\sigma}_u^2 = \sigma_u^{2,(\kappa)}$, $\nu = F^{-1}(\sigma_u^{2,(\kappa)})$, $(\mathbf{X}'\mathbf{V}^{-1}(\sigma_u^{2,(\kappa)})\mathbf{X})^{-1} = (q_{ij})_{i,j=1,\dots,p}$, κ is the final iteration of the Fisher-scoring algorithm and z_α is the α-quantile of the standard normal distribution $N(0,1)$. Observed $\hat{\beta}_i = \beta_0$, the p-value for testing the hypothesis $H_0: \beta_i = 0$ is

$$p = 2P_{H_0}(\hat{\beta}_i > |\beta_0|) = 2P(N(0,1) > |\beta_0|/\sqrt{q_{ii}}).$$

3.2.4 The EBLUP and its mean squared error

We are interested in predicting $\mu_{dt} = \mathbf{x}_{dt}\beta + u_{dt}$ with the EBLUP $\widehat{\mu}_{dt} = \mathbf{x}_{dt}\widehat{\beta} + \widehat{u}_{dt}$. No taking into account the error e_{dt}, this is equivalent to predict $y_{dt} = \mathbf{a}'\mathbf{y}$, where $\mathbf{a} = \underset{1\le\ell\le D}{\text{col}}(\underset{1\le k\le m_\ell}{\text{col}}(\delta_{d\ell}\delta_{tk}))$ is a vector having one "1" in the cell $t + \sum_{\ell=1}^{d-1} m_\ell$ and "0"'s in the remaining cells. The total $\overline{Y}_{dt}$ is estimated with $\widehat{\overline{Y}}_{dt}^{eblup} = \widehat{\mu}_{dt}$. The mean squared error of $\widehat{\overline{Y}}_{dt}^{eblup}$ is

$$MSE(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\theta) + g_2(\theta) + g_3(\theta),$$

where $\theta = \sigma_u^2$ and

$$\begin{aligned} g_1(\theta) &= \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}, \\ g_2(\theta) &= [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}] \qquad \text{y} \\ g_3(\theta) &\approx \text{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\} \end{aligned}$$

The estimator of $MSE(\widehat{\overline{Y}}_{dt}^{eblup})$ is

$$mse(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\hat{\theta}) + g_2(\hat{\theta}) + 2g_3(\hat{\theta}).$$

Calculation of $g_1(\sigma_u^2)$

We have that $g_1(\sigma_u^2) = \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}$, where $\mathbf{Z} = \mathbf{I}_{M\times M}$ and

$$\mathbf{T} = \mathbf{V}_u - \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{V}_u = \sigma_u^2\mathbf{I}_M - \sigma_u^4 \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d^{-1}).$$

We define $\mathbf{a}_d = \underset{1\le k\le m_d}{\text{col}}(\delta_{tk})$. Then, we have

$$g_1(\sigma_u^2) = \sigma_u^2\mathbf{a}_d'\mathbf{a}_d - \sigma_u^4\mathbf{a}_d'\mathbf{V}_d^{-1}\mathbf{a}_d = \frac{\sigma_u^2\sigma_{dt}^2}{\sigma_u^2 + \sigma_{dt}^2}.$$

Calculation of $g_2(\sigma_u^2)$

We have that $g_2(\sigma_u^2) = [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}]$, where

$$\begin{aligned} \mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X} &= \left[\sigma_u^2\mathbf{I}_M - \sigma_u^4 \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_d^{-1})\right] \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_{ed}^{-1}) \underset{1\le d\le D}{\text{col}}(\mathbf{X}_d) \\ &= \sigma_u^2 \underset{1\le d\le D}{\text{col}}(\mathbf{V}_{ed}^{-1}\mathbf{X}_d) - \sigma_u^4 \underset{1\le d\le D}{\text{col}}(\mathbf{V}_d^{-1}\mathbf{V}_{ed}^{-1}\mathbf{X}_d). \end{aligned}$$

Therefore,

$$\begin{aligned} g_2(\sigma_u^2) &= \left[\mathbf{a}_d'\mathbf{X}_d - \sigma_u^2\mathbf{a}_d'\mathbf{V}_{ed}^{-1}\mathbf{X}_d + \sigma_u^4\mathbf{a}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right]\mathbf{Q} \\ &\cdot \left[\mathbf{X}_d'\mathbf{a}_d - \sigma_u^2\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\mathbf{a}_d + \sigma_u^4\mathbf{X}_d'\mathbf{V}_{ed}^{-1}\mathbf{V}_d^{-1}\mathbf{a}_d\right] \end{aligned}$$

Calculation of $g_3(\sigma_u^2)$

We have that

$$g_3(\sigma_u^2) \approx \operatorname{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\},$$

where

$$\mathbf{b}' = \mathbf{a}'\mathbf{Z}\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1} = \sigma_u^2\mathbf{a}' \operatorname*{diag}_{1\le\ell\le D}(\mathbf{V}_\ell^{-1}) = \sigma_u^2 \operatorname*{col'}_{1\le\ell\le D}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-1}).$$

It holds that

$$\frac{\partial\mathbf{b}'}{\partial\sigma_u^2} = \operatorname*{col'}_{1\le\ell\le D}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-1}) - \sigma_u^2 \operatorname*{col'}_{1\le\ell\le D}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-1}\mathbf{V}_{\ell u}\mathbf{V}_\ell^{-1}), \quad \mathbf{V}_{\ell u} = \frac{\partial\mathbf{V}_\ell}{\partial\sigma_u^2} = \mathbf{I}_{m_\ell}.$$

We define

$$\begin{aligned} q &= \frac{\partial\mathbf{b}'}{\partial\sigma_u^2}\operatorname*{diag}_{1\le\ell\le D}(\mathbf{V}_\ell)\left(\frac{\partial\mathbf{b}'}{\partial\sigma_u^2}\right)' = \mathbf{a}_d'\mathbf{V}_d^{-1}\mathbf{a}_d - 2\sigma_u^2\mathbf{a}_d'\mathbf{V}_d^{-2}\mathbf{a}_d + \sigma_u^4\mathbf{a}_d'\mathbf{V}_d^{-3}\mathbf{a}_d \\ &= \frac{1}{\sigma_u^2+\sigma_{dt}^2} - \frac{2\sigma_u^2}{(\sigma_u^2+\sigma_{dt}^2)^2} + \frac{\sigma_u^4}{(\sigma_u^2+\sigma_{dt}^2)^3}, \end{aligned}$$

Finally, we get

$$g_3(\sigma_u^2) = qF^{-1}(\sigma_u^2),$$

where F is the REML Fisher amount of information calculated in the updating equation of the Fisher-scoring algorithm (cf. (3.5)).

3.3 Estimation of poverty indicators

Let us consider a finite population P_t partitioned into D domains P_{dt} at time period t, and denote their sizes by N_t and N_{dt}, $d = 1,\dots,D$. Let z_{dtj} be an income variable measured in all the units of the population and let z_t be the poverty line, so that units with $z_{dtj} < z_t$ are considered as poor at time period t. The main goal of this section is to estimate the poverty incidence (proportion of individuals under poverty) and the poverty gap in Spanish domains. These two measures belongs to the FGT family proposed by Foster et al. (1984), given by

$$Y_{\alpha;dt} = \frac{1}{N_{dt}}\sum_{j=1}^{N_{dt}} y_{\alpha;dtj}, \quad \text{where } y_{\alpha;dtj} = \left(\frac{z_t - z_{dtj}}{z_t}\right)^\alpha I(z_{dtj} < z_t), \tag{3.6}$$

$I(z_{dtj} < z_t) = 1$ if $z_{dtj} < z_t$ and $I(z_{dtj} < z_t) = 0$ otherwise. The proportion of units under poverty in the domain d and period t is thus $Y_{0;dt}$ and the poverty gap is $Y_{1;dt}$.

We use the data from the Spanish Living Conditions Survey (SLCS) corresponding to years 2004-2006 and described in Tables 1.1. We consider $D = 104$ domains obtained by crossing 52 provinces with 2 sexes. The quartiles of the distribution of the domain sample sizes are $q_0 = 17$, $q_1 = 170$, $q_2 = 293$, $q_3 = 640$, $q_4 = 2113$ 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 of the total $Y_{dt} = \sum_{j=1}^{N_{dt}} y_{dtj}$ is

$$\hat{Y}_{dt}^{dir} = \sum_{j \in S_{dt}} w_{dtj} y_{dtj}.$$

where S_{dt} is the domain sample at time period t and the w_{dtj}'s are the official calibrated sampling weights which take into account for non response. In the particular case $y_{dtj} = 1$, for all $j \in P_{dt}$, we get the estimated domain size

$$\hat{N}_{dt}^{dir} = \sum_{j \in S_{dt}} w_{dtj}.$$

Using this quantity, a direct estimator of the domain mean $\bar{Y}_{dt}$ is $\bar{y}_{dt} = \hat{Y}_{dt}^{dir} / \hat{N}_{dt}^{dir}$. 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 by

$$\hat{V}_{\pi}(\hat{Y}_{dt}^{dir}) = \sum_{j \in S_{dt}} w_{dtj}(w_{dtj} - 1)\left(y_{dtj} - \bar{y}_{dt}\right)^2 \text{ and } \hat{V}_{\pi}(\bar{y}_{dt}) = \hat{V}\left(\hat{Y}_{dt}^{dir}\right)/\hat{N}_{dt}^2. \quad (3.7)$$

The last formulas are obtained from Särndal et al. (1992), pp. 43, 185 and 391, with the simplifications $w_{dtj} = 1/\pi_{dtj}$, $\pi_{dtj,dtj} = \pi_{dtj}$ and $\pi_{dti,dtj} = \pi_{dti}\pi_{dtj}$, $i \neq j$ in the second order inclusion probabilities.

As we are interested in the cases $y_{dtj} = y_{\alpha;dtj}$, $\alpha = 0, 1$, we select the direct estimates of the poverty incidence and poverty gap at domain d and time period t (i.e. $\bar{y}_{0;dt}$ and $\bar{y}_{1;dt}$ respectively) as target variables for the time dependent area-level models 0 and 1. The considered auxiliary variables are the domain means of the category indicators of the variables AGE, EDUCATION, CITIZENSHIP and LABOR described in Section 1.2.

We first consider the linear model

$$\bar{y}_{dt} = \bar{\mathbf{X}}_{dt}\beta + u_{dt} + e_{dt}, \quad d = 1, \ldots, D$$

where $\bar{\mathbf{X}}_d$ 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 $\bar{\mathbf{X}}_d$ contains a "1", so that $p = 1 + 4 + 3 + 1 + 3 = 12$. Random effects errors are assumed to follow the distributional assumptions of model (3.1) either restricted to $\rho = 0$ (model 0) or without this restriction (model 1). As some of the explanatory variables where not significative, the starting models where simplified to include only the auxiliary variables appearing in Tables 3.3.1 and 3.3.2. As the 90% confidence intervals for ρ are 0.8214 ± 0.063 and 0.6862 ± 0.094 for $\alpha = 0$ and $\alpha = 1$ respectively, we recommend model 1 in both cases. Regression parameters and their corresponding p-values are presented in Tables 3.3.1 and 3.3.2.

model 0	*age3*	*age4*	*age5*	*edu1*	*edu2*	*cit1*	*lab2*
β	-0.3686	-2.8841	-0.3649	0.7470	0.3977	0.4808	0.9206
p	0.0143	$< 10^{-35}$	0.0142	$< 10^{-10}$	0.0005	$< 10^{-11}$	$< 10^{-5}$
model 1	*age3*	*age4*	*age5*	*edu1*	*edu2*	*cit1*	*lab2*
β	-0.209	-2.509	-0.075	0.398	0.200	0.531	0.262
p	0.262	$< 10^{-17}$	0.687	0.008	0.196	$< 10^{-7}$	0.310

Table 3.3.1. Regression parameters and p-values for $\alpha = 0$.

model 0	*constant*	*age2*	*age3*	*age4*	*age5*	*edu0*	*edu1*	*cit1*	*lab2*
β	4249.0	-4248.7	-4249.1	-4249.5	-4249.0	-4248.7	0.127	0.117	0.234
p	0.0022	0.0022	0.0022	0.0022	0.0022	0.0022	0.0001	0.0013	0.016
model 1	*constant*	*age2*	*age3*	*age4*	*age5*	*edu0*	*edu1*	*cit1*	*lab2*
β	3731.9	-3731.9	-3732.1	-3732.4	-3731.9	-3731.5	0.062	0.102	0.222
p	0.0039	0.0039	0.0039	0.0039	0.0039	0.0039	0.074	0.029	0.046

Table 3.3.2. Regression parameters and p-values for $\alpha = 1$.

By observing the signs of the regression parameters 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 greater than 25, 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 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 greater in those domains with larger proportion of population characterized by absence of studies, Spanish citizenship and unemployment.

Residuals $\hat{e}_{dt} = \bar{y}_{dt} - \bar{\mathbf{X}}_{dt}\hat{\beta} - \hat{u}_{dt}$ of fitted model 1 are plotted against the observed values $\bar{y}_{dt}$ in the Figure 3.1 for $\alpha = 0$ (top-right) and $\alpha = 1$ (top-left). The dispersion graph shows that EBLUP1 estimates are over and below direct estimates, so that the design unbiased property of the direct estimator is not completely lost by using the model 1. On the right part of the figure we observe that residuals tend to be positive, which means that the model is smoothing the value of the direct estimator larger values. We find that this is an interesting property because it protects us from the presence of outliers in the collection of direct domain estimates.

The three considered estimators of the poverty proportion and gap (direct, EBLUP0 and EBLUP1) are plotted in the Figure 3.2 for $\alpha = 0$ (top-left) and $\alpha = 1$ (top-right). The squared roots of their mean squared error estimates (RMSEs) are plotted in the Figure 3.2 for $\alpha = 0$ (bottom-left) and $\alpha = 1$ (bottom-right). We observe that the EBLUP1 is the one presenting the best results and it is thus the one we recommend. Finally full numerical information is presented in the Table 3.3.3 for the poverty proportion and in the Table 3.3.4 for the poverty gap. In these tables direct, EBLUP0 and EBLUP1 estimates are labeled by dir, eb0 and eb1 respectively.

In the Figure 3.3 the Spanish provinces are plotted in 4 colored categories depending on the values of the EBLUP1 estimates in % of the poverty proportions and the gaps, i.e. $p_d = 100 \cdot \hat{Y}^{eblup1}_{0;d,2006}$ and $g_d = 100 \cdot \hat{Y}^{eblup1}_{1;d,2006}$.

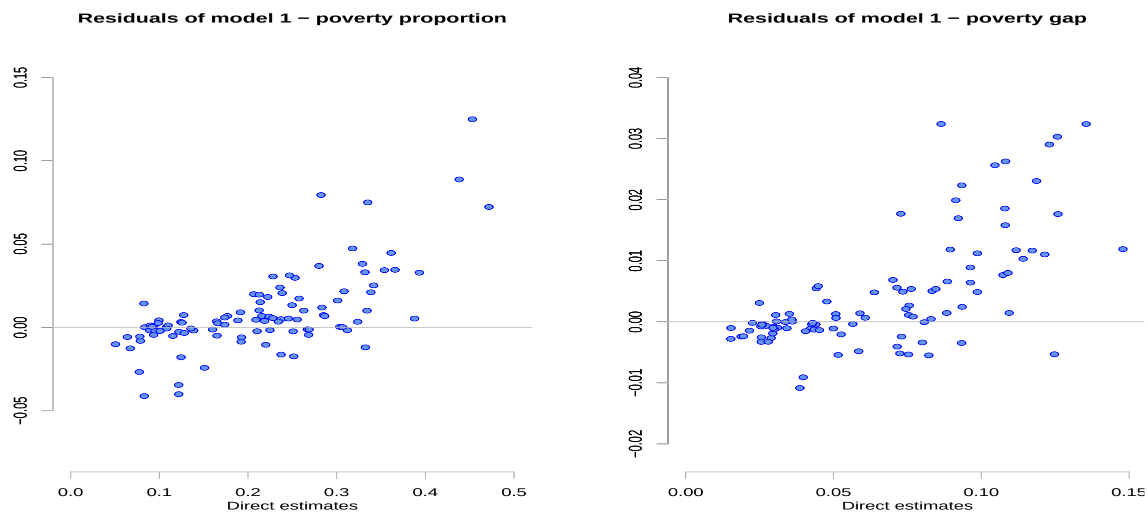


Figure 3.1: Residuals versus direct estimates.

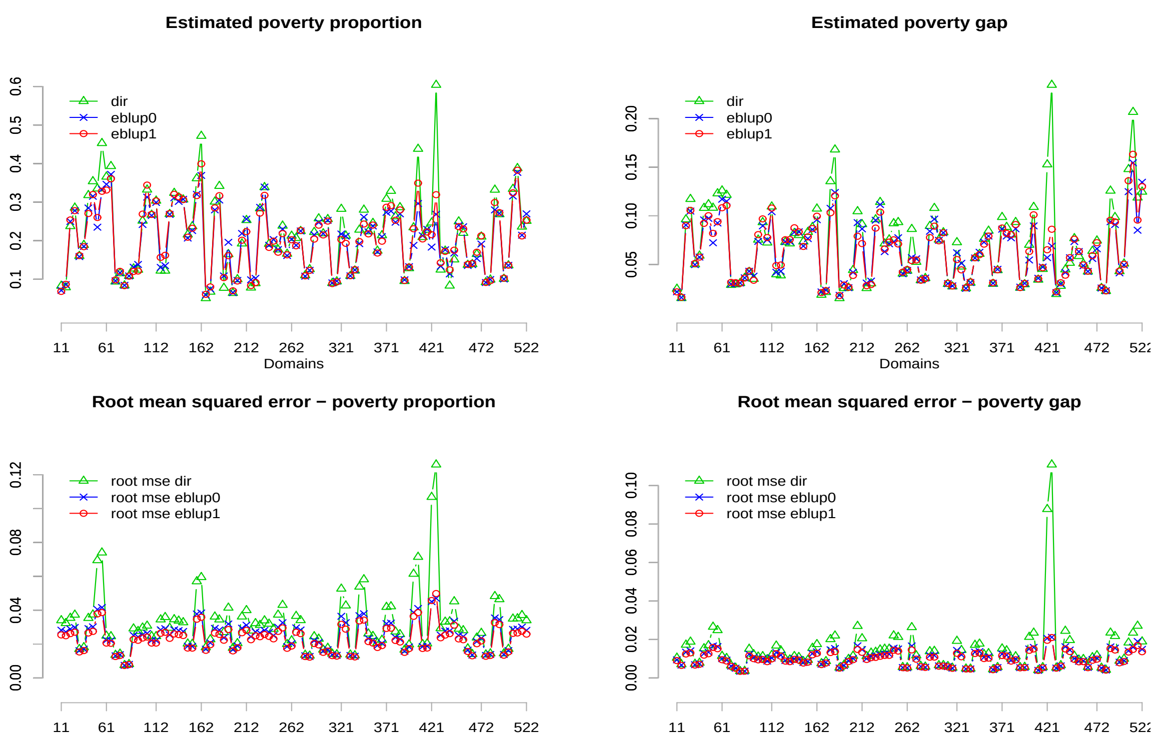


Figure 3.2: Estimates of poverty proportions and gaps (top) and their estimated RMSEs (bottom).

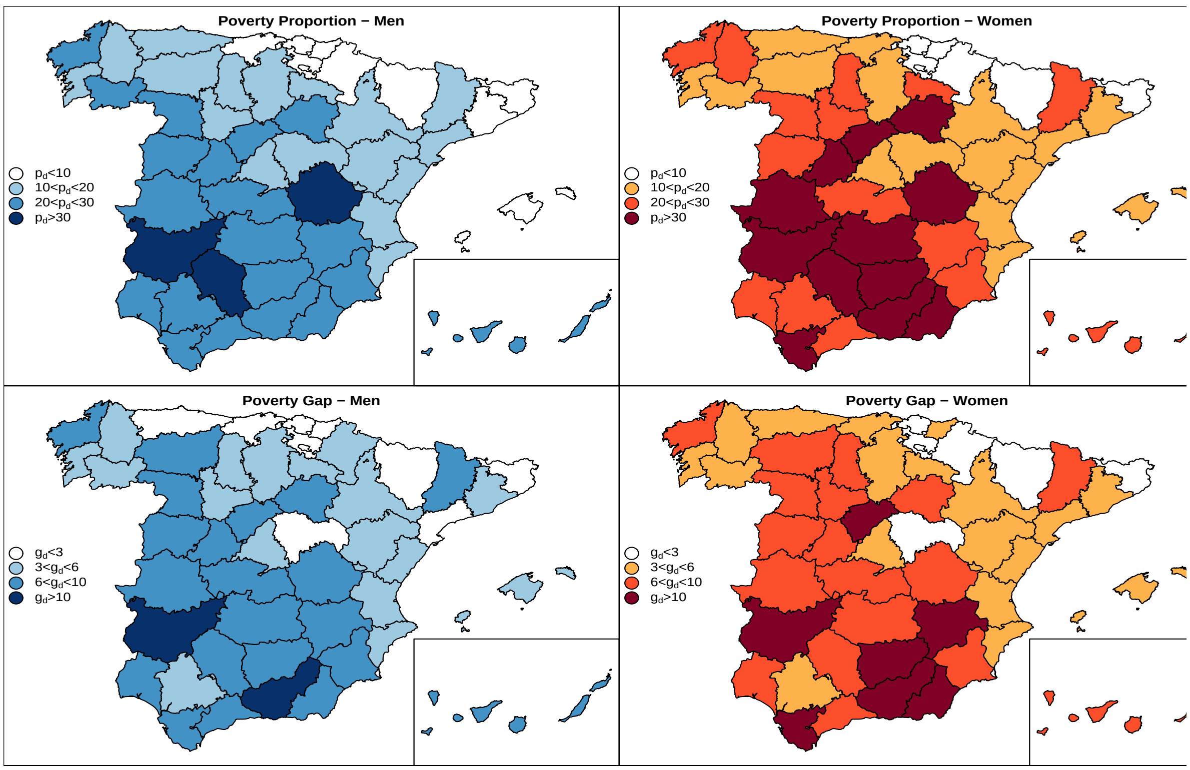


Figure 3.3: Estimates of poverty proportions (top) and gaps (bottom) for men (left) and women (right).

We observe that the Spanish regions where the proportion of the population under the poverty line is smallest are those situated in the north and east, like Cataluña, Aragón, Navarra, País Vasco, Cantabria and Baleares. On the other hand the Spanish regions with higher poverty proportion are those situated in the center-south, like Andalucía, Extremadura, Murcia, Castilla La Mancha, Canarias, Canarias, Ceuta and Melilla. In an intermediate position we can find regions that are in the center-north of Spain, like Galicia, La Rioja, Castilla León, Asturias, Comunidad Valenciana and Madrid. If we investigate how far the annual net incomes of population under the poverty line z_{2006} are from z_{2006}, we observe that in the Spanish regions situated in the center-north there exist a distance that is generally lower than the 6% of z_{2006}. However, the cited distance is in general greater than 6% of z_{2006} in the center-south. It is somehow surprising how large is the amount of Spanish provinces having a proportion greater than 30% of population with annual net incomes below z_{2006}. So it would be desirable that the Spanish Government implements policies to reduce poverty proportion in the center-south of Spain. A criticisms to the use of employed FGT poverty indicators is that they are defined by using only the income variable and do not consider the cost of living. In the case of Spain the south is poorer than the north and this is visualized in the obtained results. On the other hand to live in the south is in general cheaper than to live in the north. As the poverty line is officially calculated for the whole Spain, it is smaller than it should be in the north and greater than it should be in the south.

	men / poverty proportions / women						men / sqrt.mse / women					
d	dir	eb0	eb1	dir	eb0	eb1	dir	eb0	eb1	dir	eb0	eb1
1	0.083	0.071	0.068	0.079	0.086	0.087	0.034	0.028	0.025	0.032	0.027	0.025
2	0.237	0.250	0.254	0.285	0.278	0.278	0.035	0.029	0.026	0.037	0.030	0.027
3	0.160	0.159	0.161	0.189	0.185	0.184	0.017	0.016	0.015	0.018	0.017	0.016
4	0.318	0.285	0.270	0.354	0.315	0.320	0.035	0.029	0.027	0.037	0.030	0.027
5	0.335	0.235	0.260	0.453	0.333	0.328	0.069	0.040	0.038	0.074	0.042	0.039
6	0.366	0.346	0.331	0.393	0.372	0.361	0.025	0.022	0.021	0.025	0.022	0.020
7	0.094	0.096	0.098	0.115	0.117	0.120	0.014	0.013	0.013	0.014	0.014	0.013
8	0.083	0.084	0.083	0.108	0.108	0.108	0.008	0.007	0.007	0.008	0.008	0.008
9	0.127	0.131	0.120	0.124	0.138	0.121	0.029	0.025	0.023	0.028	0.024	0.022
10	0.252	0.242	0.269	0.332	0.313	0.344	0.030	0.026	0.024	0.031	0.026	0.024
11	0.267	0.265	0.268	0.303	0.299	0.303	0.025	0.022	0.021	0.025	0.022	0.021
12	0.122	0.131	0.157	0.122	0.135	0.162	0.034	0.028	0.026	0.036	0.029	0.027
13	0.269	0.268	0.270	0.324	0.311	0.320	0.030	0.025	0.023	0.035	0.028	0.026
14	0.312	0.302	0.314	0.307	0.307	0.306	0.034	0.028	0.026	0.033	0.027	0.025
15	0.216	0.207	0.209	0.237	0.226	0.232	0.020	0.019	0.018	0.020	0.019	0.018
16	0.362	0.320	0.317	0.472	0.369	0.399	0.057	0.038	0.035	0.059	0.038	0.036
17	0.050	0.059	0.061	0.067	0.075	0.080	0.018	0.017	0.016	0.023	0.021	0.020
18	0.301	0.280	0.285	0.342	0.305	0.317	0.036	0.029	0.026	0.034	0.028	0.026
19	0.077	0.108	0.104	0.165	0.196	0.161	0.027	0.024	0.022	0.041	0.032	0.029
20	0.064	0.065	0.070	0.100	0.097	0.095	0.018	0.017	0.016	0.020	0.019	0.018
21	0.192	0.219	0.201	0.253	0.256	0.223	0.036	0.029	0.027	0.040	0.031	0.028
22	0.078	0.098	0.083	0.089	0.102	0.091	0.028	0.024	0.023	0.032	0.027	0.025
23	0.283	0.286	0.272	0.339	0.339	0.318	0.031	0.027	0.024	0.034	0.028	0.026
24	0.192	0.188	0.183	0.193	0.203	0.199	0.032	0.027	0.025	0.029	0.025	0.023
25	0.177	0.177	0.170	0.239	0.230	0.218	0.037	0.030	0.027	0.043	0.033	0.030
26	0.166	0.161	0.163	0.212	0.204	0.202	0.020	0.019	0.017	0.022	0.020	0.019
27	0.207	0.188	0.187	0.225	0.225	0.226	0.037	0.030	0.027	0.034	0.028	0.026
28	0.110	0.108	0.109	0.126	0.123	0.123	0.014	0.013	0.013	0.013	0.013	0.012
29	0.222	0.215	0.204	0.258	0.247	0.240	0.025	0.022	0.020	0.023	0.021	0.020
30	0.219	0.218	0.215	0.256	0.253	0.251	0.017	0.016	0.015	0.018	0.017	0.016
31	0.090	0.090	0.089	0.094	0.095	0.093	0.014	0.014	0.013	0.014	0.013	0.013
32	0.282	0.217	0.203	0.213	0.212	0.193	0.053	0.036	0.032	0.043	0.033	0.029
33	0.108	0.109	0.109	0.122	0.124	0.125	0.014	0.013	0.013	0.013	0.013	0.012
34	0.228	0.192	0.198	0.280	0.261	0.243	0.054	0.037	0.034	0.058	0.038	0.034
35	0.224	0.223	0.218	0.246	0.238	0.240	0.026	0.023	0.021	0.025	0.022	0.021
36	0.174	0.171	0.168	0.214	0.210	0.199	0.021	0.019	0.018	0.022	0.020	0.019
37	0.308	0.274	0.287	0.329	0.277	0.291	0.042	0.032	0.029	0.042	0.032	0.029
38	0.263	0.248	0.253	0.286	0.270	0.280	0.027	0.024	0.022	0.026	0.023	0.021
39	0.095	0.098	0.097	0.128	0.132	0.132	0.017	0.016	0.015	0.020	0.018	0.017
40	0.234	0.188	0.231	0.438	0.296	0.349	0.061	0.039	0.036	0.071	0.041	0.039
41	0.209	0.213	0.204	0.228	0.229	0.222	0.020	0.019	0.017	0.020	0.019	0.018
42	0.247	0.183	0.216	0.604	0.268	0.319	0.107	0.045	0.046	0.126	0.047	0.050
43	0.125	0.134	0.143	0.174	0.176	0.172	0.029	0.025	0.024	0.033	0.028	0.025
44	0.083	0.114	0.124	0.151	0.168	0.175	0.033	0.028	0.026	0.045	0.033	0.031
45	0.250	0.239	0.236	0.220	0.236	0.230	0.029	0.025	0.023	0.028	0.024	0.023
46	0.137	0.136	0.138	0.139	0.138	0.141	0.017	0.016	0.016	0.014	0.014	0.013
47	0.165	0.154	0.170	0.210	0.191	0.213	0.024	0.021	0.020	0.027	0.023	0.022
48	0.092	0.091	0.092	0.099	0.099	0.096	0.014	0.013	0.013	0.014	0.014	0.013
49	0.332	0.280	0.299	0.268	0.271	0.273	0.048	0.035	0.033	0.046	0.034	0.032
50	0.101	0.100	0.103	0.136	0.136	0.136	0.014	0.014	0.013	0.017	0.016	0.015
51	0.334	0.316	0.324	0.388	0.377	0.383	0.035	0.029	0.026	0.035	0.029	0.027
52	0.236	0.214	0.212	0.251	0.270	0.253	0.037	0.030	0.028	0.034	0.028	0.026

Table 3.3.3. Estimated domain poverty proportions and their estimated squared root MSE’s by sex.

	men / poverty gaps / women						men / sqrt.mse / women					
d	dir	eb0	eb1	dir	eb0	eb1	dir	eb0	eb1	dir	eb0	eb1
1	0.025	0.022	0.022	0.015	0.016	0.016	0.010	0.009	0.009	0.007	0.007	0.007
2	0.096	0.090	0.090	0.117	0.105	0.106	0.017	0.013	0.012	0.019	0.014	0.013
3	0.050	0.050	0.051	0.059	0.058	0.058	0.007	0.007	0.007	0.008	0.007	0.007
4	0.108	0.096	0.092	0.112	0.097	0.100	0.015	0.013	0.012	0.017	0.013	0.013
5	0.108	0.072	0.082	0.123	0.093	0.094	0.027	0.017	0.016	0.025	0.016	0.015
6	0.126	0.117	0.108	0.121	0.115	0.110	0.011	0.010	0.009	0.010	0.009	0.009
7	0.029	0.030	0.032	0.029	0.030	0.031	0.006	0.006	0.006	0.005	0.005	0.005
8	0.031	0.031	0.031	0.036	0.036	0.036	0.003	0.003	0.003	0.004	0.004	0.004
9	0.042	0.043	0.043	0.035	0.038	0.034	0.015	0.012	0.012	0.012	0.011	0.010
10	0.075	0.074	0.081	0.093	0.090	0.097	0.011	0.010	0.009	0.011	0.010	0.010
11	0.072	0.075	0.078	0.109	0.104	0.108	0.010	0.009	0.008	0.012	0.010	0.010
12	0.040	0.042	0.049	0.039	0.043	0.049	0.017	0.013	0.013	0.014	0.012	0.011
13	0.073	0.074	0.075	0.072	0.074	0.076	0.010	0.009	0.009	0.010	0.009	0.008
14	0.082	0.083	0.088	0.080	0.084	0.083	0.011	0.010	0.010	0.011	0.010	0.009
15	0.073	0.069	0.069	0.083	0.077	0.078	0.009	0.008	0.008	0.009	0.009	0.008
16	0.088	0.086	0.087	0.107	0.096	0.100	0.016	0.013	0.012	0.018	0.014	0.013
17	0.019	0.022	0.021	0.022	0.024	0.023	0.008	0.007	0.007	0.009	0.008	0.008
18	0.135	0.108	0.103	0.168	0.124	0.121	0.020	0.015	0.013	0.022	0.015	0.014
19	0.015	0.018	0.018	0.026	0.030	0.029	0.005	0.005	0.005	0.007	0.006	0.006
20	0.026	0.026	0.027	0.044	0.042	0.039	0.010	0.009	0.009	0.011	0.010	0.009
21	0.105	0.093	0.079	0.091	0.087	0.071	0.027	0.017	0.015	0.021	0.015	0.013
22	0.026	0.031	0.028	0.030	0.033	0.029	0.011	0.010	0.010	0.013	0.011	0.011
23	0.096	0.095	0.087	0.114	0.112	0.104	0.013	0.011	0.011	0.015	0.012	0.011
24	0.071	0.063	0.066	0.076	0.071	0.073	0.015	0.012	0.012	0.015	0.012	0.012
25	0.092	0.073	0.075	0.093	0.077	0.071	0.022	0.015	0.014	0.021	0.015	0.014
26	0.041	0.040	0.042	0.043	0.044	0.045	0.006	0.005	0.005	0.005	0.005	0.005
27	0.086	0.057	0.054	0.053	0.056	0.055	0.026	0.017	0.015	0.012	0.010	0.010
28	0.034	0.034	0.034	0.036	0.036	0.035	0.006	0.006	0.006	0.006	0.006	0.005
29	0.090	0.084	0.078	0.108	0.097	0.089	0.014	0.012	0.011	0.014	0.012	0.011
30	0.075	0.075	0.074	0.083	0.083	0.083	0.007	0.006	0.006	0.007	0.006	0.006
31	0.030	0.030	0.031	0.027	0.028	0.028	0.006	0.006	0.006	0.005	0.005	0.005
32	0.073	0.062	0.055	0.048	0.051	0.044	0.019	0.014	0.013	0.014	0.012	0.011
33	0.025	0.025	0.026	0.031	0.031	0.032	0.005	0.005	0.005	0.005	0.005	0.005
34	0.056	0.057	0.057	0.061	0.063	0.060	0.017	0.013	0.013	0.018	0.014	0.013
35	0.076	0.073	0.071	0.085	0.079	0.079	0.012	0.011	0.010	0.013	0.011	0.010
36	0.030	0.031	0.031	0.044	0.045	0.044	0.004	0.004	0.004	0.006	0.005	0.005
37	0.099	0.086	0.087	0.089	0.079	0.082	0.015	0.012	0.012	0.014	0.012	0.011
38	0.081	0.077	0.081	0.093	0.087	0.091	0.010	0.009	0.009	0.011	0.010	0.009
39	0.026	0.027	0.026	0.030	0.031	0.031	0.006	0.005	0.005	0.006	0.005	0.005
40	0.070	0.055	0.063	0.109	0.090	0.101	0.021	0.015	0.014	0.023	0.016	0.015
41	0.034	0.036	0.035	0.045	0.047	0.047	0.004	0.004	0.004	0.006	0.005	0.005
42	0.153	0.057	0.065	0.235	0.069	0.086	0.088	0.021	0.020	0.111	0.021	0.021
43	0.019	0.021	0.022	0.028	0.030	0.031	0.005	0.005	0.005	0.007	0.006	0.006
44	0.045	0.042	0.039	0.052	0.057	0.057	0.024	0.016	0.015	0.020	0.014	0.014
45	0.077	0.074	0.076	0.059	0.063	0.063	0.011	0.010	0.010	0.009	0.009	0.008
46	0.051	0.049	0.049	0.043	0.043	0.043	0.010	0.009	0.008	0.006	0.005	0.005
47	0.064	0.056	0.059	0.074	0.066	0.072	0.011	0.010	0.009	0.012	0.010	0.010
48	0.026	0.026	0.026	0.023	0.023	0.023	0.005	0.005	0.005	0.004	0.004	0.004
49	0.126	0.094	0.095	0.099	0.091	0.094	0.024	0.016	0.015	0.022	0.016	0.015
50	0.043	0.041	0.043	0.051	0.049	0.050	0.009	0.008	0.008	0.010	0.009	0.009
51	0.148	0.125	0.136	0.207	0.155	0.163	0.018	0.014	0.013	0.023	0.016	0.015
52	0.119	0.085	0.096	0.125	0.135	0.130	0.027	0.019	0.017	0.019	0.015	0.014

Table 3.3.4. Estimated domain poverty gaps and their estimated squared root MSE's by sex.

Chapter 4

Partitioned area level time model

4.1 The partitioned model

4.1.1 The model

Let us consider the model ($model1$)

$$y_{dt} = \mathbf{x}_{dt}\beta + u_{dt} + e_{dt}, \quad d = 1,\dots,D = D_A + D_B, \quad t = 1,\dots,m_d, \tag{4.1}$$

where y_{dt} is a direct estimator of the indicator of interest for area d and time instant t, and $\mathbf{x}_{dt}$ is a vector containing the aggregated (population) values of p auxiliary variables. The index d is used for domains and the index t for time instants. We assume that the random effects u_{dt}'s are i.i.d. $N(0,\sigma_A^2)$ if $d \le D_A$ and i.i.d. $N(0,\sigma_B^2)$ if $d > D_A$. We further assume that the errors e_{dt}'s are independent $N(0,\sigma_{dt}^2)$ with known σ_{dt}^2's. Finally we assume that the u_{dt}'s and the e_{dt}'s are mutually independent. In matrix notation the model is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where vectors $\mathbf{y}$, $\mathbf{u}$ and $\mathbf{e}$ can be decomposed in the form $\mathbf{v} = (\mathbf{v}_A', \mathbf{v}_B')'$, with $\mathbf{v}_A = \underset{d \le D_A}{\text{col}}(\mathbf{v}_d)$, $\mathbf{v}_B = \underset{d > D_A}{\text{col}}(\mathbf{v}_d)$ and $\mathbf{v}_d = \underset{1 \le t \le m_d}{\text{col}}(v_{dt})$, matrix $\mathbf{X}$ can be similarly decomposed in the form $\mathbf{X} = (\mathbf{X}_A', \mathbf{X}_B')'$, with $\mathbf{X}_A = \underset{d \le D_A}{\text{col}}(\mathbf{X}_d)$, $\mathbf{X}_B = \underset{d > D_A}{\text{col}}(\mathbf{X}_d)$, $\mathbf{X}_d = \underset{1 \le t \le m_d}{\text{col}}(\mathbf{x}_{dt})$, $\mathbf{x}_{dt} = \underset{1 \le j \le p}{\text{col}'}(x_{dtj})$, $\beta = \beta_{p\times 1}$, $\mathbf{Z} = \mathbf{I}_M$, $M = M_A + M_B$, $M_A = \sum_{d \le D_A} m_d$, $M_B = \sum_{d > D_A} m_d$ an $\mathbf{I}_M$ denotes the identity $M \times M$ matrix. In this notation, $\mathbf{u} \sim N(\mathbf{0}, \mathbf{V}_u)$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{V}_e)$ are independent with covariance matrices

$$\mathbf{V}_u = \text{var}(\mathbf{u}) = \text{diag}(\sigma_A^2 \mathbf{I}_{M_A}, \sigma_B^2 \mathbf{I}_{M_B}), \quad \mathbf{V}_e = \text{var}(\mathbf{e}) = \underset{1 \le d \le D}{\text{diag}}(\mathbf{V}_{ed}), \quad \mathbf{V}_{ed} = \underset{1 \le t \le m_d}{\text{diag}}(\sigma_{dt}^2).$$

The covariance matrix of vector $\mathbf{y}$ is $\mathbf{V} = \text{var}(\mathbf{y}) = \text{diag}(\mathbf{V}_A, \mathbf{V}_B)$, where $\mathbf{V}_A = \underset{d \le D_A}{\text{diag}}(\mathbf{V}_d)$, $\mathbf{V}_B = \underset{d > D_A}{\text{diag}}(\mathbf{V}_d)$, $\mathbf{V}_d = \sigma_A^2 \mathbf{I}_{m_d} + \mathbf{V}_{ed}$ if $d \le D_A$ and $\mathbf{V}_d = \sigma_B^2 \mathbf{I}_{m_d} + \mathbf{V}_{ed}$ if $d > D_A$.

If $\sigma_A^2 > 0$ and $\sigma_B^2 > 0$ are known, the best linear unbiased estimator (BLUE) of β is

$$\widehat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

and the best linear unbiased predictor (BLUP) of $\mathbf{u}$ is

$$\widehat{\mathbf{u}} = \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\widehat{\beta}) = \text{diag}\left(\sigma_A^2\mathbf{I}_{M_A},\sigma_B^2\mathbf{I}_{M_B}\right)\underset{1\le d\le D}{\text{col}}(\mathbf{V}_d^{-1})(\mathbf{y}-\mathbf{X}\widehat{\beta}),$$

so that

$$\widehat{\mathbf{u}}_d = \begin{cases} \sigma_A^2\mathbf{V}_d^{-1}(\mathbf{y}_d-\mathbf{X}_d\widehat{\beta}), & d=1,\dots,D_A, \\ \sigma_B^2\mathbf{V}_d^{-1}(\mathbf{y}_d-\mathbf{X}_d\widehat{\beta}), & d=D_A+1,\dots,D, \end{cases}$$

or equivalently

$$\hat{u}_{dt} = \left[\frac{\sigma_A^2}{\sigma_A^2+\sigma_{dt}^2}I_{\{d\le D_A\}}(d)+\frac{\sigma_B^2}{\sigma_B^2+\sigma_{dt}^2}I_{\{d>D_A\}}(d)\right](y_{dt}-\mathbf{x}_{dt}\hat{\beta}),\ d=1,\dots,D,\ t=1,\dots,m_d.$$

4.1.2 The REML estimators

The loglikelihood of the restricted (residual) maximum likelihood method is

$$\begin{aligned} l_{reml} &= l_{reml}(\sigma_A^2,\sigma_B^2) = -\frac{M-p}{2}\log 2\pi + \frac{1}{2}\log|\mathbf{X}'\mathbf{X}| - \frac{1}{2}\log|\mathbf{V}_A| - \frac{1}{2}\log|\mathbf{V}_B| \\ &- \frac{1}{2}\log|\mathbf{X}_A'\mathbf{V}_A^{-1}\mathbf{X}_A+\mathbf{X}_B'\mathbf{V}_B^{-1}\mathbf{X}_B| - \frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y}, \end{aligned}$$

where

$$\mathbf{P} = \mathbf{V}^{-1}-\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}, \quad \mathbf{PVP}=\mathbf{P}, \quad \mathbf{PX}=\mathbf{0}.$$

Let $\theta=(\theta_1,\theta_2)=(\sigma_A^2,\sigma_B^2)$, then

$$\mathbf{V}_1 = \frac{\partial\mathbf{V}}{\partial\sigma_A^2} = \text{diag}\big(\mathbf{I}_{M_A},\underset{d>D_A}{\text{diag}}(\mathbf{0}_{m_d\times m_d})\big), \quad \mathbf{V}_2 = \frac{\partial\mathbf{V}}{\partial\sigma_B^2} = \text{diag}\big(\underset{d\le D_A}{\text{diag}}(\mathbf{0}_{m_d\times m_d}),\mathbf{I}_{M_B}\big).$$

Then

$$\mathbf{P}_a = \frac{\partial\mathbf{P}}{\partial\theta_a} = -\mathbf{P}\frac{\partial\mathbf{V}}{\partial\theta_a}\mathbf{P} = -\mathbf{P}\mathbf{V}_a\mathbf{P}, \quad a=1,2.$$

By taking partial derivatives of l_{reml} with respect to θ_a, we get the scores

$$S_a = \frac{\partial l_{reml}}{\partial\theta_a} = -\frac{1}{2}\text{tr}(\mathbf{P}\mathbf{V}_a)+\frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{y}, \quad a=1,2.$$

By taking again partial derivatives with respect to θ_a and θ_b, taking expectations and changing the sign, we get the Fisher information matrix components

$$F_{ab} = \frac{1}{2}\text{tr}(\mathbf{P}\mathbf{V}_a\mathbf{P}\mathbf{V}_b), \quad a,b=1,2.$$

To calculate the REML estimate we apply the Fisher-scoring algorithm with the updating fórmula

$$\theta^{k+1} = \theta^k + \mathbf{F}^{-1}(\theta^k)\mathbf{S}(\theta^k),$$

where **S** and **F** are the column vector of scores and the Fisher information matrix respectively. As seeds we use $\sigma_A^{2(0)} = \sigma_B^{2(0)} = \widehat{\sigma}_{uH}^2$, where $\widehat{\sigma}_{uH}^2$ is the Henderson 3 estimator under model with $\sigma_A^2 = \sigma_B^2$. The REML estimator of β is

$$\widehat{\beta}_{reml} = (\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{y}.$$

The asymptotic distributions of the REML estimators of θ and β are

$$\hat{\theta} \sim N_2(\theta, \mathbf{F}^{-1}(\theta)), \quad \hat{\beta} \sim N_p(\beta, (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}).$$

Asymptotic confidence intervals at the level $1-\alpha$ for θ_a and β_j are

$$\hat{\theta}_a \pm z_{\alpha/2}\nu_{aa}^{1/2},\ a = 1,2, \quad \hat{\beta}_j \pm z_{\alpha/2}\, q_{jj}^{1/2},\ j = 1,\dots,p,$$

where $\hat{\theta} = \theta^{\kappa}$, $\mathbf{F}^{-1}(\theta^{\kappa}) = (\nu_{ab})_{a,b=1,2}$, $(\mathbf{X}'\mathbf{V}^{-1}(\theta^{\kappa})\mathbf{X})^{-1} = (q_{ij})_{i,j=1,\dots,p}$, κ is the final iteration of the Fisher-scoring algorithm and z_α is the α-quantile of the standard normal distribution $N(0,1)$). Observed $\hat{\beta}_j = \beta_0$, the p-value for testing the hypothesis $H_0:\ \beta_j = 0$ is

$$p = 2P_{H_0}(\hat{\beta}_j > |\beta_0|) = 2P(N(0,1) > |\beta_0|/\sqrt{q_{jj}}).$$

4.1.3 The EBLUP and its mean squared error

We are interested in predicting the value of $\mu_{dt} = \mathbf{x}_{dt}\beta + u_{dt}$ by using the EBLUP $\widehat{\mu}_{dt} = \mathbf{x}_{dt}\widehat{\beta} + \widehat{u}_{dt}$. If we do not take into account the error, e_{dt}, this is equivalent to predict $y_{dt} = \mathbf{a}'\mathbf{y}$, where $\mathbf{a} = \underset{1\le\ell\le D}{\text{col}}(\underset{1\le k\le m_\ell}{\text{col}}(\delta_{d\ell}\delta_{tk}))$ is a vector having one 1 in the position $t + \sum_{\ell=1}^{d-1} m_\ell$ and 0's in the remaining cells. To estimate $\overline{Y}_{dt}$ we use $\widehat{\overline{Y}}_{dt}^{eblup} = \widehat{\mu}_{dt}$. The mean squared error of $\widehat{\overline{Y}}_{dt}^{eblup}$ is

$$MSE(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\theta) + g_2(\theta) + g_3(\theta),$$

where $\theta = (\sigma_A^2, \sigma_B^2)$,

$$\begin{aligned} g_1(\theta) &= \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}, \\ g_2(\theta) &= [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}], \\ g_3(\theta) &\approx \text{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\} \end{aligned}$$

The estimator of $MSE(\widehat{\overline{Y}}_{dt}^{eblup})$ is

$$mse(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\hat{\theta}) + g_2(\hat{\theta}) + 2g_3(\hat{\theta}).$$

Calculation of $g_1(\theta)$

In the formula of $g_1(\theta) = \mathbf{a}'\mathbf{Z}\mathbf{T}\mathbf{Z}'\mathbf{a}$, we have that $\mathbf{Z} = \mathbf{I}_M$, and $\mathbf{T} = \mathbf{V}_u - \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{V}_u = \text{diag}(\mathbf{T}_A, \mathbf{T}_B)$, where

$$\mathbf{T}_A = \sigma_A^2\mathbf{I}_{M_A} - \sigma_A^4 \underset{d\le D_A}{\text{diag}}(\mathbf{V}_d^{-1}), \quad \mathbf{T}_B = \sigma_B^2\mathbf{I}_{M_B} - \sigma_B^4 \underset{d> D_A}{\text{diag}}(\mathbf{V}_d^{-1}).$$

Let us write $\mathbf{a}_d = \underset{1\leq k\leq m_d}{\text{col}}(\delta_{tk})$. Then, $g_1(\theta)$ can be expressed in the form

$$g_1(\theta) = \begin{cases} \sigma_A^2 - \sigma_A^4 \mathbf{a}_d' \mathbf{V}_d^{-1} \mathbf{a}_d = \frac{\sigma_A^2 \sigma_{dt}^2}{\sigma_A^2+\sigma_{dt}^2} & \text{if } d \leq D_A, \\ \sigma_B^2 - \sigma_B^4 \mathbf{a}_d' \mathbf{V}_d^{-1} \mathbf{a}_d = \frac{\sigma_B^2 \sigma_{dt}^2}{\sigma_B^2+\sigma_{dt}^2} & \text{if } d > D_A. \end{cases}$$

Calculation of $g_2(\theta)$

We have that $g_2(\theta) = [\mathbf{a}'\mathbf{X} - \mathbf{a}'\mathbf{ZTZ}'\mathbf{V}_e^{-1}\mathbf{X}]\mathbf{Q}[\mathbf{X}'\mathbf{a} - \mathbf{X}'\mathbf{V}_e^{-1}\mathbf{ZTZ}'\mathbf{a}]$, where $\mathbf{ZTZ}'\mathbf{V}_e^{-1}\mathbf{X} =$

$$\begin{aligned} \mathbf{ZTZ}'\mathbf{V}_e^{-1}\mathbf{X} &= \begin{pmatrix} \left[\sigma_A^2 \mathbf{I}_{M_A} - \sigma_A^4 \underset{d\leq D_A}{\text{diag}}(\mathbf{V}_d^{-1})\right] \underset{d\leq D_A}{\text{diag}}(\mathbf{V}_{ed}^{-1}) \underset{d\leq D_A}{\text{col}}(\mathbf{X}_d) \\ \left[\sigma_B^2 \mathbf{I}_{M_B} - \sigma_B^4 \underset{d> D_A}{\text{diag}}(\mathbf{V}_d^{-1})\right] \underset{d> D_A}{\text{diag}}(\mathbf{V}_{ed}^{-1}) \underset{d> D_A}{\text{col}}(\mathbf{X}_d) \end{pmatrix} \\ &= \begin{pmatrix} \sigma_A^2 \underset{d\leq D_A}{\text{col}}(\mathbf{V}_{ed}^{-1}\mathbf{X}_d) - \sigma_A^4 \underset{d\leq D_A}{\text{col}}(\mathbf{V}_d^{-1}\mathbf{V}_{ed}^{-1}\mathbf{X}_d) \\ \sigma_B^2 \underset{d> D_A}{\text{col}}(\mathbf{V}_{ed}^{-1}\mathbf{X}_d) - \sigma_B^4 \underset{d> D_A}{\text{col}}(\mathbf{V}_d^{-1}\mathbf{V}_{ed}^{-1}\mathbf{X}_d) \end{pmatrix}. \end{aligned}$$

Therefore

$$\begin{aligned} g_2(\theta) &= \left[\mathbf{a}_d'\mathbf{X}_d - \sigma_A^2 \mathbf{a}_d'\mathbf{V}_{ed}^{-1}\mathbf{X}_d + \sigma_A^4 \mathbf{a}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right]\mathbf{Q} \\ &\cdot \left[\mathbf{X}_d'\mathbf{a}_d - \sigma_A^2 \mathbf{X}_d'\mathbf{V}_{ed}^{-1}\mathbf{a}_d + \sigma_A^4 \mathbf{X}_d'\mathbf{V}_{ed}^{-1}\mathbf{V}_d^{-1}\mathbf{a}_d\right] \quad \text{if } d \leq D_A, \\ &= \left[\mathbf{a}_d'\mathbf{X}_d - \sigma_B^2 \mathbf{a}_d'\mathbf{V}_{ed}^{-1}\mathbf{X}_d + \sigma_B^4 \mathbf{a}_d'\mathbf{V}_d^{-1}\mathbf{V}_{ed}^{-1}\mathbf{X}_d\right]\mathbf{Q} \\ &\cdot \left[\mathbf{X}_d'\mathbf{a}_d - \sigma_B^2 \mathbf{X}_d'\mathbf{V}_{ed}^{-1}\mathbf{a}_d + \sigma_B^4 \mathbf{X}_d'\mathbf{V}_{ed}^{-1}\mathbf{V}_d^{-1}\mathbf{a}_d\right] \quad \text{if } d > D_A. \end{aligned}$$

Calculation of $g_3(\theta)$

We have that

$$g_3(\theta) \approx tr\left\{(\nabla\mathbf{b}')\mathbf{V}(\nabla\mathbf{b}')' E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\},$$

where $\mathbf{b}' = \mathbf{a}'\mathbf{Z}\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1} = \mathbf{a}'\text{diag}\big(\sigma_A^2\mathbf{I}_{M_A}, \sigma_B^2\mathbf{I}_{M_B}\big) \underset{1\leq\ell\leq D}{\text{diag}}(\mathbf{V}_\ell^{-1}) = (\mathbf{b}_A', \mathbf{b}_B')$,

$$\mathbf{b}_A' = \sigma_A^2 \underset{\ell\leq D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-1}) \quad \text{and} \quad \mathbf{b}_B' = \sigma_B^2 \underset{\ell> D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-1}).$$

It holds that $\frac{\partial\mathbf{b}'}{\partial\sigma_A^2} = (\frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2}, \mathbf{0})$, $\frac{\partial\mathbf{b}'}{\partial\sigma_B^2} = (\mathbf{0}, \frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2})$, where

$$\begin{aligned} \frac{\partial\mathbf{b}_A'}{\partial\sigma_A^2} &= \underset{\ell\leq D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-1}) - \sigma_A^2 \underset{\ell\leq D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-2}), \\ \frac{\partial\mathbf{b}_B'}{\partial\sigma_B^2} &= \underset{\ell> D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-1}) - \sigma_B^2 \underset{\ell> D_A}{\text{col}'}(\delta_{d\ell}\mathbf{a}_\ell'\mathbf{V}_\ell^{-2}). \end{aligned}$$

We define

$$\begin{aligned} q_{11} &= \frac{\partial \mathbf{b}'_A}{\partial \sigma^2_A} \operatorname{diag}_{\ell \le D_A}(\mathbf{V}_\ell) \left(\frac{\partial \mathbf{b}'_A}{\partial \sigma^2_A}\right)' = \left[\mathbf{a}'_d \mathbf{V}_d^{-1}\mathbf{a}_d - 2\sigma^2_A \mathbf{a}'_d \mathbf{V}_d^{-2}\mathbf{a}_d + \sigma^4_A \mathbf{a}'_d \mathbf{V}_d^{-3}\mathbf{a}_d\right] I_{\{d\le D_A\}}(d) \\ &= \frac{\sigma^4_{dt}}{(\sigma^2_A+\sigma^2_{dt})^3} I_{\{d\le D_A\}}(d) \\ q_{22} &= \frac{\partial \mathbf{b}'_B}{\partial \sigma^2_B} \operatorname{diag}_{\ell > D_A}(\mathbf{V}_\ell) \left(\frac{\partial \mathbf{b}'_B}{\partial \sigma^2_B}\right)' = \left[\mathbf{a}'_d \mathbf{V}_d^{-1}\mathbf{a}_d - 2\sigma^2_B \mathbf{a}'_d \mathbf{V}_d^{-2}\mathbf{a}_d + \sigma^4_B \mathbf{a}'_d \mathbf{V}_d^{-3}\mathbf{a}_d\right] I_{\{d> D_A\}}(d) \\ &= \frac{\sigma^4_{dt}}{(\sigma^2_B+\sigma^2_{dt})^3} I_{\{d> D_A\}}(d). \end{aligned}$$

Finally

$$g_3(\theta) = \begin{cases} q_{11}F_{11}^{-1}, & \text{if } d \le D_A \\ q_{22}F_{22}^{-1}, & \text{if } d > D_A, \end{cases}$$

where F_{ab} is the element of the REML Fisher information matrix.

4.1.4 Testing for $H_0 : \sigma^2_A = \sigma^2_B$

Let $\hat\sigma^2_A$ and $\hat\sigma^2_B$ be the unrestricted REML estimators of σ^2_A and σ^2_B respectively. Let $\tilde\sigma^2_u$ be the REML estimator of the common value $\sigma^2_A = \sigma^2_B$ under H_0. The REML likelihood ratio statistic (LRS) for testing $H_0 : \sigma^2_A = \sigma^2_B$ is

$$\begin{aligned} \lambda &= -2[l_{REML}(\tilde\sigma^2_u) - l_{REML}(\hat\sigma^2_A, \hat\sigma^2_B)] = \log\frac{|\tilde V|}{|\hat V|} + \log\frac{|X'\tilde V^{-1}X|}{|X'\hat V^{-1}X|} + y'\tilde P y - y'\hat P y \\ &= \log|\tilde V| - \log|\hat V_A| - \log|\hat V_B| + \log|X'\tilde V^{-1}X| \\ &- \log|X'_A\hat V_A^{-1}X_A + X'_B \hat V_B^{-1}X_B| + y'\tilde P y - y'\hat P y. \end{aligned}$$

Asymptotic distribution of λ under H_0 is χ^2_1, so null hypothesis is rejected at the level α if $\lambda > \chi^2_{1,\alpha}$.

4.2 Estimation of poverty indicators

In this section we present an application of *model* 1 to real data. We compare the obtained results with the corresponding ones under the same model restricted to $H_0 : \sigma^2_A = \sigma^2_B$. In what follows this last model will be denoted by *model* 0. The main goal is to estimate the poverty incidence (proportion of individuals under poverty) and the poverty gap in Spanish domains under both models.

We use the data from the Spanish Living Conditions Survey (SLCS) corresponding to years 2004-2006 and described in Tables 1.1. We consider $D = 104$ domains obtained by crossing 52 provinces with 2 sexes. The SLCS does not produce official estimates at the domain level (provinces $\times$ sex), but the analogous direct estimator of the total $Y_{dt} = \sum_{j=1}^{N_{dt}} y_{dtj}$ of a given variable y is

$$\hat Y^{dir}_{dt} = \sum_{j\in S_{dt}} w_{dtj}\, y_{dtj}.$$

where S_{dt} is the domain sample at time period t and the w_{dtj}'s are the official calibrated sampling weights which take into account for non response. In the particular case $y_{dtj} = 1$, for all $j \in P_{dt}$, we get the estimated domain size

$$\hat{N}_{dt}^{dir} = \sum_{j \in S_{dt}} w_{dtj}.$$

Using this quantity, a direct estimator of the domain mean $\bar{Y}_{dt}$ is $\bar{y}_{dt} = \hat{Y}_{dt}^{dir}/\hat{N}_{dt}^{dir}$. The direct estimates of the domain means are used as responses in the area-level time model. The design-based variances of these estimators are approximated by

$$\hat{V}_\pi(\hat{Y}_{dt}^{dir}) = \sum_{j \in S_{dt}} w_{dtj}(w_{dtj}-1)\left(y_{dtj}-\bar{y}_{dt}\right)^2 \text{ and } \hat{V}_\pi(\bar{y}_{dt}) = \hat{V}\left(\hat{Y}_{dt}^{dir}\right)/\hat{N}_{dt}^2.$$

As we are interested in the cases $y_{dtj} = y_{\alpha;dtj}$ (cf. (1.1)), $\alpha = 0, 1$, we select the direct estimates of the poverty incidence and poverty gap at domain d and time period t (i.e. $\bar{y}_{0;dt}$ and $\bar{y}_{1;dt}$ respectively) as target variables for the time dependent area-level models 0 and 1. The considered auxiliary variables are the domain means of the category indicators of the variables AGE, EDUCATION, CITIZENSHIP and LABOR described in Section 1.2.

We first consider the linear model

$$\bar{y}_{dt} = \bar{\mathbf{X}}_{dt}\beta + u_{dt} + e_{dt}, \quad d = 1, \ldots, D = D_A + D_B$$

where the population is partitioned in subsets of sizes D_A for sex=1 (men) and D_B for sex=2 (women) and $\bar{\mathbf{X}}_d$ is the $1 \times p$ vector containing the population (aggregated) mean values of the selected categories of the explanatory variables. Under the non-restricted *model* 1, random effects errors are assumed to follow the distributional assumptions of (4.1), otherwise the restriction $H_0 : \sigma_A^2 = \sigma_B^2$ requires to assume random effects u_{dt}'s i.i.d. $N(0, \sigma_u^2)$ for the *model* 0.

The selection of auxiliary variables is done under *model* 1. The first position of $\bar{\mathbf{X}}_d$ contains a "1", because of the presence of the intercept. The starting set of $1+4+3+1+3 = 12$ x-variables is reduced step by step by testing if the beta parameters are significatively different from zero. We take also into account the socio-economic sense of the selected x-variables. Finally, we have decided to use the same set of auxiliary x-variables for both poverty indicators $\alpha = 0$ and $\alpha = 1$. At the end of the procedure the set of $p = 1+2+1+1+1 = 6$ auxiliary variables appearing in Tables 4.2.1 and 4.2.2 is obtained. For the sake of comparability the same set of auxiliary variables are also used for *model* 1. Regression parameters and their corresponding p-values are presented in Tables 4.2.1 and 4.2.2.

model 1	*constant*	*age3*	*age4*	*edu2*	*cit1*	*lab2*
β	0.5327	-0.4993	-2.9722	-0.2162	0.4391	0.8940
$p-value$	0.0000	0.0024	0.0000	0.0030	0.0000	0.0001
model 0	*constant*	*age3*	*age4*	*edu2*	*cit1*	*lab2*
β	0.5298	-0.3909	-3.2626	-0.2192	0.4310	1.2334
$p-value$	0.0000	0.1251	0.0000	0.0253	0.0001	0.0001

Table 4.2.1. Regression parameters and p-values for $\alpha = 0$.

model 1	*constant*	*age3*	*age4*	*edu2*	*cit1*	*lab2*
β	0.1655	-0.1237	-1.0236	-0.0590	0.1343	0.3518
$p-value$	0.0008	0.0798	0.0000	0.0653	0.0004	0.0003
model 0	*constant*	*age3*	*age4*	*edu2*	*cit1*	*lab2*
β	0.1152	-0.0686	-0.8770	-0.0594	0.1311	0.5040
$p-value$	0.1000	0.5226	0.0000	0.1593	0.0038	0.0001

Table 4.2.2. Regression parameters and p-values for $\alpha = 1$.

By observing the signs of the regression parameters for $\alpha = 0$ (poverty proportions) in both models, we interpret that there is an inverse relation between poverty proportion and the categories *age3*, *age4* and *edu2* referred to the explanatory variables *age* and *education*. That is, poverty incidence tends to be smaller in those domains with larger proportion of population in the subset defined by age between 25 - 49 and 50 - 64, and in the subset defined by education in the category of secondary studied completed. On the other hand, poverty proportion tends to be larger in those domains with larger proportion of people in the subsets defined by category *cit1* of non Spanish citizenship (may be because immigrants tend to go to regions with greater richness where it is easier to find job) and with higher proportion of unemployed people by category *lab2*.

By doing the same exercise with the signs of the regression parameters for $\alpha = 1$ (poverty gaps) in both models, we can interpret them as in the previous case. Regarding the p-values, the performance of the beta parameters for *model* 0 is worse than for *model* 1. In particular, the beta parameters of *constant*, *age*3 and *edu*2 have a high p-value, so the corresponding estimates are unreliable.

The confidence intervals (CIs) for the β's at the level of confidence 90% are presented in Table 4.2.3 when $\alpha = 0$. The CIs for the β's at the level of confidence 90% are presented in Table 4.2.4 when $\alpha = 1$. The CI lower limit is labeled by INF and the upper limit by SUP. The CIs can be used to test the hypothesis $H_0 : \beta_j = 0$. Under the column labeled by $0 \in$CI appears FALSE if H_0 is rejected and TRUE if it is not rejected. In both Tables, H_0 is rejected for almost all the regression parameters.

	model 1			*model* 0		
ITEMS	INF	SUP	$0 \in$CI	INF	SUP	$0 \in$CI
constant	0.3424	0.7230	FALSE	0.2537	0.8060	FALSE
age3	-0.7695	-0.2292	FALSE	-0.8101	0.0284	TRUE
age4	-3.3487	-2.5957	FALSE	-3.8350	-2.6903	FALSE
edu2	-0.3359	-0.0964	FALSE	-0.3803	-0.0581	FALSE
cit1	0.2934	0.5848	FALSE	0.2547	0.6073	FALSE
lab2	0.5164	1.2716	FALSE	0.7211	1.7457	FALSE

Table 4.2.3. 90% confidence intervals for β's when $\alpha = 0$.

	model 1			*model* 0		
ITEMS	INF	SUP	$0 \in$CI	INF	SUP	$0 \in$CI
constant	0.0847	0.2463	FALSE	-0.0000	0.2305	TRUE
age3	-0.2398	-0.0075	FALSE	-0.2450	0.1078	TRUE
age4	-1.1837	-0.8635	FALSE	-1.1140	-0.6401	FALSE
edu2	-0.1117	-0.0064	FALSE	-0.1288	0.0100	TRUE
cit1	0.0724	0.1963	FALSE	0.0566	0.2056	FALSE
lab2	0.1909	0.5127	FALSE	0.2899	0.7181	FALSE

Table 4.2.4. 90% confidence intervals for β's when $\alpha = 1$.

Table 4.2.5 gives the CIs for the σ^2's at the level 90% for *model* 1 and *model* 0 and for $\alpha = 0$ and $\alpha = 1$. The CIs for σ_A^2, σ_B^2 and σ_u^2 do not contain the origin 0 in any case, so the variances are significatively positive. Table 4.2.6 presents the CIs for the difference of variances $\sigma_A^2 - \sigma_B^2$. Variances σ_A^2 and σ_B^2 can not be considered as different at the confidence level 90%. However, if we reduce the level to 78% for α_0 and to 70% for α_1, then we can do it.

	$\alpha = 0$			$\alpha = 1$		
model 1	INF	SUP	$0 \in$CI	INF	SUP	$0 \in$CI
σ_A^2	0.0018	0.0030	FALSE	0.0003	0.0006	FALSE
σ_B^2	0.0024	0.0039	FALSE	0.0004	0.0007	FALSE
model 0	INF	SUP	$0 \in$CI	INF	SUP	$0 \in$CI
σ_u^2	0.0019	0.0032	FALSE	0.0003	0.0006	FALSE

Table 4.2.5. 90% confidence intervals for variances.

$\alpha = 0$	INF	SUP	$0 \in$CI	$\alpha = 1$	INF	SUP	$0 \in$CI
90%	-0.0017	0.0002	TRUE	90%	-0.0003	0.0006	TRUE
78%	-0.0015	-0.0000	FALSE	70%	-0.0002	-0.0000	FALSE

Table 4.2.6. Confidence intervals for $\sigma_A^2 - \sigma_B^2$ under model 1.

In case of $\alpha = 0$ the REML likelihood ratio statistic (LRS) for testing $H_0 : \sigma_A^2 = \sigma_B^2$ takes the value 1.3631 and its corresponding p-value is 0.2430. In case of $\alpha = 1$ the value of the REML likelihood ratio statistic (LRS) for testing $H_0 : \sigma_A^2 = \sigma_B^2$ is 0.9502 and the corresponding p-value is 0.3297. In both cases we cannot reject the null hypothesis of equality of variances. Nevertheless, the p-values are rather small and this fact make us think that by slightly increasing the sample sizes (number of years in the study) the p-values will become lower than 0.05.

Residuals $\hat{e}_{dt} = \bar{y}_{dt} - \bar{\mathbf{X}}_{dt}\hat{\beta} - \hat{u}_{dt}$ of fitted *model* 1 and *model* 0 are plotted against the observed values $\bar{y}_{dt}$ in the Figure 4.1 for $\alpha = 0$ (on the left) and $\alpha = 1$ (right). Dispersion graphs show that there is a great difference, between *model* 1 and *model* 0, in the pattern of the plots. In particular, *model* 0 present heteroscedastic residuals and *model* 1 not. Further, *model* 1 presents a better fit to the direct estimates. Therefore we can recommend *model* 1.

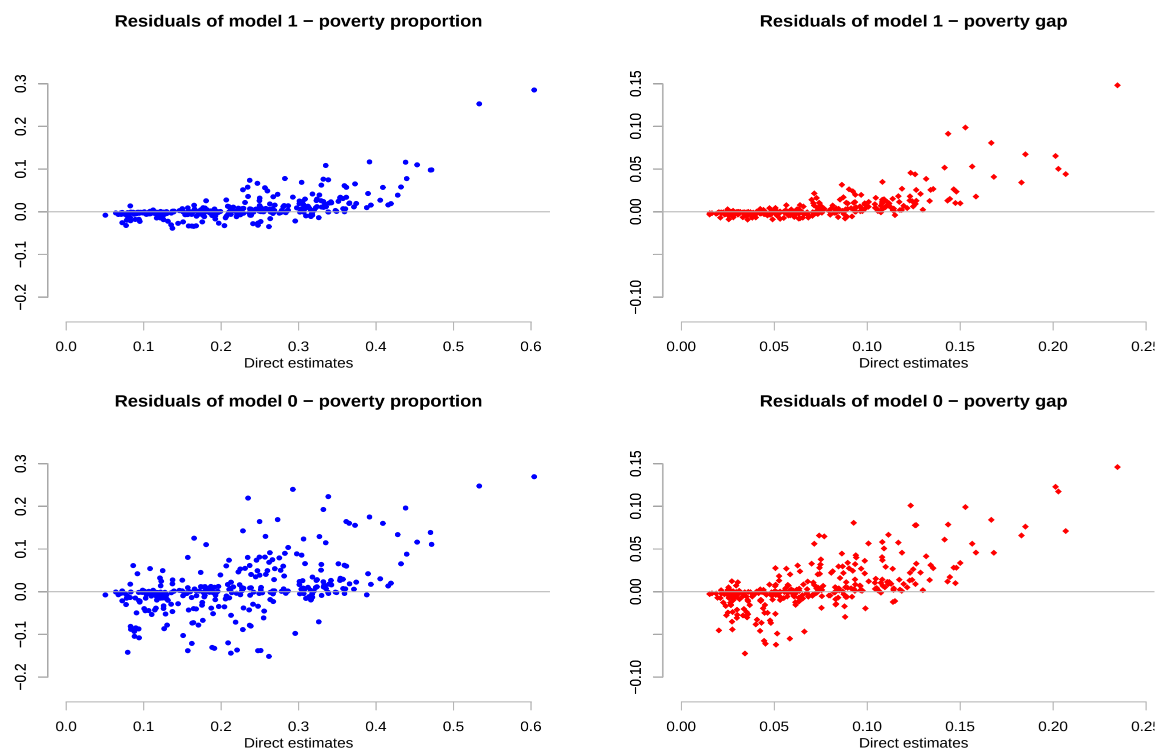


Figure 4.1: Residuals versus direct estimates.

Tables 4.2.7 and 4.2.8 present the full numerical information of the estimates of poverty proportions ($\alpha = 0$) and poverty gaps ($\alpha = 1$). The EBLUP estimates under models 1 and 0 are labeled by EB_1, EB_0 respectively. The direct estimates are labeled by DIR. The squared root of MSE's are labeled by $RMSE_\star$ for the direct estimator and by $RMSE_1$ and $RMSE_0$ for the EBLUPs under *model* 1 and *model* 0 respectively. Numerical results are sorted by sex. Regarding the reduction of the MSE when passing from direct to EBLUP estimates, we observe that *model* 0 performs better in a higher number of domains. On the other hand, the average of the MSEs across the domains in both models are very close. For $\alpha = 0$ the average of the MSEs for *model* 1 is 0.029011 while for *model* 0 is 0.026675. For $\alpha = 1$ the average of the MSEs for *model* 1 is 0.011936 while for *model* 0 is 0.010798.

In Tables 4.2.7 and 4.2.8 the Spanish provinces are listed as follows: 1 Álava, 2 Albacete, 3 Alicante, 4 Almería, 5 Ávila, 6 Badajoz, 7 Baleares, 8 Barcelona, 9 Burgos, 10 Cáceres, 11 Cádiz, 12 Castellón, 13 Ciudad Real, 14 Córdoba, 15 Coruña La, 16 Cuenca, 17 Gerona, 18 Granada, 19 Guadalajara, 20 Guipúzcoa, 21 Huelva, 22 Huesca, 23 Jaén, 24 León, 25 Lérida, 26 La Rioja, 27 Lugo, 28 Madrid, 29 Málaga, 30 Murcia, 31 Navarra, 32 Orense, 33 Asturias (Oviedo), 34 Palencia, 35 Palmas Las, 36 Pontevedra, 37 Salamanca, 38 Santa Cruz de Tenerife, 39 Cantabria (Santander), 40 Segovia, 41 Sevilla, 42 Soria, 43 Tarragona, 44 Teruel, 45 Toledo, 46 Valencia, 47 Valladolid, 48 Vizcaya, 49 Zamora, 50 Zaragoza, 51 Ceuta, 52 Melilla.

	Men						Women					
d	DIR	EB_1	EB_0	$RMSE_\star$	$RMSE_1$	$RMSE_0$	DIR	EB_1	EB_0	$RMSE_\star$	$RMSE_1$	$RMSE_0$
1	0.083	0.069	0.065	0.034	0.029	0.029	0.079	0.088	0.088	0.032	0.029	0.028
2	0.237	0.245	0.246	0.035	0.029	0.029	0.285	0.283	0.284	0.037	0.032	0.030
3	0.160	0.160	0.160	0.017	0.018	0.016	0.189	0.187	0.187	0.018	0.020	0.017
4	0.318	0.276	0.284	0.035	0.030	0.030	0.354	0.320	0.322	0.037	0.032	0.031
5	0.335	0.227	0.220	0.069	0.040	0.042	0.453	0.343	0.337	0.074	0.045	0.042
6	0.366	0.345	0.347	0.025	0.023	0.023	0.393	0.378	0.376	0.025	0.024	0.022
7	0.094	0.095	0.095	0.014	0.016	0.013	0.115	0.117	0.117	0.014	0.017	0.014
8	0.083	0.084	0.084	0.008	0.013	0.007	0.108	0.109	0.109	0.008	0.014	0.008
9	0.127	0.131	0.129	0.029	0.026	0.025	0.124	0.138	0.139	0.028	0.026	0.024
10	0.252	0.242	0.241	0.030	0.026	0.026	0.332	0.322	0.321	0.031	0.028	0.027
11	0.267	0.263	0.265	0.025	0.023	0.022	0.303	0.297	0.297	0.025	0.024	0.022
12	0.122	0.132	0.131	0.034	0.029	0.029	0.122	0.136	0.137	0.036	0.031	0.030
13	0.269	0.266	0.267	0.030	0.026	0.026	0.324	0.312	0.312	0.035	0.030	0.029
14	0.312	0.299	0.302	0.034	0.029	0.028	0.307	0.307	0.311	0.033	0.029	0.028
15	0.216	0.205	0.204	0.020	0.021	0.019	0.237	0.230	0.228	0.020	0.021	0.019
16	0.362	0.304	0.302	0.057	0.038	0.039	0.472	0.374	0.361	0.059	0.041	0.040
17	0.050	0.058	0.058	0.018	0.019	0.017	0.067	0.074	0.074	0.023	0.023	0.021
18	0.301	0.282	0.284	0.036	0.030	0.030	0.342	0.319	0.315	0.034	0.030	0.029
19	0.077	0.109	0.107	0.027	0.025	0.024	0.165	0.199	0.204	0.041	0.034	0.032
20	0.064	0.067	0.065	0.018	0.019	0.017	0.100	0.102	0.100	0.020	0.021	0.019
21	0.192	0.209	0.211	0.036	0.030	0.030	0.253	0.253	0.256	0.040	0.033	0.032
22	0.078	0.098	0.093	0.028	0.025	0.025	0.089	0.107	0.104	0.032	0.029	0.028
23	0.283	0.286	0.287	0.031	0.027	0.027	0.339	0.339	0.343	0.034	0.030	0.029
24	0.192	0.185	0.181	0.032	0.028	0.027	0.193	0.201	0.201	0.029	0.027	0.025
25	0.177	0.167	0.165	0.037	0.030	0.030	0.239	0.233	0.229	0.043	0.035	0.033
26	0.166	0.160	0.159	0.020	0.020	0.019	0.212	0.207	0.205	0.022	0.023	0.020
27	0.207	0.179	0.146	0.037	0.030	0.030	0.225	0.222	0.171	0.034	0.030	0.029
28	0.110	0.112	0.120	0.014	0.016	0.013	0.126	0.126	0.138	0.013	0.017	0.013
29	0.222	0.214	0.176	0.025	0.023	0.022	0.258	0.250	0.217	0.023	0.023	0.021
30	0.219	0.218	0.290	0.017	0.018	0.017	0.256	0.254	0.301	0.018	0.020	0.018
31	0.090	0.092	0.174	0.014	0.017	0.015	0.094	0.098	0.202	0.014	0.017	0.014
32	0.282	0.204	0.192	0.053	0.036	0.037	0.213	0.213	0.268	0.043	0.035	0.033
33	0.108	0.107	0.054	0.014	0.016	0.013	0.122	0.123	0.108	0.013	0.017	0.013
34	0.228	0.177	0.086	0.054	0.037	0.037	0.280	0.268	0.220	0.058	0.041	0.038
35	0.224	0.217	0.208	0.026	0.024	0.023	0.246	0.239	0.172	0.025	0.024	0.022
36	0.174	0.169	0.162	0.021	0.021	0.019	0.214	0.211	0.216	0.022	0.023	0.021
37	0.308	0.279	0.242	0.042	0.032	0.033	0.329	0.303	0.279	0.042	0.034	0.033
38	0.263	0.245	0.171	0.027	0.025	0.024	0.286	0.272	0.183	0.026	0.025	0.023
39	0.095	0.098	0.119	0.017	0.018	0.016	0.128	0.132	0.180	0.020	0.021	0.019
40	0.234	0.177	0.154	0.061	0.039	0.040	0.438	0.322	0.242	0.071	0.044	0.042
41	0.209	0.215	0.329	0.020	0.020	0.019	0.228	0.231	0.317	0.020	0.021	0.019
42	0.247	0.180	0.198	0.107	0.045	0.047	0.604	0.319	0.335	0.126	0.052	0.048
43	0.125	0.127	0.075	0.029	0.026	0.026	0.174	0.171	0.128	0.033	0.030	0.028
44	0.083	0.104	0.164	0.033	0.028	0.028	0.151	0.175	0.253	0.045	0.036	0.034
45	0.250	0.235	0.085	0.029	0.026	0.025	0.220	0.231	0.224	0.028	0.026	0.025
46	0.137	0.136	0.119	0.017	0.019	0.017	0.139	0.139	0.138	0.014	0.017	0.014
47	0.165	0.150	0.040	0.024	0.023	0.022	0.210	0.198	0.136	0.027	0.025	0.024
48	0.092	0.093	0.050	0.014	0.016	0.014	0.099	0.101	0.095	0.014	0.017	0.015
49	0.332	0.256	0.139	0.048	0.035	0.035	0.268	0.268	0.240	0.046	0.037	0.035
50	0.101	0.102	0.140	0.014	0.017	0.014	0.136	0.139	0.138	0.017	0.019	0.016
51	0.334	0.316	0.320	0.035	0.029	0.029	0.388	0.378	0.395	0.035	0.031	0.029
52	0.236	0.238	0.279	0.037	0.030	0.030	0.251	0.274	0.389	0.034	0.030	0.028

Table 4.2.7. Estimated domain poverty proportions ($\alpha = 0$) and their estimated RMSE's.

	Men						Women					
d	DIR	EB_1	EB_0	$RMSE_\star$	$RMSE_1$	$RMSE_0$	DIR	EB_1	EB_0	$RMSE_\star$	$RMSE_1$	$RMSE_0$
1	0.025	0.022	0.023	0.010	0.010	0.010	0.015	0.017	0.018	0.007	0.008	0.007
2	0.096	0.091	0.090	0.017	0.014	0.014	0.117	0.106	0.103	0.019	0.015	0.014
3	0.050	0.050	0.050	0.007	0.008	0.007	0.059	0.059	0.059	0.008	0.008	0.007
4	0.108	0.094	0.094	0.015	0.013	0.013	0.112	0.102	0.100	0.017	0.014	0.014
5	0.108	0.073	0.071	0.027	0.017	0.017	0.123	0.105	0.101	0.025	0.017	0.016
6	0.126	0.117	0.117	0.011	0.011	0.010	0.121	0.117	0.116	0.010	0.010	0.009
7	0.029	0.030	0.030	0.006	0.007	0.006	0.029	0.030	0.030	0.005	0.007	0.005
8	0.031	0.031	0.031	0.003	0.006	0.003	0.036	0.036	0.036	0.004	0.006	0.004
9	0.042	0.042	0.042	0.015	0.013	0.012	0.035	0.040	0.041	0.012	0.011	0.011
10	0.075	0.074	0.073	0.011	0.010	0.010	0.093	0.093	0.092	0.011	0.011	0.010
11	0.072	0.074	0.074	0.010	0.009	0.009	0.109	0.104	0.104	0.012	0.011	0.010
12	0.040	0.043	0.042	0.017	0.013	0.013	0.039	0.043	0.042	0.014	0.013	0.012
13	0.073	0.075	0.074	0.010	0.010	0.009	0.072	0.074	0.074	0.010	0.010	0.009
14	0.082	0.083	0.083	0.011	0.010	0.010	0.080	0.084	0.084	0.011	0.010	0.010
15	0.073	0.069	0.069	0.009	0.009	0.008	0.083	0.079	0.079	0.009	0.009	0.009
16	0.088	0.086	0.083	0.016	0.013	0.013	0.107	0.100	0.097	0.018	0.014	0.014
17	0.019	0.021	0.021	0.008	0.008	0.007	0.022	0.023	0.024	0.009	0.009	0.008
18	0.135	0.109	0.108	0.020	0.015	0.015	0.168	0.127	0.123	0.022	0.016	0.016
19	0.015	0.018	0.018	0.005	0.007	0.005	0.026	0.030	0.030	0.007	0.008	0.006
20	0.026	0.026	0.026	0.010	0.010	0.009	0.044	0.042	0.042	0.011	0.011	0.010
21	0.105	0.088	0.089	0.027	0.017	0.017	0.091	0.087	0.087	0.021	0.016	0.015
22	0.026	0.030	0.029	0.011	0.010	0.010	0.030	0.034	0.033	0.013	0.012	0.011
23	0.096	0.096	0.094	0.013	0.012	0.012	0.114	0.113	0.112	0.015	0.013	0.012
24	0.071	0.064	0.063	0.015	0.013	0.013	0.076	0.074	0.074	0.015	0.013	0.013
25	0.092	0.068	0.068	0.022	0.015	0.016	0.093	0.082	0.079	0.021	0.016	0.015
26	0.041	0.040	0.040	0.006	0.007	0.005	0.043	0.044	0.044	0.005	0.007	0.005
27	0.086	0.055	0.042	0.026	0.017	0.017	0.053	0.054	0.034	0.012	0.011	0.010
28	0.034	0.035	0.054	0.006	0.007	0.006	0.036	0.036	0.062	0.006	0.007	0.006
29	0.090	0.080	0.056	0.014	0.012	0.012	0.108	0.096	0.067	0.014	0.012	0.012
30	0.075	0.075	0.094	0.007	0.007	0.007	0.083	0.083	0.092	0.007	0.008	0.007
31	0.030	0.030	0.061	0.006	0.007	0.006	0.027	0.028	0.062	0.005	0.007	0.005
32	0.073	0.057	0.069	0.019	0.014	0.014	0.048	0.051	0.081	0.014	0.013	0.012
33	0.025	0.025	0.020	0.005	0.006	0.005	0.031	0.032	0.030	0.005	0.006	0.005
34	0.056	0.049	0.029	0.017	0.014	0.013	0.061	0.067	0.067	0.018	0.014	0.014
35	0.076	0.074	0.069	0.012	0.011	0.011	0.085	0.081	0.055	0.013	0.012	0.011
36	0.030	0.031	0.051	0.004	0.006	0.004	0.044	0.045	0.063	0.006	0.007	0.005
37	0.099	0.090	0.062	0.015	0.013	0.013	0.089	0.086	0.095	0.014	0.013	0.012
38	0.081	0.077	0.059	0.010	0.010	0.009	0.093	0.089	0.061	0.011	0.011	0.010
39	0.026	0.027	0.033	0.006	0.007	0.005	0.030	0.031	0.036	0.006	0.007	0.006
40	0.070	0.056	0.037	0.021	0.015	0.015	0.109	0.094	0.058	0.023	0.017	0.016
41	0.034	0.036	0.107	0.004	0.006	0.004	0.045	0.047	0.106	0.006	0.007	0.005
42	0.153	0.054	0.054	0.088	0.021	0.021	0.235	0.086	0.089	0.111	0.023	0.021
43	0.019	0.021	0.027	0.005	0.007	0.005	0.028	0.030	0.038	0.007	0.008	0.006
44	0.045	0.044	0.070	0.024	0.016	0.016	0.052	0.056	0.101	0.020	0.015	0.015
45	0.077	0.073	0.012	0.011	0.011	0.010	0.059	0.063	0.038	0.009	0.010	0.009
46	0.051	0.049	0.043	0.010	0.009	0.009	0.043	0.043	0.053	0.006	0.007	0.006
47	0.064	0.056	0.040	0.011	0.010	0.010	0.074	0.068	0.053	0.012	0.011	0.010
48	0.026	0.026	0.023	0.005	0.007	0.005	0.023	0.023	0.035	0.004	0.006	0.005
49	0.126	0.082	0.048	0.024	0.016	0.016	0.099	0.089	0.077	0.022	0.016	0.015
50	0.043	0.042	0.053	0.009	0.009	0.008	0.051	0.051	0.055	0.010	0.010	0.009
51	0.148	0.124	0.120	0.018	0.014	0.014	0.207	0.163	0.136	0.023	0.017	0.016
52	0.119	0.092	0.073	0.027	0.017	0.017	0.125	0.118	0.098	0.019	0.015	0.014

Table 4.2.8. Estimated domain poverty gaps ($\alpha = 1$) and their estimated RMSE's.

In Table 4.2.9 we present the normalized Euclidean distances between the direct and the EBLUPs estimates in both cases $\alpha = 0$ and $\alpha = 1$. We use the formula

$$D(\mathbf{x},\mathbf{y}) = D(x_1,\ldots,x_D;y_1,\ldots,y_D) = \left(\frac{1}{D}\sum_{d=1}^{D}(x_d - y_d)^2\right)^{1/2}.$$

We observe that EBLUP estimates under *model* 1 are closer to direct estimates than the corresponding estimates obtained under *model* 0.

$\alpha = 0$				$\alpha = 1$			
Model 1		Model 0		Model 1		Model 0	
Men	Women	Men	Women	Men	Women	Men	Women
0.0297	0.0488	0.0623	0.0671	0.0186	0.0228	0.0284	0.0305

Table 4.2.9: Normalized Euclidean distances for $\alpha = 0$ and $\alpha = 1$.

The three estimators of the poverty proportion, direct, EBLUP for *model* 1 (EBLUP1) and EBLUP for *model* 0 (EBLUP0) are plotted in the Figure 4.2. Results for men and women are plotted in the left and right hand side respectively. Estimated poverty proportion are plotted in the top of Figure 4.2 and their estimated root mean squared errors in the bottom. Figure 4.2 shows some differences in the behavior of EBLUP1 and EBLUP0 estimates. We observe that EBLUP1 estimates approximate direct estimates better than the EBLUP0 ones for both cases, $\alpha = 0$ and $\alpha = 1$. Nevertheless, EBLUP0 and EBLUP1 follow the basic pattern of direct estimates and they are are almost always close enough to them. So the models are producing EBLUPs that captures part of the good properties of the direct estimates, like being approximately unbiased with respect to the sampling design distribution. Figure 4.2 also shows that MSEs of EBLUPs are lower that the ones of direct estimates. So it is worthwhile to use the model based-estimators. Figure 4.3 plots the direct, EBLUP1 and EBLUP0 estimates of the poverty gaps. Similar comments and conclusions as for Figure 4.2 can be given by observing Figure 4.3.

Figure 4.4 plots the results obtained for the Spanish provinces in 4 colored categories depending on the values of the EBLUP1 estimates in % of the poverty proportions and the gaps, i.e. $p_d = 100 \cdot \hat{Y}^{EBLUP1}_{0;d,2006}$ and $g_d = 100 \cdot \hat{Y}^{EBLUP1}_{1;d,2006}$.

In the Table 4.2.9 the Spanish provinces are classified in 4 categories depending on the values of the EBLUP1 estimates in % of the poverty proportions and the gaps, i.e. $p_d = 100 \cdot \hat{Y}^{EBLUP1}_{0;d,2006}$ and $g_d = 100 \cdot \hat{Y}^{EBLUP1}_{1;d,2006}$. The poverty proportion categories are P_1 if $p_d < 10$, P_2 if $10 < p_d < 20$, P_3 if $20 < p_d < 30$ and P_4 if $p_d > 30$. The poverty gap categories are G_1 if $g_d < 3$, G_2 if $3 < G_d < 6$, G_3 if $6 < g_d < 10$ and G_4 if $g_d > 10$. The same results are plotted in Figure 4.4.

We observe that the Spanish regions where the proportion of the population under the poverty line is smallest are those situated in the north and east, like Cataluña, Aragón, Navarra, País Vasco, Cantabria and Baleares. On the other hand the Spanish regions with higher poverty proportion are those situated in the center-south, like Andalucía, Extremadura, Murcia, Castilla La Mancha, Canarias, Canarias, Ceuta and Melilla. In an intermediate position we can find regions that are in the center-north of Spain, like Galicia, La Rioja, Castilla León, Asturias, Comunidad Valenciana and Madrid. If we investigate how far the annual net incomes of population under the poverty line z_{2006} are from z_{2006}, we observe that in the

Figure 4.2: Estimates of poverty proportion (top) and squared root of their estimated MSEs (bottom) respectively for MEN (on the left) and WOMEN (on the right).

Spanish regions situated in the center-north there exist a distance that is generally lower than the 6% of z_{2006}. However, the cited distance is in general greater than 6% of z_{2006} in the center-south.

It is somehow surprising how large is the amount of Spanish provinces having a proportion greater than 30% of population with annual net incomes below z_{2006}. So it would be desirable that the Spanish Government implements policies to reduce poverty proportion in the center-south of Spain. A criticisms to the use of employed FGT poverty indicators is that they are defined by using only the income variable and do not consider the cost of living. In the case of Spain the south is poorer than the north and this is visualized in the obtained results. On the other hand to live in the south is in general cheaper than to live in the north. As the poverty line is officially calculated for the whole Spain, it is smaller than it should be in the north and greater than it should be in the south. Nevertheless, these comments may moderate but not annul the validity of the given conclusions. The obtained results are valuable tools for taking decisions as they show the basic poverty situation per provinces and sex in Spain.

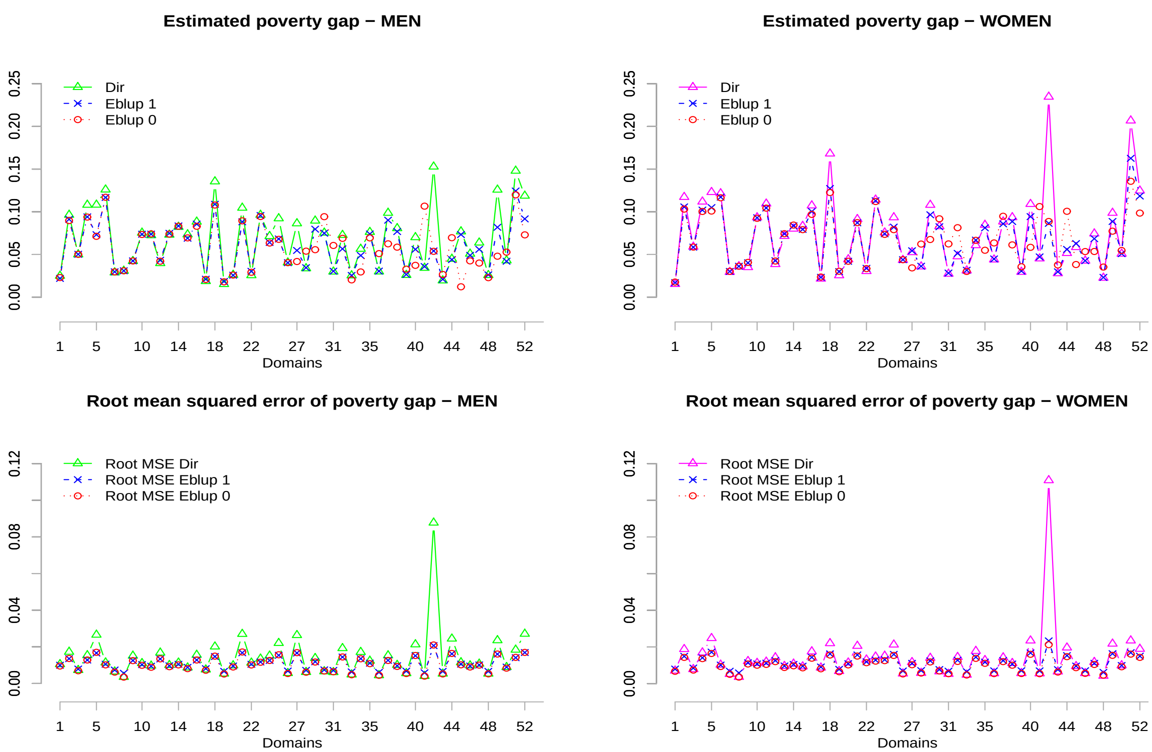


Figure 4.3: Estimates of poverty gap (top) and squared root of their estimated MSEs (bottom) respectively for MEN (on the left) and WOMEN (on the right).

	Range % of p_d	**Provinces included in the range for Poverty Proportion**
men	$p_d < 10$	1, 7, 8, 17, 20, 22, 31, 39, 48
	$10 < p_d < 20$	3, 9, 12, 19, 24, 25, 26, 27, 28, 33, 34, 36, 40, 42, 43, 44, 46, 47, 50
	$20 < p_d < 30$	2, 4, 5, 10, 11, 13, 14, 15, 18, 21, 23, 29, 30, 32, 35, 37, 38, 41, 45, 49
	$p_d > 30$	6, 16
women	$p_d < 10$	1, 17, 31
	$10 < p_d < 20$	3, 7, 8, 9, 12, 19, 20, 22, 28, 33, 39, 43, 44, 46, 47, 48, 50
	$20 < p_d < 30$	2, 11, 15, 21, 24, 25, 26, 27, 29, 30, 32, 34, 35, 36, 38, 41, 45, 49
	$p_d > 30$	4, 5, 6, 10, 13, 14, 16, 18, 23, 37, 40, 42
	Range % of g_d	**Provinces included in the range for Poverty Gap**
men	$g_d < 3$	1, 7, 17, 19, 20, 33, 39, 43, 48
	$3 < g_d < 6$	3, 8, 9, 12, 22, 26, 27, 28, 31, 32, 34, 36, 40, 41, 42, 44, 46, 47, 50
	$6 < g_d < 10$	2, 4, 5, 10, 11, 13, 14, 15, 16, 21, 23, 24, 25, 29, 30, 35, 37, 38, 45, 49
	$g_d > 10$	6, 18
women	$g_d < 3$	1, 17, 19, 31, 43, 48
	$3 < g_d < 6$	3, 7, 8, 9, 12, 20, 22, 26, 27, 28, 32, 33, 36, 39, 41, 44, 46, 50
	$6 < g_d < 10$	10, 13, 14, 15, 21, 24, 25, 29, 30, 34, 35, 37, 38, 40, 42, 45, 47, 49
	$g_d > 10$	2, 4, 5, 6, 11, 16, 18, 23

Table 4.2.10. Spanish provinces classified by poverty proportion (up) and gap (bottom) in %.

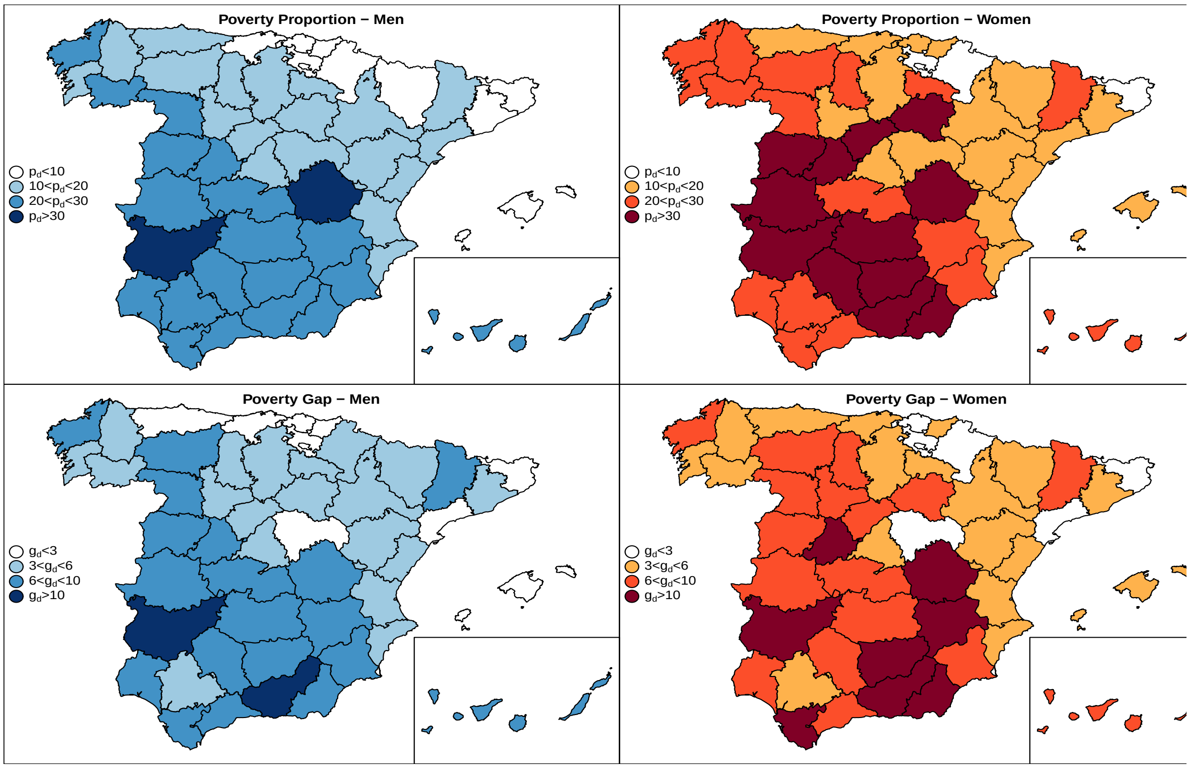


Figure 4.4: Estimates of poverty proportions (top) and gaps (bottom) for men (left) and women (right).

Chapter 5

EB prediction of poverty measures with unit level models

5.1 EB method for poverty estimation

Consider a population U partitioned in D domains or areas $U_1,\dots,U_D$ of sizes $N_1,\dots,N_D$. A sample $s_d \subset U_d$ of size n_d has been drawn from each domain d, $d=1,\dots,D$. Let E_{dj} be the value of a quantitative welfare measure for j-th individual within d-th domain and z a poverty line defined for the population. Our target parameters are the FGT poverty measures for domain d, given by

$$F_{\alpha d} = \frac{1}{N_d}\sum_{j=1}^{N_d}\left(\frac{z-E_{dj}}{z}\right)^{\alpha} I(E_{dj}<z), \quad \alpha=0,1,2. \tag{5.1}$$

Calculation of the BP of $F_{\alpha d}$ requires to express $F_{\alpha d}$ in terms of a domain vector $\mathbf{y}_d$, for which the conditional distribution of the out-of-sample vector $\mathbf{y}_{dr}$ given sample data $\mathbf{y}_{ds}$ is known. The distribution of the welfare variables E_{dj} is seldom Normal due to the typical right-skewness of these kind of economical variables. However, here we suppose that there is a one-to-one transformation $Y_{dj}=T(E_{dj})$ following a Normal distribution. In particular, we will assume that Y_{dj} follows the nested error model of Battese, Harter and Fuller (1999),

$$\begin{aligned} &Y_{dj} = \mathbf{x}_{dj}\beta + u_d + e_{dj}, \qquad j=1,\dots,N_d, \quad d=1,\dots D,\\ &u_d \sim \text{ iid } N(0,\sigma_u^2), \quad e_{dj} \sim \text{ iid } N(0,\sigma_e^2) \end{aligned} \tag{5.2}$$

where $\mathbf{x}_{dj}$ is a row vector with the values of p explanatory variables, u_d is a random area-specific effect and e_{dj} are residual errors. Let $\mathbf{y}_d=(\mathbf{y}_{ds}',\mathbf{y}_{dr}')'$ be the vector containing the values of the transformed variables Y_{dj} for the sample and out-of-sample units within domain d. Then $F_{\alpha d}$ is function of $\mathbf{y}_d$, that is

$$F_{\alpha d} = \frac{1}{N_d}\sum_{j=1}^{N_d}\left(\frac{z-T^{-1}(Y_{dj})}{z}\right)^{\alpha} I(T^{-1}(Y_{dj})<z) =: h_\alpha(\mathbf{y}_d), \quad \alpha=0,1,2.$$

Thus, the FGT poverty measure of order α is a non-linear function $h_\alpha(\mathbf{y}_d)$ of $\mathbf{y}_d$. Then the BP of $F_{\alpha d}$ is given by

$$\hat{F}_{dj}^{B} = E_{\mathbf{y}_{dr}}[h_\alpha(\mathbf{y}_d)|\mathbf{y}_{ds}] = \int_{I\!R} h_\alpha(\mathbf{y}_d) f(\mathbf{y}_{dr}|\mathbf{y}_{ds})\, d\mathbf{y}_{dr}, \tag{5.3}$$

where $f(\mathbf{y}_{dr}|\mathbf{y}_{ds})$ is the joint density of $\mathbf{y}_{dr}$ given the observed data vector $\mathbf{y}_{ds}$. Due to the complexity of the function $h_\alpha(\cdot)$, there is not explicit expression for the expectation in (5.3), but this expectation can be approximated by Monte Carlo. For this, generate L replicates $\{\mathbf{y}_d^{(\ell)}; \ell = 1,\ldots,L\}$ of $\mathbf{y}_d$ from the distribution of $\mathbf{y}_{dr}|\mathbf{y}_{ds}$. Then, an approximation to the best predictor of $F_{\alpha d}$ is

$$\hat{F}_{\alpha d}^{B} \approx \frac{1}{L}\sum_{\ell=1}^{L} h_\alpha(\mathbf{y}_d^{(\ell)}).$$

Typically, the mean vector μ_d and the covariance matrix $\mathbf{V}_d$ depend on an unknown vector of parameters θ. Then the conditional density $f(\mathbf{y}_{dr}|\mathbf{y}_{ds})$ depends on θ, and we make this explicit by writing $f(\mathbf{y}_{dr}|\mathbf{y}_{ds},\theta)$. We take an estimator $\hat{\theta}$ of θ such as the maximum likelihood (ML) or restricted ML (REML)estimator. Then the expectation can be approximated by generating values from the estimated density $f(\mathbf{y}_{dr}|\mathbf{y}_{ds},\hat{\theta})$. The result is the EBP, denoted $\hat{F}_{\alpha d}^{EB}$.

5.2 Parametric bootstrap for MSE estimation

The parametric bootstrap of González-Manteiga et al. (2008) has been used to derive an estimator of the MSE of the EB estimator of the poverty measures. This method works as follows.

(1) Fit the nested-error model (6.9) by ML, REML or Henderson method III, deriving model parameter estimates $\hat{\beta}$, $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$.

(2) Generate bootstrap domain effects from:

$$u_d^* \stackrel{iid}{\sim} N(0,\hat{\sigma}_u^2), \quad d = 1,\ldots,D.$$

(3) Generate, independently of $u_1^*,\ldots,u_D^*$, disturbances:

$$e_{dj}^* \stackrel{iid}{\sim} N(0,\hat{\sigma}_e^2), \quad j = 1,\ldots,N_d,\ d = 1,\ldots,D$$

(4) Generate a bootstrap population from the model

$$y_{dj}^* = \mathbf{x}_{dj}^T\hat{\beta} + u_d^* + e_{dj}^*, \quad j = 1,\ldots,N_d,\ d = 1,\ldots,D.$$

(5) Calculate target quantities for the bootstrap population

$$F_{\alpha d}^* = \frac{1}{N_d}\sum_{j=1}^{N_d} F_{\alpha dj}^*, \quad F_{\alpha dj}^* = h_\alpha(y_{dj}^*), \quad d = 1,\ldots,D.$$

(6) Take the elements y_{dj}^* with indices contained in the sample s: $\mathbf{y}_s^*$. Fit the model to bootstrap sample $\mathbf{y}_s^*$: $\hat{\sigma}_\upsilon^{2*}$, $\hat{\sigma}_e^{2*}$, $\hat{\beta}^*$

(7) Obtain the bootstrap EBP (through the Monte Carlo approximation), $\hat{F}_{\alpha d}^{EB*}$.

(8) Repeat (2)–(7) B times. Let $F^*_{\alpha d}(b)$ be true value and $\hat{F}^{EB*}_{\alpha b}(b)$ the EBP for bootstrap sample b, $b = 1,\ldots,B$.

(9) The bootstrap estimator of the MSE is given by

$$\mathrm{mse}_B(\hat{F}^{EB}_{\alpha d}) = B^{-1}\sum_{b=1}^{B}\left\{\hat{F}^{EB*}_{\alpha d}(b) - F^*_{\alpha d}(b)\right\}^2.$$

5.3 Application with the Spanish EU-SILC data

The EB method has been used to obtain poverty indicators for Spanish provinces by gender, using data from the 2006 *Spanish Survey on Income and Living Conditions* (SILC). In this case, the welfare variable is the equivalized income, computed on the basis of the annual net income with standard protocols of the Spanish Statistical Institute (INE). A positive quantity was added to make it always positive and then a logarithm transformation has been taken. This is the response variable considered in the nested-error regression model. The auxiliary variables are:

- Five categories of variable age,
- Spanish nationality (yes/no),
- Education level with four categories,
- Employment with three categories: *unemployed*, *employed* and *inactive*.

In each auxiliary variable, one of the categories was considered as the reference one and an intercept was included. As in the EB method, the total number of people with the same values in abscises must be known, these were estimated by means of the sampling weights corresponding to the sample units in the SILC.

We used a parametric bootstrap estimator $mse(\hat{F}^{EB}_{\alpha d})$ with $B = 500$ replicates to estimate the MSEs of the poverty measures. Tables 5.3 and 5.4 report the EB estimators, $\hat{F}^{EB}_{\alpha d}$, and their estimated coefficients of variation (CVs) for the poverty incidence ($\alpha = 0$) and the poverty gap ($\alpha = 1$) for few representative domains (provinces $\times$ gender), where $cv(\hat{F}^{EB}_{\alpha d}) = \{mse(\hat{F}^{EB}_{\alpha d})\}^{1/2}/\hat{F}^{EB}_{\alpha d}$. Direct estimators and their estimated variances were also calculated following standard formulas in sampling theory, but taking as observations the quantities

$$F_{\alpha dj} = \left(\frac{z - E_{dj}}{z}\right)^{\alpha} I(E_{dj} < z), \quad j \in s_d$$

and using the SILC sampling weights attached to the units w_{dj}, namely

$$\hat{F}^{w}_{\alpha d} = \frac{1}{N_d}\sum_{j\in s_d} w_{dj}F_{\alpha dj} \quad \text{and} \quad var(\hat{F}^{w}_{\alpha d}) = \frac{1}{N_d^2}w_{dj}(w_{dj}-1)F^2_{\alpha dj},$$

Tables 5.3 and 5.4 report the direct and EB estimators and their corresponding estimated coefficients of variation. Observe that the CVs of the EB estimators are much smaller than those of direct estimators

Province	Gender	N_d	n_d	$\hat{F}_{0d}^{w}$	$\hat{F}_{0d}^{EB}$	var$(\hat{F}_{0d}^{w})$	mse$(\hat{F}_{0d}^{EB})$	cv$(\hat{F}_{0d}^{w})$	cv$(\hat{F}_{0d}^{EB})$
Soria	F	17211	17	60,41	31,48	158,6708	27,0518	20,85	16,52
Tarragona	M	264627	129	12,46	14,86	8,5695	5,7605	23,50	16,15
Córdoba	F	364583	230	30,66	33,32	10,7598	5,0252	10,70	6,73
Badajoz	M	351985	472	36,58	36,56	6,1853	1,7031	6,80	3,57
Barcelona	F	2752431	1483	10,82	13,10	0,6605	0,4944	7,51	5,37

Table 5.1: Population size, sample size, direct and EB estimates of poverty incidences ($\times 100$), estimated MSEs of direct and EB estimators ($\times 10^4$) and CVs of direct and EB estimators ($\times 100$) for the Spanish domains with sample size closest to minimum, first quartile, median, third quartile and maximum.

Province	Gender	N_d	n_d	$\hat{F}_{1d}^{w}$	$\hat{F}_{1d}^{EB}$	var$(\hat{F}_{1d}^{w})$	mse$(\hat{F}_{1d}^{EB})$	cv$(\hat{F}_{1d}^{w})$	cv$(\hat{F}_{1d}^{EB})$
Soria	F	17211	17	23,46	11,84	122,9756	5,5980	47,27	19,99
Tarragona	M	264627	129	1,95	4,53	0,2800	1,0997	27,15	23,14
Córdoba	F	364583	230	8,01	12,26	1,1694	1,1819	13,50	8,87
Badajoz	M	351985	472	12,59	14,11	1,2979	0,3086	9,05	3,94
Barcelona	F	2752431	1483	3,60	3,92	0,1297	0,1027	10,00	8,17

Table 5.2: Population size, sample size, direct and EB estimates of poverty gaps ($\times 100$), estimated MSEs of direct and EB estimators ($\times 10^4$) and CVs of direct and EB estimators ($\times 100$) for the Spanish domains with sample size closest to minimum, first quartile, median, third quartile and maximum.

for almost all domains and the reduction in CV tends to be greater for domains with smaller sample sizes. As national statistical offices establish a maximum publishable CV, in these data, the estimated CVs of direct estimators of poverty incidences exceeded the level of 10% for 87 (out of the 104) domains while those of the EB estimators exceeded this level for only 28 domains. An increase of the level to 20%, make the direct estimators to have greater CV for 25 domains, although the CV of EB estimators exceeded 20% only for one domain.

We show here cartograms (see Figures 5.1 and 5.2) of the estimated poverty incidences and the poverty gaps in the Spanish provinces for males and females obtained using the EB method. Here, the highest level in poverty is located in general in the south and west parts of Spain:

(i) For men: critical poverty incidences (over 30%) Almería and Córdoba (south); Badajoz, Ávila, Salamanca and Zamora (west) and Cuenca (east).

(ii) For women: poverty incidence increases in most provinces, becoming critical also, in the south (Granada, Jaén, Albacete and Ciudad Real) and in the north (Palencia and Soria).

In Lleida (north-east) the poverty level seems to be an outlier because this province is included in Catalonia, which is commonly considered as a rich community. The poverty gap measures the degree of poverty instead of the quantity of people under poverty. Therefore in a region with a great amount of people income under the poverty line but close to it, the poverty gap will be near zero. In this way, provinces with income over 12.5% under the poverty line are also included in the group with critical values of poverty incidence. In this case, Lleida does not show a significant gap in comparison with the rest of the provinces. This province contains many people whose equivalized income is under the poverty line, but close to it.

Table 4.3: Results on poverty incidence: Spanish SILC data. Columns respectively denote province, gender, population size, sample size, direct estimate of poverty incidence, EB estimate, estimated variance of direct estimator, estimated MSE of EB estimator, CV of direct estimator, CV of EB estimator and ratio of CVs of direct estimators over EB estimators. Estimated poverty incidences and CVs in percentage.

Province	Gender	N_d	n_d	$\hat{F}_{0d}^{w}$	$\hat{F}_{0d}^{EB}$	$var(\hat{F}_{0d}^{w})$	$mse(\hat{F}_{0d}^{EB})$	cv$(\hat{F}_{0d}^{w})$	cv$(\hat{F}_{0d}^{EB})$	Ratio
Álava	M	99354	95	8,27	12,84	11,5481	6,8045	41,11	20,32	2,02
Álava	F	108422	96	7,87	12,50	10,4377	6,0004	41,08	19,60	2,10
Albacete	M	184058	163	23,74	29,22	12,5291	4,6175	14,91	7,35	2,03
Albacete	F	186503	183	28,52	33,74	13,8958	4,6180	13,07	6,37	2,05
Alicante	M	929288	526	16,00	19,45	2,8864	1,4660	10,62	6,23	1,71
Alicante	F	931405	552	18,85	22,59	3,2665	1,6009	9,59	5,60	1,71
Almería	M	341228	204	31,78	32,88	12,4891	3,6423	11,12	5,80	1,92
Almería	F	318857	193	35,39	35,72	13,9795	5,0203	10,57	6,27	1,68
Ávila	M	56601	56	33,50	31,48	48,2337	12,0614	20,73	11,03	1,88
Ávila	F	61708	60	45,29	38,51	54,7503	13,2846	16,34	9,46	1,73
Badajoz	M	351985	472	36,58	36,56	6,1853	1,7031	6,80	3,57	1,90
Badajoz	F	346810	515	39,33	39,13	6,0257	1,9473	6,24	3,57	1,75
Baleares	M	477561	609	9,36	11,55	1,8329	1,0415	14,47	8,84	1,64
Baleares	F	472843	660	11,52	14,05	2,0379	1,1302	12,40	7,57	1,64
Barcelona	M	2617681	1358	8,35	10,49	0,5676	0,5241	9,02	6,90	1,31
Barcelona	F	2752431	1483	10,82	13,10	0,6605	0,4944	7,51	5,37	1,40
Burgos	M	215155	168	12,73	16,72	8,4949	3,7359	22,89	11,56	1,98
Burgos	F	211240	167	12,43	18,33	7,6372	4,0972	22,23	11,04	2,01
Cáceres	M	169833	261	25,15	24,69	8,8116	2,0992	11,80	5,87	2,01
Cáceres	F	184785	302	33,23	28,24	9,4933	2,6895	9,27	5,81	1,60
Cádiz	M	642053	373	26,69	26,88	6,1005	2,0133	9,25	5,28	1,75
Cádiz	F	681522	422	30,34	31,63	6,0606	2,3161	8,11	4,81	1,69
Castellón	M	201428	113	12,19	14,79	11,8952	6,4886	28,29	17,22	1,64
Castellón	F	197726	123	12,18	17,35	12,8760	6,0077	29,46	14,13	2,09
Ciudad Real	M	265393	260	26,88	28,39	8,7218	3,0603	10,99	6,16	1,78
Ciudad Real	F	254508	239	32,37	30,18	11,9596	3,5984	10,68	6,29	1,70
Córdoba	M	356218	217	31,21	30,16	11,3893	3,6360	10,81	6,32	1,71
Córdoba	F	364583	230	30,66	33,32	10,7598	5,0252	10,70	6,73	1,59
La Coruña	M	509141	457	21,57	24,66	4,0965	1,5492	9,38	5,05	1,86
La Coruña	F	563190	533	23,70	25,36	4,0267	1,7892	8,47	5,27	1,61
Cuenca	M	92275	96	36,16	35,26	32,4762	6,6763	15,76	7,33	2,15
Cuenca	F	86760	87	47,17	35,35	35,3979	9,1829	12,61	8,57	1,47
Gerona	M	307975	145	5,05	13,29	3,2613	4,5124	35,77	15,98	2,24
Gerona	F	245519	138	6,72	15,38	5,1708	5,6725	33,82	15,48	2,18
Granada	M	371735	188	30,10	29,16	13,0768	3,4234	12,01	6,35	1,89
Granada	F	424598	229	34,17	36,34	11,8751	3,3397	10,08	5,03	2,01
Guadalajara	M	87591	92	7,73	12,74	7,2795	6,3392	34,89	19,76	1,77
Guadalajara	F	79560	86	16,46	15,83	17,1614	9,0549	25,17	19,01	1,32
Guipúzcoa	M	323719	279	6,39	11,30	3,1379	2,4878	27,71	13,96	1,98
Guipúzcoa	F	348524	291	9,95	14,56	4,1817	2,2501	20,54	10,30	1,99
Huelva	M	223158	121	19,24	29,06	13,1036	4,9763	18,82	7,68	2,45
Huelva	F	214587	123	25,31	29,13	15,9706	5,6059	15,79	8,13	1,94
Huesca	M	96617	125	7,79	17,11	7,7903	6,1259	35,81	14,47	2,48
Huesca	F	91147	105	8,92	18,99	10,4123	8,0476	36,17	14,94	2,42

Jaén	M	380752	233	28,34	28,60	9,7503	2,8775	11,02	5,93	1,86
Jaén	F	356344	230	33,86	32,31	11,4904	4,0100	10,01	6,20	1,62
León	M	204462	209	19,15	22,60	10,1969	3,7717	16,67	8,59	1,94
León	F	225753	228	19,28	24,17	8,3597	3,6393	15,00	7,89	1,90
Lérida	M	214123	127	17,67	25,74	13,9178	6,1157	21,11	9,61	2,20
Lérida	F	218051	133	23,86	27,36	18,5469	5,7767	18,05	8,79	2,05
La Rioja	M	149238	519	16,57	18,57	3,9800	1,3831	12,04	6,33	1,90
La Rioja	F	147554	500	21,25	21,45	4,8900	1,5570	10,41	5,82	1,79
Lugo	M	175462	169	20,66	24,51	13,3649	3,9145	17,69	8,07	2,19
Lugo	F	167892	177	22,47	26,87	11,5973	4,0344	15,16	7,47	2,03
Madrid	M	2816184	893	10,98	12,06	1,8430	0,6187	12,36	6,52	1,90
Madrid	F	3011923	996	12,57	13,91	1,7183	0,7045	10,43	6,04	1,73
Málaga	M	693871	361	22,22	27,95	6,0082	2,0308	11,03	5,10	2,16
Málaga	F	702667	397	25,76	32,45	5,3280	2,1898	8,96	4,56	1,97
Murcia	M	668714	868	21,87	25,35	2,7740	1,0275	7,61	4,00	1,90
Murcia	F	660107	902	25,55	28,70	3,1275	1,0872	6,92	3,63	1,91
Navarra	M	286947	525	8,98	9,13	2,0382	1,4047	15,91	12,98	1,23
Navarra	F	289947	603	9,40	11,40	1,9199	1,2109	14,75	9,66	1,53
Orense	M	120257	118	28,23	25,07	27,7280	5,9930	18,66	9,77	1,91
Orense	F	137587	140	21,27	22,12	18,2438	4,8094	20,08	9,92	2,03
Oviedo	M	511169	742	10,82	16,01	1,8852	0,8239	12,69	5,67	2,24
Oviedo	F	546817	864	12,20	16,59	1,7551	0,8935	10,86	5,70	1,91
Palencia	M	75638	71	22,82	26,16	29,0001	10,4552	23,60	12,36	1,91
Palencia	F	72558	72	28,00	30,13	33,8491	10,9066	20,78	10,96	1,90
Las Palmas	M	592262	458	22,39	24,65	6,8535	1,6151	11,69	5,16	2,27
Las Palmas	F	580265	485	24,57	25,40	6,0901	1,5201	10,05	4,85	2,07
Pontevedra	M	494161	434	17,36	19,15	4,3008	1,6196	11,94	6,64	1,80
Pontevedra	F	525627	462	21,37	22,66	5,0313	1,8652	10,50	6,03	1,74
Salamanca	M	151335	166	30,83	31,46	17,5220	3,8624	13,58	6,25	2,17
Salamanca	F	152234	162	32,90	33,56	17,8204	4,0296	12,83	5,98	2,15
Tenerife	M	366253	370	26,29	24,14	7,4825	1,5902	10,41	5,22	1,99
Tenerife	F	376690	392	28,64	26,36	6,6434	2,0065	9,00	5,37	1,67
Santander	M	267290	424	9,49	16,00	2,7620	1,3983	17,52	7,39	2,37
Santander	F	279191	443	12,82	16,93	3,8771	1,6784	15,36	7,65	2,01
Segovia	M	62518	57	23,43	19,24	37,7420	10,9102	26,22	17,17	1,53
Segovia	F	63217	58	43,80	26,74	51,0263	12,0322	16,31	12,97	1,26
Sevilla	M	816795	472	20,90	19,61	3,9714	1,5750	9,54	6,40	1,49
Sevilla	F	853057	491	22,80	24,04	4,1109	1,4935	8,89	5,08	1,75
Soria	M	26431	24	24,67	26,33	114,1075	20,6661	43,30	17,26	2,51
Soria	F	17211	17	60,41	31,48	158,6708	27,0518	20,85	16,52	1,26
Tarragona	M	264627	129	12,46	14,86	8,5695	5,7605	23,50	16,15	1,45
Tarragona	F	255490	139	17,36	19,28	10,9952	5,1967	19,10	11,82	1,62
Teruel	M	53380	66	8,30	17,13	11,1166	8,4195	40,16	16,94	2,37
Teruel	F	65002	78	15,09	22,26	20,3901	10,1119	29,92	14,29	2,09
Toledo	M	288335	278	24,96	26,22	8,3920	1,8713	11,60	5,22	2,22
Toledo	F	305241	272	21,99	22,50	7,8891	2,7841	12,77	7,42	1,72
Valencia	M	1169258	686	13,70	17,89	3,0194	0,9395	12,68	5,42	2,34
Valencia	F	1197478	742	13,88	20,78	1,9784	1,1622	10,14	5,19	1,95
Valladolid	M	305496	292	16,51	15,34	5,6491	2,2156	14,40	9,70	1,48
Valladolid	F	322530	306	21,02	18,29	7,0314	2,3520	12,62	8,38	1,50
Vizcaya	M	576042	515	9,18	10,01	1,9012	1,2668	15,02	11,24	1,34
Vizcaya	F	590094	532	9,86	11,57	2,0773	1,1750	14,61	9,37	1,56
Zamora	M	101433	109	33,19	34,67	23,2980	7,3875	14,54	7,84	1,86
Zamora	F	98337	100	26,82	32,84	21,5443	7,9639	17,30	8,59	2,01

Zaragoza	M	466651	555	10,07	15,42	2,0890	1,2316	14,35	7,20	1,99
Zaragoza	F	462937	574	13,57	15,34	2,9199	0,9892	12,59	6,48	1,94
Ceuta	M	35705	223	33,41	30,26	12,2208	3,4820	10,46	6,17	1,70
Ceuta	F	40426	247	38,79	33,15	12,3241	3,8045	9,05	5,88	1,54
Melilla	M	30595	179	23,61	19,27	13,7321	3,7828	15,70	10,09	1,56
Melilla	F	27498	180	25,10	25,45	11,5507	5,5790	13,54	9,28	1,46

Table 4.4: Results on poverty gap: Spanish SILC data. Columns respectively denote province, gender, population size, sample size, direct estimate of poverty gap, EB estimate, estimated variance of direct estimator, estimated MSE of EB estimator, CV of direct estimator, CV of EB estimator and ratio of CVs of direct estimators over EB estimators. Estimated poverty gaps and CVs in percentage.

Province	Gender	N_d	n_d	$\hat{F}_{1d}^{w}$	$\hat{F}_{1d}^{EB}$	$var(\hat{F}_{1d}^{w})$	$mse(\hat{F}_{1d}^{EB})$	cv$(\hat{F}_{1d}^{w})$	cv$(\hat{F}_{1d}^{EB})$	Ratio
Álava	M	99354	95	2,49	3,75	1,0904	1,1907	41,94	29,09	1,44
Álava	F	108422	96	1,53	3,65	0,4942	1,1315	45,97	29,12	1,58
Albacete	M	184058	163	9,63	10,53	2,9626	0,9286	17,87	9,15	1,95
Albacete	F	186503	183	11,72	12,68	3,5333	0,9250	16,03	7,59	2,11
Alicante	M	929288	526	5,00	6,29	0,5269	0,2937	14,53	8,62	1,68
Alicante	F	931405	552	5,89	7,55	0,6127	0,3407	13,28	7,73	1,72
Almería	M	341228	204	10,81	12,30	2,3507	0,6839	14,19	6,73	2,11
Almería	F	318857	193	11,18	13,64	2,8714	1,0880	15,16	7,65	1,98
Ávila	M	56601	56	10,82	11,64	7,0443	2,3237	24,54	13,09	1,87
Ávila	F	61708	60	12,30	15,40	6,1424	2,8775	20,15	11,01	1,83
Badajoz	M	351985	472	12,59	14,11	1,2979	0,3086	9,05	3,94	2,30
Badajoz	F	346810	515	12,15	15,46	1,0543	0,4007	8,45	4,09	2,06
Baleares	M	477561	609	2,88	3,34	0,4130	0,1955	22,28	13,24	1,68
Baleares	F	472843	660	2,94	4,23	0,2716	0,2227	17,72	11,17	1,59
Barcelona	M	2617681	1358	3,07	3,00	0,1224	0,0997	11,38	10,53	1,08
Barcelona	F	2752431	1483	3,60	3,92	0,1297	0,1027	10,00	8,17	1,22
Burgos	M	215155	168	4,22	5,21	2,2735	0,6704	35,72	15,70	2,27
Burgos	F	211240	167	3,50	5,81	1,4983	0,8377	34,93	15,75	2,22
Cáceres	M	169833	261	7,54	8,52	1,2188	0,3704	14,65	7,14	2,05
Cáceres	F	184785	302	9,33	10,13	1,2620	0,5620	12,03	7,40	1,63
Cádiz	M	642053	373	7,24	9,38	0,9284	0,4024	13,31	6,76	1,97
Cádiz	F	681522	422	10,95	11,65	1,4154	0,5101	10,87	6,13	1,77
Castellón	M	201428	113	3,97	4,48	2,8120	1,2242	42,19	24,68	1,71
Castellón	F	197726	123	3,86	5,51	2,0386	1,3159	36,97	20,83	1,77
Ciudad Real	M	265393	260	7,30	10,07	0,9995	0,6338	13,70	7,91	1,73
Ciudad Real	F	254508	239	7,15	10,86	0,9134	0,7758	13,36	8,11	1,65
Córdoba	M	356218	217	8,22	10,82	1,2822	0,6983	13,77	7,72	1,78
Córdoba	F	364583	230	8,01	12,26	1,1694	1,1819	13,50	8,87	1,52
La Coruña	M	509141	457	7,34	8,47	0,7480	0,2867	11,78	6,32	1,86
La Coruña	F	563190	533	8,33	8,72	0,8716	0,3791	11,20	7,06	1,59
Cuenca	M	92275	96	8,83	13,41	2,4195	1,4071	17,62	8,84	1,99
Cuenca	F	86760	87	10,73	13,36	3,0724	2,0791	16,33	10,80	1,51
Gerona	M	307975	145	1,87	3,95	0,5700	0,7954	40,35	22,56	1,79
Gerona	F	245519	138	2,15	4,67	0,7537	1,0857	40,30	22,29	1,81
Granada	M	371735	188	13,55	10,56	4,0423	0,6923	14,84	7,88	1,88
Granada	F	424598	229	16,81	14,02	4,8343	0,7568	13,08	6,20	2,11
Guadalajara	M	87591	92	1,52	3,80	0,2823	1,2206	34,88	29,10	1,20
Guadalajara	F	79560	86	2,55	4,90	0,4615	1,7868	26,63	27,28	0,98
Guipúzcoa	M	323719	279	2,60	3,25	0,9591	0,4277	37,69	20,09	1,88
Guipúzcoa	F	348524	291	4,42	4,38	1,3093	0,4294	25,90	14,95	1,73

Huelva	M	223158	121	10,46	10,37	7,2743	0,9412	25,78	9,36	2,75
Huelva	F	214587	123	9,13	10,40	4,2187	1,0980	22,49	10,07	2,23
Huesca	M	96617	125	2,56	5,39	1,2775	1,2615	44,18	20,86	2,12
Huesca	F	91147	105	3,04	6,06	1,7064	1,5781	42,92	20,72	2,07
Jaén	M	380752	233	9,63	10,28	1,8186	0,5968	14,01	7,51	1,86
Jaén	F	356344	230	11,41	11,94	2,1644	0,8385	12,89	7,67	1,68
León	M	204462	209	7,14	7,58	2,2850	0,7474	21,16	11,41	1,85
León	F	225753	228	7,56	8,31	2,2879	0,8014	20,00	10,77	1,86
Lérida	M	214123	127	9,22	9,08	4,8531	1,2797	23,88	12,46	1,92
Lérida	F	218051	133	9,34	9,77	4,5156	1,2979	22,75	11,66	1,95
La Rioja	M	149238	519	4,05	5,97	0,3139	0,2546	13,83	8,46	1,64
La Rioja	F	147554	500	4,34	7,14	0,2958	0,3245	12,52	7,98	1,57
Lugo	M	175462	169	8,64	8,40	6,9390	0,7199	30,50	10,10	3,02
Lugo	F	167892	177	5,26	9,40	1,3626	0,8026	22,20	9,54	2,33
Madrid	M	2816184	893	3,37	3,58	0,3812	0,1145	18,33	9,45	1,94
Madrid	F	3011923	996	3,59	4,26	0,3350	0,1442	16,14	8,92	1,81
Málaga	M	693871	361	8,95	9,90	1,9024	0,4162	15,41	6,52	2,37
Málaga	F	702667	397	10,80	12,04	1,9554	0,4561	12,95	5,61	2,31
Murcia	M	668714	868	7,54	8,74	0,4296	0,2175	8,69	5,34	1,63
Murcia	F	660107	902	8,30	10,31	0,4393	0,2373	7,99	4,73	1,69
Navarra	M	286947	525	2,99	2,53	0,3732	0,2389	20,45	19,28	1,06
Navarra	F	289947	603	2,73	3,31	0,2752	0,2450	19,23	14,96	1,29
Orense	M	120257	118	7,28	8,66	3,6924	1,1440	26,41	12,36	2,14
Orense	F	137587	140	4,77	7,44	2,0973	0,9954	30,34	13,41	2,26
Oviedo	M	511169	742	2,54	4,95	0,2335	0,1618	19,02	8,12	2,34
Oviedo	F	546817	864	3,11	5,14	0,2300	0,1857	15,41	8,38	1,84
Palencia	M	75638	71	5,65	9,10	2,9335	2,1179	30,32	15,99	1,90
Palencia	F	72558	72	6,08	10,92	3,1612	2,5262	29,27	14,56	2,01
Las Palmas	M	592262	458	7,63	8,40	1,5179	0,3192	16,15	6,73	2,40
Las Palmas	F	580265	485	8,46	8,78	1,6326	0,3139	15,11	6,38	2,37
Pontevedra	M	494161	434	3,00	6,09	0,1949	0,3100	14,73	9,14	1,61
Pontevedra	F	525627	462	4,40	7,57	0,3114	0,3888	12,68	8,24	1,54
Salamanca	M	151335	166	9,87	11,50	2,3273	0,7451	15,46	7,51	2,06
Salamanca	F	152234	162	8,85	12,74	2,0322	0,8510	16,11	7,24	2,22
Tenerife	M	366253	370	8,07	8,20	1,0073	0,2909	12,44	6,58	1,89
Tenerife	F	376690	392	9,35	9,18	1,2484	0,4136	11,95	7,00	1,71
Santander	M	267290	424	2,59	4,94	0,3045	0,2634	21,30	10,38	2,05
Santander	F	279191	443	2,95	5,29	0,3210	0,3526	19,20	11,22	1,71
Segovia	M	62518	57	7,01	6,30	4,5293	2,1717	30,36	23,41	1,30
Segovia	F	63217	58	10,90	9,64	5,5058	2,9114	21,52	17,71	1,22
Sevilla	M	816795	472	3,42	6,34	0,1609	0,2819	11,72	8,38	1,40
Sevilla	F	853057	491	4,53	8,14	0,3045	0,3112	12,19	6,85	1,78
Soria	M	26431	24	15,28	9,13	76,9805	3,9189	57,42	21,68	2,65
Soria	F	17211	17	23,46	11,84	122,9756	5,5980	47,27	19,99	2,37
Tarragona	M	264627	129	1,95	4,53	0,2800	1,0997	27,15	23,14	1,17
Tarragona	F	255490	139	2,79	6,16	0,4304	1,1317	23,52	17,26	1,36
Teruel	M	53380	66	4,48	5,49	5,9649	1,6368	54,54	23,29	2,34
Teruel	F	65002	78	5,16	7,38	3,8559	1,9972	38,09	19,15	1,99
Toledo	M	288335	278	7,69	9,18	1,3151	0,3438	14,92	6,39	2,33
Toledo	F	305241	272	5,85	7,58	0,8997	0,6082	16,21	10,28	1,58
Valencia	M	1169258	686	5,08	5,70	0,9538	0,1722	19,24	7,28	2,64
Valencia	F	1197478	742	4,26	6,78	0,3187	0,2456	13,25	7,31	1,81
Valladolid	M	305496	292	6,38	4,71	1,1430	0,4269	16,75	13,87	1,21
Valladolid	F	322530	306	7,45	5,82	1,3767	0,4782	15,76	11,89	1,33

Vizcaya	M	576042	515	2,57	2,80	0,2783	0,2338	20,49	17,27	1,19
Vizcaya	F	590094	532	2,26	3,35	0,1756	0,2177	18,56	13,92	1,33
Zamora	M	101433	109	12,58	13,10	5,5333	1,5147	18,71	9,40	1,99
Zamora	F	98337	100	9,86	12,18	4,7252	1,7185	22,04	10,76	2,05
Zaragoza	M	466651	555	4,29	4,77	0,7891	0,2377	20,69	10,23	2,02
Zaragoza	F	462937	574	5,08	4,72	0,9837	0,1956	19,53	9,38	2,08
Ceuta	M	35705	223	14,79	11,09	3,3694	0,7296	12,41	7,70	1,61
Ceuta	F	40426	247	20,68	12,52	5,5107	0,8832	11,35	7,50	1,51
Melilla	M	30595	179	11,87	6,22	7,3207	0,7442	22,80	13,86	1,64
Melilla	F	27498	180	12,47	8,82	3,5770	1,1392	15,16	12,10	1,25

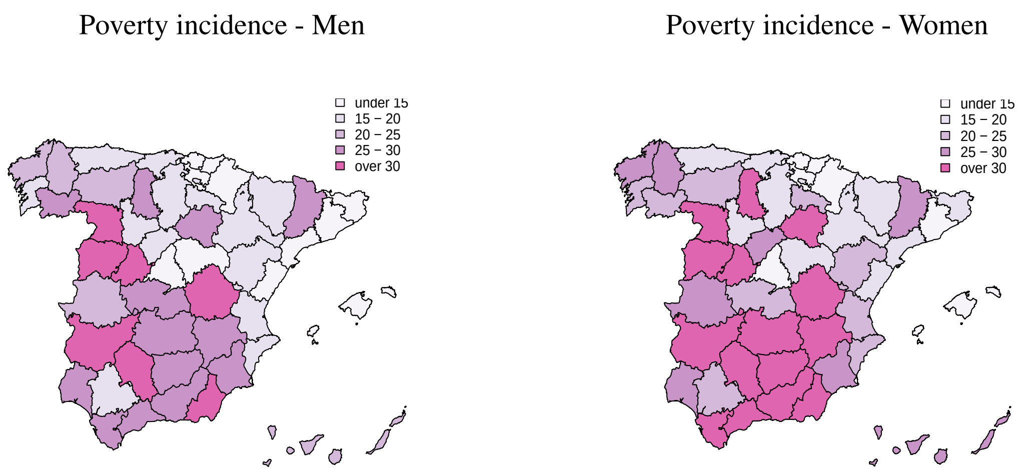


Figure 5.1: Cartograms of estimated percent poverty incidences in Spanish provinces for Men and Women.

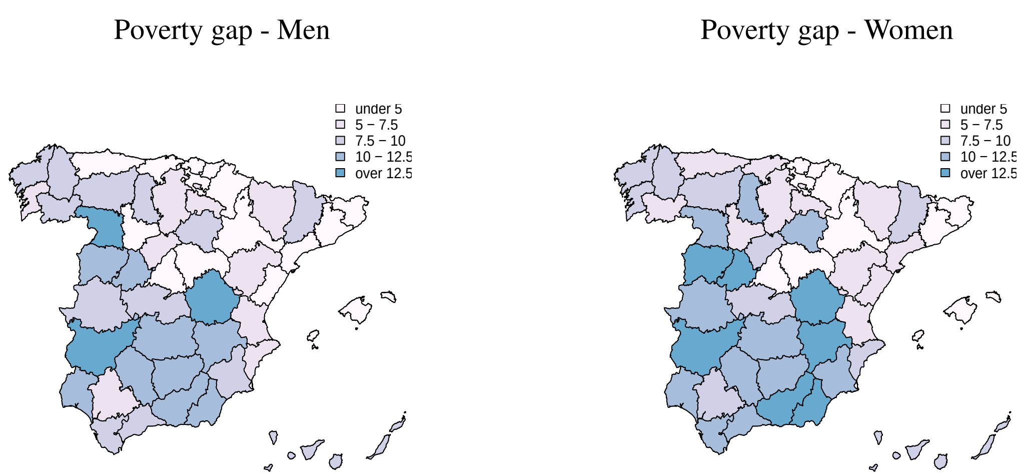


Figure 5.2: Cartograms of estimated percent poverty gaps in Spanish provinces for Men and Women.

Chapter 6

Fast EB method for estimation of fuzzy poverty measures

6.1 Fuzzy monetary and supplementary indicators

Let $U = \{1, \dots, N\}$ be a finite population of size N, where E_i is the value of a welfare variable (e.g. equivalised income) for individual i. Let us consider the empirical distribution function of $\{E_1, \dots, E_N\}$, defined as

$$F_E(x) = \frac{1}{N}\sum_{j=1}^{N} I\{E_j \leq x\}, \quad x \in \mathbf{R},$$

where $I\left\{E_j \leq x\right\} = 1$ if $E_j \leq x$ and 0 otherwise. Consider also the (empirical) Lorenz curve, given by

$$L_E(x) = \frac{\sum_{j=1}^{N} E_j I\left\{E_j \leq x\right\}}{\sum_{j=1}^{N} E_j}, \quad x \in \mathbf{R}.$$

Following the Integrated Fuzzy and Relative (IFR) approach of Betti et al. (2006), the Fuzzy Monetary Index (FMI) for individual i is defined as

$$\begin{aligned} FM_i &= \left\{\frac{N}{N-1}(1 - F_E(E_i))\right\}^{\alpha-1} \{1 - L_E(E_i)\} \\ &= \left\{\frac{1}{N-1}\sum_{j=1}^{N} I\left\{E_j > E_i\right\}\right\}^{\alpha-1} \left\{\frac{\sum_{j=1}^{N} E_j I\left\{E_j > E_i\right\}}{\sum_{j=1}^{N} E_j}\right\}, \quad i \in U. \end{aligned}$$

Here, $1 - F_E(E_i)$ is the proportion of individuals that are less poor than individual i. This gives a degree of poverty of individual i and it was proposed by Cheli e Lemmi (1995) as a poverty indicator. Observe that $N(1 - F_E(E_i))/(N-1)$ is equal to 1 when individual i is the poorest. Moreover, $1 - L_E(E_i)$

is the share of the total welfare of all individuals that are less poor than this individual, indicator that was proposed by Betti and Verma (1999). The average FMI for the population is given by

$$FM = \frac{1}{N}\sum_{i=1}^{N} FM_i \tag{6.1}$$

Observe that the FMI for individual i depends on the whole population of welfare values, $\{E_1,\ldots,E_N\}$.

Consider now a score variable S_i for i-th individual defined using the IFR approach, instead of a welfare variable E_i. These scores S_i are obtained by applying a multidimensional approach that takes into account a variety of non-monetary indicators of deprivation. Then the Fuzzy Supplementary Index (FSI) for individual i is defined analogously to the FMI, but in terms of the scores $\{S_1,\ldots,S_N\}$, as

$$\begin{aligned} FS_i &= \left\{\frac{N}{N-1}(1-F_S(S_i))\right\}^{\alpha-1}\{1-L_S(S_i)\} \\ &= \left\{\frac{1}{N-1}\sum_{j=1}^{N} I\{S_j > S_i\}\right\}^{\alpha-1}\left\{\frac{\sum_{j=1}^{N} S_j I\{S_j > S_i\}}{\sum_{j=1}^{N} S_j}\right\}, \quad i \in U. \end{aligned}$$

Here, $F_S(x)$ is the empirical distribution function and $L_S(x)$ the Lorenz curve of the score variables $\{S_1,\ldots,S_N\}$. Similarly, $1-F_S(S_i)$ is the proportion of individuals who are less deprived than individual i and $1-L_S(S_i)$ is the share of the total lack of deprivation score assigned to all individuals less deprived than individual i. The average FSI for the population is given by

$$FS = \frac{1}{N}\sum_{i=1}^{N} FS_i \tag{6.2}$$

Now consider that the population U is partitioned into D domains or areas $U_1,\ldots,U_D$ of sizes $N_1,\ldots,N_D$. Let E_{dj} be the welfare for individual j within domain d. The average fuzzy monetary index for domain d is

$$FM_d = \frac{1}{N_d}\sum_{j=1}^{N_d} FM_{dj}, \quad d = 1,\ldots,D, \tag{6.3}$$

where FM_{dj} is the FMI for j-th individual from d-th domain.

A random sample $s \subseteq U$ of size $n \le N$ is drawn from the population. Let s_d be the subsample from domain d, $d = 1,\ldots,D$. A design-based estimator of the average FMI for domain d, FM_d, is

$$\widehat{FM}_d^{DB} = \frac{\sum_{j\in s_d} w_{dj}\widehat{FM}_{dj}^{DB}}{\sum_{j\in s_d} w_{dj}}, \quad d = 1,\ldots,D, \tag{6.4}$$

where w_{dj} is the sampling weight for individual j within domain d and

$$\widehat{FM}_{dj}^{DB} = \left\{\frac{\sum_{\ell=1}^{D}\sum_{i\in s_d} w_{\ell i} I\{E_{\ell i} > E_{dj}\}}{\sum_{\ell=1}^{D}\sum_{i\in s_d} w_{\ell i}}\right\}^{\alpha-1}\left\{\frac{\sum_{\ell=1}^{D}\sum_{i\in s_d} w_{\ell i} E_{\ell i} I\{E_{\ell i} > E_{dj}\}}{\sum_{\ell=1}^{D}\sum_{i\in s_d} w_{\ell i} E_{\ell i}}\right\}. \tag{6.5}$$

Observe that $\widehat{FM}_{dj}^{DB}$ is not a direct estimator because it uses the sample data from the whole population and not only from domain d. The average FSI for domain d is given by

$$FS_d = \frac{1}{N_d}\sum_{j=1}^{N_d} FS_{dj}, \quad d = 1,\ldots,D. \tag{6.6}$$

Finally, a design-based estimator of FS_d would be

$$\widehat{FS}_d^{DB} = \frac{\sum_{j\in s_d} w_{dj}\widehat{FS}_{dj}^{DB}}{\sum_{j\in s_d} w_{dj}}, \quad d = 1,\ldots,D. \tag{6.7}$$

where

$$\widehat{FS}_{dj}^{DB} = \left\{\frac{\sum_{\ell=1}^{D}\sum_{i\in s_d} w_{\ell i} I\{S_{\ell i} > S_{dj}\}}{\sum_{\ell=1}^{D}\sum_{i\in s_d} w_{\ell i}}\right\}^{\alpha-1}\left\{\frac{\sum_{\ell=1}^{D}\sum_{i\in s_d} w_{\ell i} S_{\ell i} I\{S_{\ell i} > S_{dj}\}}{\sum_{\ell=1}^{D}\sum_{i\in s_d} w_{\ell i} S_{\ell i}}\right\}. \tag{6.8}$$

In these poverty indicators, the parameter α can be fixed to the value such that the FM and FS indicators coincide with the head count ratio computed for the official poverty line (60% of the median).

6.2 Fast EB method for estimation of fuzzy poverty measures

In order to apply the EB method of Molina and Rao (2010) to estimate the domain average FMI, FM_d, we need to express this indicator in terms of a population vector $\mathbf{y} = (\mathbf{y}_s', \mathbf{y}_r')'$, for which the conditional distribution of the non-sampled part $\mathbf{y}_r$ given the sample data $\mathbf{y}_s$ is known. The distribution of the welfare variable E_{dj} is seldom Normal. However, many times it is possible to find a transformation whose distribution is approximately Normal. Suppose that there exists a one-to-one transformation $Y_{dj} = T(E_{dj})$ of the welfare variable E_{dj}, which follows a Normal distribution. Concretely, we assume that Y_{dj} follows the nested error linear regression model of Battese, Harter and Fuller (1988), defined as

$$\begin{aligned} &Y_{dj} = \mathbf{x}_{dj}\beta + u_d + e_{dj}, \qquad j = 1,\ldots,N_d, \quad d = 1,\ldots D, \\ &u_d \sim \text{ iid } N(0,\sigma_u^2), \quad e_{dj} \sim \text{ iid } N(0,\sigma_e^2) \end{aligned} \tag{6.9}$$

where $\mathbf{x}_{dj}$ is a row vector with the values of p explanatory variables, u_d is a random area-specific effect and e_{dj} are residual errors. Let $\mathbf{y}_d = (Y_{d1},\ldots,Y_{dN_D})'$ be vector of responses for domain d and $\mathbf{y} = (\mathbf{y}_1',\ldots,\mathbf{y}_D')'$ be the full population vector. Then, observe that the individual FMIs can be expressed as

$$\begin{aligned} FM_{dj} &= \left\{\frac{1}{N-1}\sum_{\ell=1}^{D}\sum_{i=1}^{N_\ell} I\left\{T^{-1}(Y_{\ell i}) > T^{-1}(Y_{dj})\right\}\right\}^{\alpha-1} \\ &\times \left\{\frac{\sum_{\ell=1}^{D}\sum_{i=1}^{N_\ell} T^{-1}(Y_{\ell i}) I\left\{T^{-1}(Y_{\ell i}) > T^{-1}(Y_{dj})\right\}}{\sum_{\ell=1}^{D}\sum_{i=1}^{N_\ell} T^{-1}(Y_{\ell i})}\right\}, \quad j = 1,\ldots,N_d,\ d = 1,\ldots,D. \end{aligned}$$

This means that the average FMI for domain d is a non-linear function of the population vector $\mathbf{y}$, that is,

$$FM_d = \frac{1}{N_d}\sum_{j=1}^{N_d} FM_{dj} = h_d(\mathbf{y}), \quad d = 1,\ldots,D.$$

Let us separate the population vector of responses $\mathbf{y}$ in the sample and non-sample parts, that is, $\mathbf{y} = (\mathbf{y}_s', \mathbf{y}_r')'$, where $\mathbf{y}_s$ corresponds to the sample and $\mathbf{y}_r$ to the non-sample. Then the BP of FM_d is

$$\widehat{FM}_d^B = E_{\mathbf{y}_r}(FM_d|\mathbf{y}_s) = E_{\mathbf{y}_r}(h_d(\mathbf{y})|\mathbf{y}_s). \tag{6.10}$$

This expectation can be empirically approximated by Monte Carlo simulation. For this, first fit the nested-error model (6.9) to the sample data $\mathbf{y}_s$, to obtain estimates $\hat{\beta}$, $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$ of the model parameters β, σ_u^2 and σ_e^2 respectively. Obtain also the EB predictor $\hat{u}_d$ of u_d, given by $E(u_d|\mathbf{y}_s)$ with unknown parameters replaced by estimated values. Then, using those estimates, generate a large number L of vectors $\mathbf{y}_r$ from the estimated conditional distribution $\mathbf{y}_r|\mathbf{y}_s$. Let $\mathbf{y}_r^{(l)}$ be the vector generated in l-th generation. We attach this vector to the sample vector to obtain the full population vector $\mathbf{y}^{(l)} = (\mathbf{y}_s', (\mathbf{y}_r^{(l)})')'$. Using the elements of $\mathbf{y}^{(l)}$, we calculate the domain parameter of interest $FM_d^{(l)} = h_d(\mathbf{y}^{(l)})$, $d = 1,\ldots,D$. Then, a Monte Carlo approximation to the EB predictor of FM_d is given by

$$\widehat{FM}_d^{EB} \approx \frac{1}{L}\sum_{l=1}^{L} FM_d^{(l)}, \quad d = 1,\ldots,D. \tag{6.11}$$

Observe that for each population $l = 1,\ldots,L$, instead of generating a multivariate normal vector of size $N-n$, we just need to generate univariate values Y_{dj} from

$$Y_{dj} = \mathbf{x}_{dj}\hat{\beta} + \hat{u}_d + v_d + \varepsilon_{di},\ v_d \sim N(0, \hat{\sigma}_u^2(1-\hat{\gamma}_d)),\ \varepsilon_{dj} \sim N(0,\hat{\sigma}_e^2),\ j \in U_d - s_d,\ d = 1,\ldots,D, \tag{6.12}$$

where $\gamma_d = \sigma_u^2(\sigma_u^2 + \sigma_e^2/n_d)^{-1}$ and n_d is the sample size in domain d. Still, for large populations and/or complex indicators, the EB method can be unfeasible. FMIs require sorting of all population elements, and this needs to be repeated for $l = 1,\ldots,L$. This is too time consuming for large N and large L. Here we propose a faster version the EB estimator that is based on replacing the true value of the domain average FMI in population l, $FM_d^{(l)}$, by the design-based estimator given in (6.4). Since the design-based estimator is obtained from a sample drawn from l-th population, this avoids the task of generation of the full population of responses (we need to generate only the responses for the sample elements) and the sorting of all the population elements. Concretely, for each Monte Carlo replication l, we take a sample $s(l) \subseteq U$ using the same sampling scheme and the same sample size allocation as in the original sample s. We take the values of the auxiliary variables corresponding to the units in $s(l)$, that is, we take $\mathbf{x}_{dj}$, $j \in s_d(l)$, where $s_d(l)$ is the subsample from d-th domain. Then we generate the corresponding responses Y_{dj}, $j \in s_d(l)$, for $d = 1,\ldots,D$, as in (6.12). Let us denote the vector containing those values as $\mathbf{y}_{s(l)}$. With $\mathbf{y}_{s(l)}$, calculate the design-based estimator as in (6.4) and (6.5), that is, obtain

$$\widehat{FM}_d^{DB}(l) = \frac{\sum_{j\in s_d(l)} w_{dj}\widehat{FM}_{dj}^{DB}(l)}{\sum_{j\in s_d(l)} w_{dj}}, \quad d = 1,\ldots,D, \tag{6.13}$$

where

$$\widehat{FM}_{dj}^{DB}(l) = \left\{ \frac{\sum_{\ell=1}^{D} \sum_{i \in s_d(l)} w_{\ell i} I\{E_{\ell i} > E_{dj}\}}{\sum_{\ell=1}^{D} \sum_{i \in s_d(l)} w_{\ell i}} \right\}^{\alpha-1} \left\{ \frac{\sum_{\ell=1}^{D} \sum_{i \in s_d(l)} w_{\ell i} E_{\ell i} I\{E_{\ell i} > E_{dj}\}}{\sum_{\ell=1}^{D} \sum_{i \in s_d(l)} w_{\ell i} E_{\ell i}} \right\}.$$

Finally, the fast EB estimator of FM_d is given by

$$\widehat{FM}_{d}^{FEB} = \frac{1}{L} \sum_{l=1}^{L} \widehat{FM}_{dj}^{DB}(l), \quad d = 1, \ldots, D.$$

6.3 Application to EU-SILC data from Tuscany

The fast EB method described in the previous section was applied to estimate the poverty incidence or head count ratio (HCR), the fuzzy monetary (FM) and the fuzzy supplementary (FS) indicators in Tuscany provinces. Data from the 2004 Italian SILC survey was used.

The regional sample from Tuscany is based on a stratified two stage sample design: in each province, the municipalities are the primary sampling units (PSUs), and these are then divided into strata according to their population size. From these PSUs, households (SSUs) are selected by means of systematic sampling. Some provinces, generally the smaller ones, may have very few sampled municipalities and many municipalities are not represented in the sample. For example, in the 2004 survey, only 53 municipalities out of the total of 287 are present in the sample. Then, small area estimation techniques can be required given large errors of direct estimators at province level or the impossibility to compute them at municipalities level.

In our analysis the small areas of interest are the 10 Tuscany provinces, with sample sizes ranging from 155 (Province of Grosseto) to 1403 (Province of Firenze). The regional sample size is of 4426 individuals. The welfare variable for the individuals is the equivalized annual net income. In order to overcome the problem of negative values of this variable, we followed the recommendation of Eurostat: applying a bottom coding strategy to the lowest values of the distribution. In particular, all values below 15% of the median household income have been set equal to the 15% of the median. This strategy does not affect the poverty line and consequently, neither the design-based estimators (Eurostat, 2006; Ciampalini *et al*., 2009; Neri *et al*., 2009). The equivalized annual net income has been transformed by taking logarithm to obtain a distribution approximately normal. This transformed variable acts as the response in the nested-error regression model. As auxiliary variables, we have considered the indicators of 5 quinquennial groupings of variable age, the indicator of having Italian nationality, the indicators of 3 levels of the variable education level and 3 categories of the variable employment.

The poverty line for the calculation of HCR is computed as the 60% of the weighted median of the individual equivalent income at Regional level and is equal to 9,372.24 Euros.

Design-based and fast EB estimators of the HCR, FM and FS indicators were calculated. We took $\alpha = 2$, avoiding any numerical link with the traditional approach, because the primary objective of this analysis is to develop methodologies for estimating fuzzy measures in small domains rather than numerical comparisons with the conventional approach. Values of design-based and fast EB estimators of HCR and their associated coefficients of variation (CV) are shown in Table 6.1 for each Tuscany province. The average over provinces is 16.4%. The poorer provinces concentrate mainly

Province	N_d	n_d	$\widehat{HCR}_d^{DB}$	$\widehat{HCR}_d^{fastEB}$	cv($\widehat{HCR}_d^{DIR}$)	cv($\widehat{HCR}_d^{fastEB}$)
Arezzo	304121	416	0.087	0.130	19.09	12.42
Firenze	1119377	1403	0.133	0.144	8.32	10.89
Grosseto	149082	155	0.124	0.147	26.10	11.10
Livorno	290122	339	0.131	0.149	15.61	10.30
Siena	278495	338	0.110	0.156	19.31	10.80
Prato	319320	416	0.170	0.159	14.06	10.77
Pistoia	267076	344	0.174	0.169	13.75	9.87
Pisa	335777	399	0.168	0.178	15.54	8.88
Lucca	265293	315	0.215	0.182	13.30	8.88
Massa Carrara	251471	301	0.260	0.224	13.09	7.30
Average			**0.157**	**0.164**		

Table 6.1: Population size, sample size, direct and new EB estimators of HCR, CVs of direct and new EB estimators ($\times$100) for Tuscany Provinces.

in the north-west of Tuscany. The province of Massa has the highest percentage of poor individuals (22.4%) followed by Lucca (18.2%) and Pisa (17.8%). On the other hand, the provinces of Arezzo (13.0%) and Firenze (14.4%) are the richest. The MSEs of fast EB estimators of HCR are calculated using the parametric bootstrap with $B = 500$ replicates. The coefficient of variation is given by $\text{cv}(\widehat{HCR}_d^{fastEB}) = \{\text{mse}(\widehat{HCR}_d^{fastEB})^{1/2} \big/ \widehat{HCR}_d^{fastEB}\}$. Results in Table 6.1 show that the CVs of fast EB estimators are much smaller than those of design-based estimators and the reduction in CV tends to be greater for domains with smaller sample size. The only exception is Province of Firenze with a large sample size, for which the CV of the fast EB estimator is larger than that of the design-based estimator.

Table 6.2 shows respectively the design-based and fast EB estimators of the FM and FS indicators and the combination of the two. The fast EB estimators of the FMI provide the same picture concerning the monetary poverty in the small areas as the HCR. The values of the FMI are larger than those of the HCR since in each province there is a concentration of individuals with equivalised income just above the poverty line. The provinces of Arezzo (36.5%) and Firenze (38.0%) remain the richest, whereas the province of Massa has the largest percentage of poor individuals (47.5%) followed by Lucca (42.6%) and Pisa (42.3%). The FSIs have been also estimated for Tuscany provinces. In this case, the welfare variable is the score, constructed as explained in Section 6.1. As response variable in the nested-error regression model, we have taken the clog-log transformation of the score. We used the same auxiliary variables employed for the estimation of HCR and FMI.

Concerning non-monetary deprivation as measured by the FSIs, the ranking of some of the provinces is completely opposite as that obtained from the monetary poverty measures, see Table 6.2. For example, the province of Massa has large values of the FMI and small values of the FSI, whereas the opposite holds for Firenze and Livorno provinces.

The non-monetary dimension is combined with the monetary dimension in order to obtain measures of manifest (MAN) and latent (LAT) deprivation, which correspond respectively to the intersection and the union of fuzzy sets. Note from Table 5.2 that there are some differences between provinces. The overlapping takes the smallest value of 27.6% for Grosseto and the largest of 39.4% for Pistoia. In general, the MAN/LAT ratio is smaller for areas with lower levels of deprivation, and larger for areas

Province	n_d	$\widehat{FM}_d^{\alpha DB}$	$\widehat{FM}_d^{\alpha fastEB}$	$\widehat{FS}_d^{\alpha DB}$	$\widehat{FS}_d^{\alpha fastEB}$	LAT	MAN	MAN/LAT
Arezzo	416	0.354	0.365	0.262	0.297	0.494	0.167	0.338
Firenze	1403	0.376	0.380	0.371	0.368	0.542	0.206	0.380
Grosseto	155	0.390	0.383	0.158	0.203	0.460	0.127	0.276
Livorno	339	0.379	0.392	0.370	0.372	0.552	0.212	0.384
Siena	338	0.381	0.396	0.306	0.321	0.526	0.191	0.363
Prato	416	0.402	0.404	0.332	0.347	0.545	0.206	0.377
Pistoia	344	0.424	0.414	0.411	0.380	0.570	0.225	0.394
Pisa	399	0.433	0.423	0.353	0.348	0.558	0.212	0.381
Lucca	315	0.424	0.426	0.398	0.360	0.566	0.220	0.388
Massa Carrara	301	0.496	0.475	0.330	0.316	0.578	0.213	0.368
Average		**0.406**	**0.406**	**0.329**	**0.331**			

Table 6.2: Sample size, direct and new EB estimators of FM and FS indicators, latent (LAT) and manifest (MAN) deprivation, ratio MAN/LAT for Tuscany Provinces.

with higher levels. A large value of this ratio means that different types of deprivation overlap and this means that in areas where levels of relative deprivation are already high, deprivation in the income and non-monetary domains is more likely to afflict the same individuals in the population. On the other hand, low values imply the absence of such overlap at micro level.

Figures 6.1 and 6.2 show respectively the cartograms of the estimated head count ratio, fuzzy monetary indicators, fuzzy supplementary indicators and manifest/latent ratio in Tuscany provinces constructed using the new EB estimators.

6.4 Concluding remarks

The application of the EB method to Tuscany data showed some limitations, due mainly to the non-normality of the distribution of the response variable. This seems to cause some bias in the EB estimator. Mean square errors of fast EB estimators of fuzzy poverty measures have to be calculated, and bootstrap is in this case too computationally expensive.

Concerning fuzzy measures, as explained before, in our analysis the parameter α has been fixed to 2. Determining the value of α to make the FM and FS indicators numerically identical to the head count ratio deserves also further study.

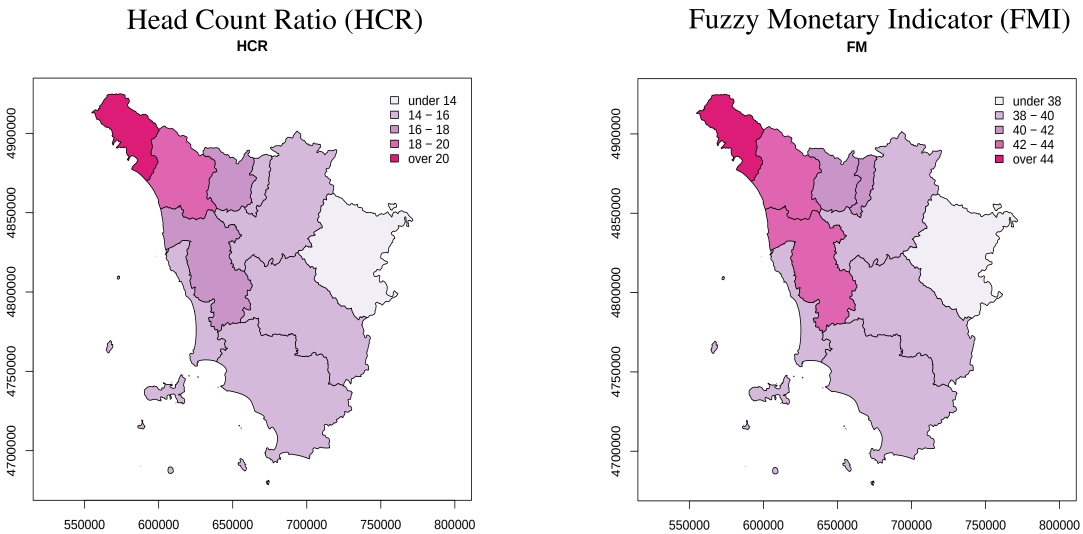


Figure 6.1: Cartograms of estimated percent head count ratio and fuzzy monetary indicators in Tuscany provinces.

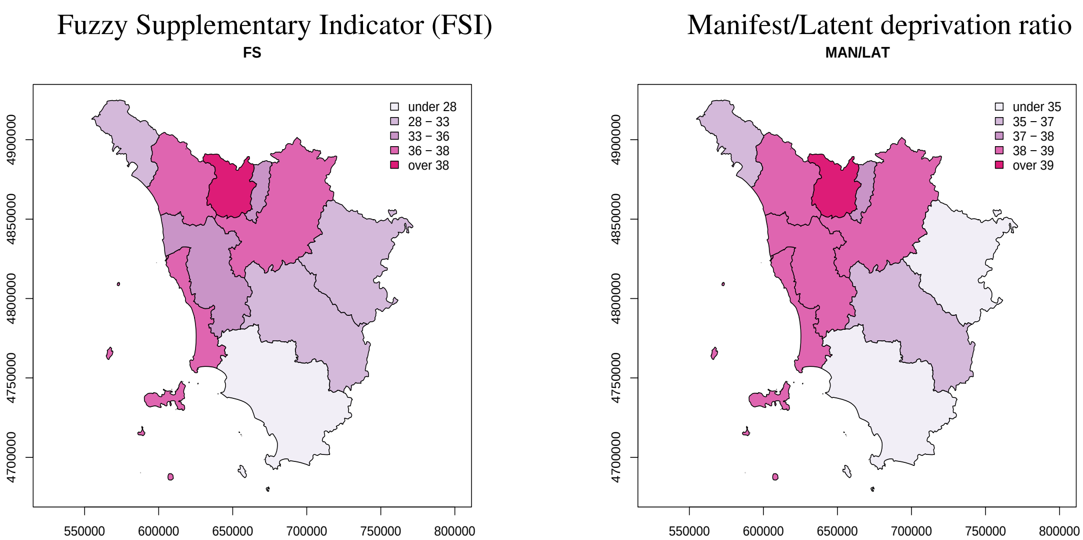


Figure 6.2: Cartograms of estimated percent fuzzy supplementary indicators and manifest/latent deprivation ratio in Tuscany provinces.

Chapter 7

Spatial Fay-Herriot models

7.1 The model

Consider the Spatial Fay-Herriot model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{v} + \mathbf{e}, \tag{7.1}$$

where $\mathbf{y} = (y_1, \dots, y_D)'$, where y_d is the direct estimator of the quantity of interest $\theta_d = \mathbf{x}_d\beta + v_d$ for domain d, $d = 1, \dots, D$, $\mathbf{v} = (v_1, \dots, v_D)'$ is a vector with random effects of the domains, $\mathbf{e} = (e_1, \dots, e_D)'$ is a vector of domain errors, $\mathbf{X} = (\mathbf{x}'_1, \dots, \mathbf{x}'_D)'$ is the matrix with the values of p auxiliary variables for the domains and $\Psi = \text{diag}(\psi_1, \dots, \psi_D)$ is the matrix with the design-based variances of direct estimators. Here, $\mathbf{v}$ is the result of a SAR process with unknown autoregression parameter ρ and proximity matrix $\mathbf{W}$ (see Anselin (1988) and Cressie(1993)), i.e.,

$$\mathbf{v} = \rho\mathbf{W}\mathbf{v} + \mathbf{u}. \tag{7.2}$$

The vector $\mathbf{u} = (u_1, \dots, u_D)'$ has mean vector $\mathbf{0}$ and covariance matrix $\sigma_u^2\mathbf{I}_D$, where $\mathbf{I}_D$ denotes the $D \times D$ identity matrix and σ_u^2 is an unknown parameter. We consider that the proximity matrix $\mathbf{W}$ is defined in row standardized form; that is, $\mathbf{W}$ is row stochastic and that the matrix $(\mathbf{I}_D - \rho\mathbf{W})$ is non-singular. Then, $\rho \in (-1, 1)$ is called spatial autocorrelation parameter Banerjee et al.(2004). Hereafter, the vector of variance components will be denoted $\omega = (\omega_1, \omega_2)' = (\sigma_u^2, \rho)'$. Observe that $\mathbf{v}$ has mean vector $\mathbf{0}$ and covariance matrix equal to

$$\mathbf{G}(\omega) = \sigma_u^2[(\mathbf{I}_D - \rho\mathbf{W})'(\mathbf{I}_D - \rho\mathbf{W})]^{-1}. \tag{7.3}$$

Since $\mathbf{e}$ is independent of $\mathbf{v}$, the covariance matrix of $\mathbf{y}$ is equal to

$$\mathbf{V}(\omega) = \mathbf{G}(\omega) + \Psi.$$

The Spatial BLUP of the quantity of interest, $\theta_d = \mathbf{x}_d\beta + v_d$ is

$$\tilde{\theta}_d(\omega) = \mathbf{x}_d\tilde{\beta}(\omega) + \mathbf{b}'_d\mathbf{G}(\omega)\mathbf{V}^{-1}(\omega)[\mathbf{y} - \mathbf{X}\tilde{\beta}(\omega)], \tag{7.4}$$

where

$$\tilde{\beta}(\omega) = [\mathbf{X}'\mathbf{V}^{-1}(\omega)\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}(\omega)\mathbf{y}.$$

and $\mathbf{b}'_d$ is the $1 \times D$ vector $(0, \dots, 0, 1, 0, \dots, 0)$ with 1 in the d-th position. The Spatial BLUP $\tilde{\theta}_d(\omega)$ depends on the unknown vector of variance components $\omega = (\sigma_u^2, \rho)'$. The Spatial EBLUP is obtained by replacing ω in expression (7.4) by a consistent estimator $\hat{\omega} = (\hat{\sigma}_u^2, \hat{\rho})'$, that is, $\tilde{\theta}_d(\hat{\omega})$.

7.2 Analytical approximation of the MSE

Singh et al. (2005) obtained an approximately unbiased MSE estimator of the Spatial EBLUP. When $\hat{\omega} = (\hat{\sigma}_u^2, \hat{\rho})'$ is obtained by ML, the estimator is

$$mse_{ML}^{PR}[\tilde{\theta}_d(\hat{\omega})] = g_{1d}(\hat{\omega}) + g_{2d}(\hat{\omega}) + 2g_{3d}(\hat{\omega}) - \mathbf{b}_{ML}^T(\hat{\omega})\nabla g_{1d}(\hat{\omega}) - g_{4d}(\hat{\omega}), \quad (7.5)$$

where

$$\begin{aligned} g_{1d}(\omega) &= \mathbf{b}_d'[\mathbf{G}(\omega) - \mathbf{G}(\omega)\mathbf{V}^{-1}(\omega)\mathbf{G}(\omega)]\mathbf{b}_d, \\ g_{2d}(\omega) &= \mathbf{b}_d'[\mathbf{I}_D - \mathbf{G}(\omega)\mathbf{V}^{-1}(\omega)]\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}(\omega)\mathbf{X})^{-1}\mathbf{X}'[\mathbf{I}_D - \mathbf{V}^{-1}(\omega)\mathbf{G}(\omega)]\mathbf{b}_d, \\ g_{3d}(\omega) &= \text{trace}\left\{\mathbf{L}_d(\omega)\mathbf{V}(\omega)\mathbf{L}_d'(\omega)I^{-1}(\omega)\right\}. \end{aligned}$$

Here, we have $\mathbf{C}(\rho) = (\mathbf{I}_D - \rho\mathbf{W})'(\mathbf{I}_D - \rho\mathbf{W})$ and

$$\mathbf{L}_d(\omega) = \begin{pmatrix} \mathbf{b}_d'\left[\mathbf{C}^{-1}(\rho)\mathbf{V}^{-1}(\omega) - \sigma_u^2\mathbf{C}^{-1}(\rho)\mathbf{V}^{-1}(\omega)\mathbf{C}^{-1}(\rho)\mathbf{V}^{-1}(\omega)\right] \\ \mathbf{b}_d'\left[\mathbf{A}(\omega)\mathbf{V}^{-1}(\omega) - \sigma_u^2\mathbf{C}^{-1}(\rho)\mathbf{V}^{-1}(\omega)\mathbf{A}(\omega)\mathbf{V}^{-1}(\omega)\right] \end{pmatrix}.$$

Moreover, $\nabla g_{1d}(\omega) = \partial g_{1d}(\omega)/\partial\omega$ is the gradient of $g_{1d}(\omega)$ and $\mathbf{b}_{ML}(\hat{\omega})$ is the bias of the ML estimator $\hat{\omega}$ up to order $o(D^{-1})$. This bias is equal to $\mathbf{b}_{ML}(\hat{\omega}) = I^{-1}(\hat{\omega})\mathbf{h}(\hat{\omega})/2$, where $I^{-1}(\hat{\omega})$ is the Fisher information matrix of $\omega = (\sigma_u^2, \rho)'$ evaluated at $\hat{\omega} = (\hat{\sigma}_u^2, \hat{\rho})'$ and $\mathbf{h}(\hat{\omega}) = (h_1(\hat{\omega}), h_2(\hat{\omega}))'$ with

$$h_k(\omega) = \text{trace}\left\{\left[\mathbf{X}'\mathbf{V}^{-1}(\omega)\mathbf{X}\right]^{-1}\frac{\partial\left[\mathbf{X}'\mathbf{V}^{-1}(\omega)\mathbf{X}\right]}{\partial\omega_k}\right\}, \quad k = 1, 2.$$

Finally,

$$g_{4d}(\omega) = \frac{1}{2}\sum_{k=1}^{2}\sum_{\ell=1}^{2}\mathbf{b}_d'\Psi\mathbf{V}^{-1}(\omega)\frac{\partial^2\mathbf{V}(\omega)}{\partial\omega_k\partial\omega_\ell}\mathbf{V}^{-1}(\omega)\Psi\, I_{k\ell}^{-1}(\omega)\mathbf{b}_d.$$

When $\hat{\omega}$ is obtained by REML method, Singh et al. (2005) give the estimator

$$mse^{SSK}[\tilde{\theta}_d(\hat{\omega})] = g_{1d}(\hat{\omega}) + g_{2d}(\hat{\omega}) + 2g_{3d}^{PR}(\hat{\omega}) - g_{4d}(\hat{\omega}). \quad (7.6)$$

7.3 Parametric bootstrap estimation of the MSE

The parametric bootstrap of González-Manteiga et al. (2007) is used to derive an estimator of the MSE of the Spatial EBLUP. This method works as follows:

(1) Obtain the estimates $\hat{\omega} = (\hat{\sigma}_u^2, \hat{\rho})'$ and $\hat{\beta} = \tilde{\beta}(\hat{\omega})$ by fitting the Spatial FH model (7.1) to the initial data $\mathbf{y} = (y_1, \ldots, y_D)'$.

(2) Generate a vector $\mathbf{t}_1^*$ whose D elements are independent $N(0,1)$. Build bootstrap vectors $\mathbf{u}^* = \hat{\sigma}_u\mathbf{t}_1^*$ and $\mathbf{v}^* = (\mathbf{I}_D - \hat{\rho}\mathbf{W})^{-1}\mathbf{u}^*$, and calculate $\theta^* = \mathbf{X}\hat{\beta} + \mathbf{v}^*$, where $\hat{\beta}$ and $\hat{\omega}$ are viewed as the true values of the parameters.

(3) Generate a vector $\mathbf{t}_2^*$ with D independent $N(0,1)$ elements, which is independent of $\mathbf{t}_1^*$. Then, construct the vector of random errors as $\mathbf{e}^* = \Psi^{1/2}\mathbf{t}_2^*$.

(4) Obtain bootstrap data $\mathbf{y}^*$ directly applying the model, $\mathbf{y}^* = \theta^* + \mathbf{e}^* = \mathbf{X}\hat{\beta} + \mathbf{v}^* + \mathbf{e}^*$.

(5) Fit the Spatial FH model (7.1) to the bootstrap data $\mathbf{y}^*$ using $\hat{\beta}$ and $\hat{\omega}$ as the true values of β and ω. The estimates of the "true" $\hat{\beta}$ and $\hat{\omega}$ are obtained based on bootstrap data $\mathbf{y}^*$, by calculating the estimator of $\hat{\beta}$ at the "true" $\hat{\omega}$,

$$\tilde{\beta}^*(\hat{\omega}) = \left[\mathbf{X}'\mathbf{V}^{-1}(\hat{\omega})\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{V}^{-1}(\hat{\omega})\mathbf{y}^*;$$

then, obtain the estimator $\hat{\omega}^*$ based on $\mathbf{y}^*$. Finally, the estimator of $\hat{\beta}$ calculated at $\hat{\omega}^*$ is $\tilde{\beta}^*(\hat{\omega}^*)$.

(6) Calculate the bootstrap Spatial BLUP from bootstrap data $\mathbf{y}^*$ using $\hat{\omega}$ as the true value of ω,

$$\tilde{\theta}_d^*(\hat{\omega}) = \mathbf{x}_d\tilde{\beta}^*(\hat{\omega}) + \mathbf{b}_d'\mathbf{G}(\hat{\omega})\mathbf{V}(\hat{\omega})^{-1}[\mathbf{y}^* - \mathbf{X}\tilde{\beta}^*(\hat{\omega})].$$

Then, compute the bootstrap Spatial EBLUP using $\hat{\omega}^*$ in place of the "true" $\hat{\omega}$ as,

$$\tilde{\theta}_d^*(\hat{\omega}^*) = \mathbf{x}_d\tilde{\beta}^*(\hat{\omega}^*) + \mathbf{b}_d'\mathbf{G}(\hat{\omega}^*)\mathbf{V}^{-1}(\hat{\omega}^*)[\mathbf{y}^* - \mathbf{X}\tilde{\beta}^*(\hat{\omega}^*)].$$

(7) Repeat steps (2)–(6) B times. In the b-th bootstrap replication, $\theta_d^{*(b)}$ is the quantity of interest for d-th area, $\hat{\omega}^{*(b)}$ the bootstrap estimate of ω, $\tilde{\theta}_d^{*(b)}(\hat{\omega})$ the bootstrap Spatial BLUP and $\tilde{\theta}_d^{*(b)}(\hat{\omega}^{*(b)})$ is the bootstrap Spatial EBLUP for d-th area.

(8) A parametric bootstrap estimator of $g_{3d}(\omega)$ is

$$g_{3d}^{PB}(\hat{\omega}) = B^{-1}\sum_{b=1}^{B}\left[\tilde{\theta}_d^{*(b)}(\hat{\omega}^{*(b)}) - \tilde{\theta}_d^{*(b)}(\hat{\omega})\right]^2,$$

and a naive parametric bootstrap estimator of the full MSE is is given by

$$mse^{naPB}[\tilde{\theta}_d(\hat{\omega})] = B^{-1}\sum_{b=1}^{B}\left[\tilde{\theta}_d^{*(b)}(\hat{\omega}^{*(b)}) - \theta_d^{*(b)}\right]^2. \quad (7.7)$$

We can also obtain an alternative estimator of the MSE by adding the analytical estimates $g_{1d}(\hat{\omega})$ and $g_{2d}(\hat{\omega})$, the bootstrap estimate $g_{3d}^{PB}(\hat{\omega})$, and a bootstrap bias correction of $g_{1d}(\hat{\omega}) + g_{2d}(\hat{\omega})$ to obtain a MSE estimate similar to the one of Pfeffermann and Tiller (2005). The alternative final estimator is

$$mse^{bcPB}[\tilde{\theta}_d(\hat{\omega})] = 2\left[g_{1d}(\hat{\omega}) + g_{2d}(\hat{\omega})\right] - B^{-1}\sum_{b=1}^{B}\left[g_{1d}(\hat{\omega}^{*(b)}) + g_{2d}(\hat{\omega}^{*(b)})\right] + g_{3d}^{PB}(\hat{\omega}). \quad (7.8)$$

7.4 Nonparametric bootstrap

Instead of a parametric bootstrap, we can use a nonparametric bootstrap for the estimation of the MSE. In this method, the random effects $\{u_1^*, \ldots, u_D^*\}$ and the random errors $\{e_1^*, \ldots, e_D^*\}$ are obtained by re-sampling respectively from the empirical distribution of the previously standardized predicted random effects $\{\hat{u}_1, \ldots, \hat{u}_D\}$ and residuals $\{\hat{r}_1, \ldots, \hat{r}_D\}$ respectively, where $r_d = y_d - \tilde{\theta}_d(\hat{\omega})$, $d = 1, \ldots, D$. The nonparametric bootstrap is more robust to the non-normality of any of the random components of the model since it does not assume any distribution for them.

Under model (7.1)–(7.2), the BLUPs of $\mathbf{u}$ and $\mathbf{v}$ are respectively

$$\tilde{\mathbf{v}}(\omega) = \mathbf{G}(\omega)\mathbf{V}^{-1}(\omega)[\mathbf{y} - \mathbf{X}\tilde{\beta}(\omega)], \quad \tilde{\mathbf{u}}(\omega) = (\mathbf{I} - \rho\mathbf{W})\tilde{\mathbf{v}}(\omega),$$

with $\tilde{\mathbf{u}}(\omega)$ covariance matrix given by

$$\mathbf{V}_{\mathbf{u}}(\omega) = (\mathbf{I} - \rho\mathbf{W})\mathbf{G}(\omega)\mathbf{P}(\omega)\mathbf{G}(\omega)(\mathbf{I} - \rho\mathbf{W}'),$$

where

$$\mathbf{P}(\omega) = \mathbf{V}^{-1}(\omega) - \mathbf{V}^{-1}(\omega)\mathbf{X}\left[\mathbf{X}'\mathbf{V}^{-1}(\omega)\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{V}^{-1}.$$

Furthermore, the vector of residuals is

$$\tilde{\mathbf{r}}(\omega) = \mathbf{y} - \mathbf{X}\tilde{\beta}(\omega) - \tilde{\mathbf{v}}(\omega) = (y_1 - \tilde{\theta}_1(\omega), \ldots, y_D - \tilde{\theta}_D(\omega))',$$

with covariance matrix

$$\mathbf{V}_{\mathbf{r}}(\omega) = \Psi\mathbf{P}(\omega)\Psi.$$

We obtain the spectral decomposition of $\hat{\mathbf{V}}_{\mathbf{u}} = \mathbf{V}_{\mathbf{u}}(\hat{\omega})$ as

$$\hat{\mathbf{V}}_{\mathbf{u}} = \mathbf{Q}_{\mathbf{u}}\Delta_{\mathbf{u}}\mathbf{Q}'_{\mathbf{u}},$$

where $\Delta_{\mathbf{u}}$ is a diagonal matrix with the $D-p$ non-zero eigenvalues of $\hat{\mathbf{V}}_{\mathbf{u}}$ and $\mathbf{Q}_{\mathbf{u}}$ is the matrix with the corresponding eigenvectors in the columns. Keep in view that $\tilde{\mathbf{u}}(\omega)$ lies in a $D-p$ dimension space. Second, we square the matrix $\hat{\mathbf{V}}_{\mathbf{u}}^{-1/2} = \mathbf{Q}_{\mathbf{u}}\Delta_{\mathbf{u}}^{-1/2}\mathbf{Q}'_{\mathbf{u}}$ to obtain a generalized inverse of $\hat{\mathbf{V}}_{\mathbf{u}}$. The transformed $\hat{\mathbf{u}}$ is obtained as

$$\hat{\mathbf{u}}^S = \hat{\mathbf{V}}_{\mathbf{u}}^{-1/2}\hat{\mathbf{u}}.$$

The covariance matrix of $\hat{\mathbf{u}}^S$ is $Var(\hat{\mathbf{u}}^S) = \mathbf{Q}_{\mathbf{u}}\mathbf{Q}'_{\mathbf{u}}$, which is close to an identity matrix.

The nonparametric bootstrap procedure works by replacing steps (2) and (3) of the parametric bootstrap by the new steps (2') and (3') as:

(2') Calculate predictors of $\mathbf{v}$ and $\mathbf{u}$ using the estimates $\hat{\omega} = (\hat{\sigma}_u^2, \hat{\rho})'$ and $\hat{\beta} = \tilde{\beta}(\hat{\omega})$ obtained in step (1) in the following way:

$$\hat{\mathbf{v}} = \mathbf{G}(\hat{\omega})\mathbf{V}(\hat{\omega})^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}), \quad \hat{\mathbf{u}} = (\mathbf{I} - \hat{\rho}\mathbf{W})\hat{\mathbf{v}} = (\hat{u}_1, \ldots, \hat{u}_D)'.$$

Make $\hat{\mathbf{u}}^S = \hat{\mathbf{V}}_{\mathbf{u}}^{-1/2}\hat{\mathbf{u}} = (\hat{u}_1^S, \ldots, \hat{u}_D^S)'$, where $\hat{\mathbf{V}}_{\mathbf{u}}^{1/2}$ is the square root of the generalized inverse of $\hat{\mathbf{V}}_{\mathbf{u}}$ obtained by the spectral decomposition. Then, re-scale the elements $\hat{u}_d^S$ to obtain elements with sample mean exactly equal to zero and sample variance $\hat{\sigma}_u^2$. The transformation is

$$\hat{u}_d^{SS} = \frac{\hat{\sigma}_u(\hat{u}_d^S - D^{-1}\sum_{\ell=1}^D \hat{u}_\ell^S)}{\sqrt{D^{-1}\sum_{k=1}^D(\hat{u}_k^S - D^{-1}\sum_{\ell=1}^D \hat{u}_\ell^S)^2}}, \quad d = 1, \ldots, D.$$

Build the vector $\mathbf{u}^* = (u_1^*, \ldots, u_D^*)'$. Its elements are obtained by extracting a simple random sample with replacement of size D from the set $\{\hat{u}_1^{SS}, \ldots, \hat{u}_D^{SS}\}$. Proceed by obtaining $\mathbf{v}^* = (\mathbf{I} - \hat{\rho}\mathbf{W})^{-1}\mathbf{u}^*$ and calculating the bootstrap quantity of interest $\theta^* = \mathbf{X}\hat{\beta} + \mathbf{v}^* = (\theta_1^*, \ldots, \theta_D^*)'$

(3') Compute the vector of residuals $\hat{\mathbf{r}} = \mathbf{y} - \mathbf{X}\hat{\beta} - \hat{\mathbf{v}} = (\hat{r}_1, \ldots, \hat{r}_D)'$ and proceed with its standardization $\hat{\mathbf{r}}^S = \hat{\mathbf{V}}_{\mathbf{r}}^{-1/2}\hat{\mathbf{r}} = (\hat{r}_1^S, \ldots, \hat{r}_D^S)'$, where $\hat{\mathbf{V}}_{\mathbf{r}} = \Psi\mathbf{P}(\hat{\omega})\Psi$ is the estimated covariance matrix and $\hat{\mathbf{V}}_{\mathbf{r}}^{-1/2}$ is a root square of the generalized inverse derived from the spectral decomposition of $\hat{\mathbf{V}}_{\mathbf{r}}$. Once more, re-standardize these values in the following way

$$\hat{r}_d^{SS} = \frac{\hat{r}_d^S - D^{-1}\sum_{\ell=1}^D \hat{r}_\ell^S}{\sqrt{D^{-1}\sum_{k=1}^D (\hat{r}_k^S - D^{-1}\sum_{\ell=1}^D \hat{r}_\ell^S)^2}}, \quad d = 1, \ldots, D.$$

Finally, build $\mathbf{r}^* = (r_1^*, \ldots, r_D^*)'$ by extracting a simple random sample with replacement of size D from the set $\{\hat{r}_1^{SS}, \ldots, \hat{r}_D^{SS}\}$ and let $\mathbf{e}^* = (e_1^*, \ldots, e_D^*)'$, where $e_d^* = \psi_d^{1/2} r_d^*$, $d = 1, \ldots, D$.

This procedure leads to naive and bias-corrected nonparametric bootstrap estimators analogous to (7.7) and (7.8), which are denoted as $mse^{naNPB}[\tilde{\theta}_d(\hat{\omega})]$ and $mse^{bcNPB}[\tilde{\theta}_d(\hat{\omega})]$ respectively.

When the normality assumption is suspected to be violated either for the random effects or for the errors but not for both, it is possible to combine step (2') with (3), or step (2) with (3') of the two bootstrap procedures. This comes out in a semiparametric bootstrap that avoids the normality assumption on the desired component of the model.

7.5 Application with Spanish EU-SILC data

We applied the Spatial Fay-Herriot model to find small area estimators of FGT poverty measures in Spanish provinces. For this, we used the SLCS data described in Section 1.1 for year 2006 to calculate direct estimators $F_{\alpha d}^w$ of FGT poverty measures with $\alpha = 0$ (poverty incidence) and $\alpha = 1$ (poverty gap), together with their corresponding estimated sampling variances. Let E_{dj} be the equivalized income of j-th individual in the Spanish SILC for d-th Spanish province and w_{dj} the SLCS sampling weight for that same individual, $j = 1, \ldots, n_d$, $d = 1, \ldots, D$. Let z be the poverty line obtained for Spain in year 2006. Consider as individual observations, the values

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

then, obtain direct estimators of FGT poverty measures, together with their estimated sampling variances, using standard sampling theory formulas applied to $F_{\alpha dj}$, that is,

$$\hat{F}_{\alpha d}^w = \frac{1}{N_d}\sum_{j \in s_d} w_{dj}F_{\alpha dj} \quad \text{and} \quad \text{var}(\hat{F}_{\alpha d}^w) = \frac{1}{N_d^2} w_{dj}(w_{dj} - 1)F_{\alpha dj}^2, \quad \alpha = 0, 1.$$

Here, the variance has been obtained under the assumption that $w_{dj} = 1/\pi_{dj}$, where π_{dj} is the inclusion probability of j-th unit from d-th province in the SLCS sample, and that the joint probabilities of inclusion of j-th and i-th individuals in d-th province are $\pi_{dj,di} = \pi_{dj}\pi_{di}$.

Then, a different Fay-Herriot model was considered for estimating poverty incidence ($\alpha = 0$) and poverty gap ($\alpha = 1$). In the model for the poverty incidence, we consider $y_d = \hat{F}_{0d}^w$ and $\psi_d = \text{var}(\hat{F}_{0d}^w)$, $d = 1, \ldots, D$, while in the model for the poverty gap, we consider $y_d = \hat{F}_{1d}^w$ and $\psi_d = \text{var}(\hat{F}_{1d}^w)$, $d = 1, \ldots, D$. In order to apply a Fay-Herriot model, we need covariates at the province level. For both models, we have considered as covariates, the province proportions of people in the following categories:

- Five categories of variable age;
- Gender (Male/Female);
- Spanish nationality (yes/no);
- Four categories of Education level;
- Three categories of Employment (unemployed, employed, inactive).

The true values of these proportions in each province are not available, but they were estimated from the Spanish EAPS survey of year 2006, which has a total sample size of over 155,000 observations, see Section 1.2. These proportions have small variability due to the large sample size of this survey and we consider them as the covariates in our Spatial Fay-Herriot model. However, after fitting the model for both poverty incidence and poverty gap, the only proportions that were significant were the following:

- age4: Age in between 50-64;
- nat1: Having Spanish nationality;
- educ1: Education level of until primary school;
- situ2: Unemployed.

Concerning the proximity matrix, we constructed it by assigning value 1 to each pair of provinces that share some boundary and that belong to the same Autonomous Community and 0 otherwise.

Results of the model fit for the poverty incidence are given in Table 7.1. The signs of the coefficients indicate that the poverty incidence decreases for people in group age 50-64 and increases for Spanish people that have as much primary education and that are unemployed. As for the nationality, the positive coefficient can be explained from the fact that the foreigners surveyed by the SLCS are only those with better life conditions than average Spaniards. Note that the estimated spatial correlation is $\rho = 0.449$, which is quite significant. This indicates that direct estimators for neighbor provinces within the same region are correlated.

Variable	Coef	Std. Error	t value	p value
Const	6.8E-05	3.7E-04	1.8E-01	8.5E-01
age4	-2.632	0.450	-5.853	4.8E-09
nat1	0.357	0.107	3.322	8.9E-04
educ1	0.688	0.119	5.793	6.9E-09
situ2	2.555	0.613	4.167	3.1E-05
σ_u^2	ρ	loglike	AIC	BIC
0.000985	0.449	87.813	-161.627	-147.97

Table 7.1: Results of the Spatial Fay-Herriot model fitting for poverty incidence.

Results of the model fit for the poverty gap are given in Table 7.2. In this table the signs of the coefficients are the same as in the model for the poverty incidence, which means that the effect of the different categories on the poverty gap is similar to the effect on the poverty incidence. However, the

Variable	Coef	Std. Error	t value	p value
Const	-9.753E-06	2.254E-04	-0.043	0.965
age4	-1.035	0.267	-3.875	1.068E-04
nat1	0.156	0.063	2.475	0.013
educ1	0.225	0.067	3.354	0.001
situ2	0.611	0.360	1.698	0.090
σ_u^2	ρ	loglike	AIC	BIC
0.000463	0.348	119.453	-224.906	-211.248

Table 7.2: Results of the Spatial Fay-Herriot model fitting for poverty gap.

variables nat1, educ1 and situ2 are less significant than before. Observe also that the spatial correlation is smaller than in the model for the poverty incidence.

The standard Fay-Herriot model without spatial correlation was also fitted to both poverty incidence and gap. Table 6.3 lists direct, EBLUP obtained from the standard Fay-Herriot and Spatial EBLUP small area estimates of poverty incidece, together with their estimated MSEs and percent coefficients of variation. MSEs of EBLUP estimators have been obtained using the analytical formula of Prasad and Rao (1990). To estimate the MSE of the Spatial EBLUP, the parametric bootstrap described in Section 7.3 has been applied, using the bias-corrected parametric bootstrap MSE estimator (7.8). Comparing the CVs of direct and EBLUP estimators, see that there is a considerable reduction for some of the provinces, and this reduction is larger for provinces with smaller sample sizes such as Ávila, Castellón, Soria and Teruel. Concerning the Spatial EBLUPs, observe that their CVs are smaller than those of the BLUPs for all except four provinces, and in those four provinces the CVs are very similar. This means that taking into account the spatial correlation between neighbor provinces in the same region improves the final small area estimators of poverty incidence for most provinces.

Concerning the poverty gap, see in Table 6.4 that again the use of a Fay-Herriot model improves considerably the CVs of the estimators as compared with direct estimators. However, comparing the CVs of the Spatial EBLUP with those of the EBLUP, even if the Spatial EBLUP still improves the EBLUP for most provinces, it seems that taking into account the spatial correlation in this case does not help much to increase the efficiency of EBLUP estimators.

Figure 7.1 plots the values of the SEBLUP estimates of the poverty incidences (left) and poverty gaps (right) against the corresponding direct estimators. Observe in the left plot that the points approximately follow the line $y = x$, which means that SEBLUP poverty incidence estimators are approximately design unbiased, except for four provinces, namely Ávila, Cuenca, Segovia and Soria. These provinces are among the 7 provinces with largest direct estimators of poverty incidence and at the same time with smallest sample sizes. The large direct estimators for these four provinces, with more than 30% of people under the poverty line, are not reliable due to the small sample sizes, and then the SEBLUP estimators are smoothing them. For the poverty gaps, the points follow more closely the line $y = x$ except for Granada, which has an atypically large value of the direct estimator of the poverty gap. Since the sample size in Granada is not so small, this might be an outlier that is not well fitted by the model.

Cartograms with estimated percent poverty incidences and poverty gaps for Spanish provinces, obtained using the Spatial EBLUP, are given in Figure 7.2. Observe that the provinces with largest values of estimated poverty incidences are not necessarily those with largest poverty gaps, that is, the provinces

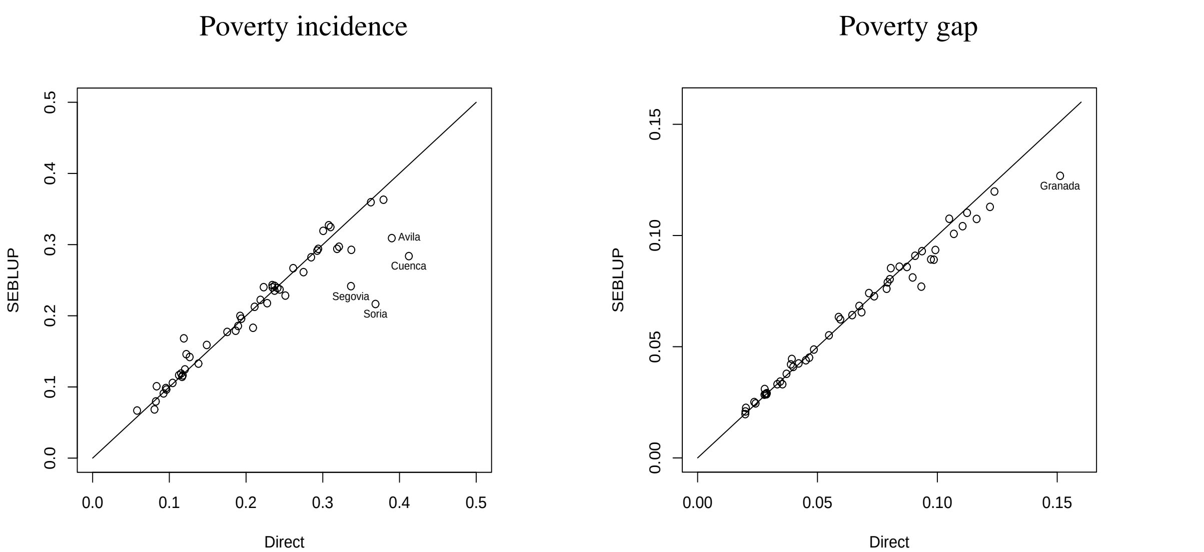


Figure 7.1: SEBLUP versus direct estimates of poverty incidences (left) and poverty gaps (right) in Spanish provinces.

with largest proportion of people under the poverty line (Badajoz, Córdoba, Jaen and Zamora) are not necessarily the ones that have the poorest people among the poor (Granada and Soria).

Table 6.3: Results on poverty incidence for Spanish SILC data: province, sample size, direct, EBLUP and SEBLUP estimates (%), estimated variance of direct estimator, estimated MSEs of EBLUP and SEBLUP ($\times 10^4$), CVs of direct, EBLUP and SEBLUP (%).

Province	n_d	$\hat{F}_{0d}^{w}$	$\hat{F}_{0d}^{EBLUP}$	$\hat{F}_{0d}^{SEBLUP}$	var$(\hat{F}_{0d}^{w})$	$mse(\hat{F}_{0d}^{EB})$	$mse(\hat{F}_{0d}^{SEBLUP})$	cv$(\hat{F}_{0d}^{w})$	cv$(\hat{F}_{0d}^{EBLUP})$	cv$(\hat{F}_{0d}^{SEBLUP})$
Álava	191	8.06	7.15	6.84	6.00	4.67	4.22	30.40	30.24	30.03
Albacete	346	26.15	27.11	26.69	8.19	5.63	5.19	10.94	8.75	8.53
Alicante	1078	17.54	17.65	17.73	1.86	1.68	1.67	7.77	7.35	7.29
Almería	397	33.73	29.11	29.26	9.24	6.02	5.87	9.01	8.43	8.28
Ávila	116	39.00	29.96	30.92	41.22	11.84	10.91	16.46	11.48	10.68
Badajoz	987	37.90	36.04	36.31	4.75	3.75	3.65	5.75	5.38	5.26
Baleares	1269	10.42	10.56	10.57	1.05	0.99	0.98	9.82	9.41	9.36
Barcelona	2841	9.62	9.70	9.63	0.34	0.33	0.33	6.03	5.92	5.95
Burgos	335	12.65	13.45	14.20	4.60	3.61	3.55	16.96	14.12	13.27
Cáceres	563	29.27	27.82	29.15	6.51	4.68	4.39	8.72	7.78	7.19
Cádiz	795	28.48	28.87	28.22	4.14	3.32	3.15	7.14	6.32	6.29
Castellon	236	12.22	14.05	14.61	7.16	4.89	4.77	21.90	15.73	14.96
Ciudad Real	499	29.42	29.99	29.40	6.72	4.77	4.38	8.81	7.29	7.12
Cordoba	447	30.78	32.85	32.72	7.83	5.44	4.92	9.09	7.10	6.78
Coruña La	990	22.75	21.88	21.77	2.45	2.12	2.09	6.88	6.66	6.64
Cuenca	183	41.21	33.14	28.39	31.34	11.21	8.06	13.58	10.10	10.00
Gerona	283	5.79	6.62	6.68	2.13	1.92	1.88	25.19	20.92	20.54
Granada	417	32.11	29.43	29.71	9.10	5.68	4.93	9.40	8.10	7.47
Guadalajara	178	11.89	16.30	16.82	6.50	4.64	4.57	21.46	13.21	12.71
Guipúzcoa	570	8.24	7.99	7.95	2.03	1.82	1.75	17.28	16.89	16.64

Huelva	244	22.29	24.51	24.02	8.62	5.60	5.07	13.17	9.65	9.37
Huesca	230	8.34	9.92	10.09	4.88	3.73	3.57	26.47	19.48	18.71
Jaén	463	31.01	32.31	32.47	7.00	5.01	4.83	8.53	6.93	6.77
Leon	437	19.22	19.67	19.99	5.55	4.08	3.92	12.26	10.27	9.91
Lérida	260	20.91	20.09	18.32	10.39	6.35	5.86	15.41	12.55	13.21
Rioja La	1019	18.97	18.66	18.56	2.71	2.32	2.26	8.68	8.15	8.10
Lugo	346	21.10	20.47	21.26	7.97	5.53	5.16	13.38	11.49	10.68
Madrid	1889	11.74	11.66	11.63	1.03	0.96	0.96	8.62	8.43	8.42
Málaga	758	24.09	24.08	23.90	3.61	2.93	2.80	7.89	7.11	7.00
Murcia	1770	23.72	23.61	23.53	1.85	1.66	1.64	5.74	5.45	5.44
Navarra	1128	9.24	9.11	9.09	1.08	1.02	1.01	11.23	11.06	11.07
Orense	258	23.72	23.73	24.21	15.82	8.00	7.08	16.77	11.92	10.99
Oviedo	1606	11.54	11.80	11.88	1.02	0.96	0.96	8.75	8.32	8.24
Palencia	143	25.13	23.16	22.84	18.48	8.47	7.47	17.11	12.57	11.96
Palmas Las	943	23.41	23.76	24.31	4.20	3.42	3.23	8.76	7.78	7.39
Pontevedra	896	19.40	19.19	19.59	2.97	2.50	2.43	8.89	8.24	7.96
Salamanca	328	31.87	29.08	29.40	12.23	6.74	6.32	10.98	8.93	8.55
Tenerife	762	27.49	26.59	26.13	4.76	3.69	3.45	7.94	7.22	7.10
Santander	867	11.24	11.56	11.66	1.89	1.70	1.67	12.24	11.27	11.09
Segovia	115	33.67	23.62	24.15	36.42	10.70	9.64	17.92	13.85	12.85
Sevilla	963	21.87	22.48	22.24	2.55	2.24	2.16	7.31	6.66	6.60
Soria	41	36.88	21.25	21.66	147.78	13.33	11.87	32.97	17.18	15.91
Tarragona	268	14.87	15.97	15.91	5.51	4.09	3.96	15.79	12.67	12.51
Teruel	144	12.03	12.45	12.46	9.04	6.03	5.38	25.00	19.72	18.61
Toledo	550	23.44	24.07	23.99	5.01	3.86	3.75	9.55	8.16	8.07
Valencia	1428	13.79	13.59	13.29	1.49	1.37	1.34	8.84	8.61	8.73
Valladolid	598	18.65	17.69	17.90	3.89	3.15	3.09	10.57	10.04	9.83
Vizcaya	1047	9.54	9.63	9.84	1.08	1.02	1.01	10.89	10.47	10.20
Zamora	209	30.06	30.96	31.94	14.90	9.91	9.66	12.84	10.17	9.73
Zaragoza	1129	11.67	11.72	11.42	1.38	1.28	1.24	10.05	9.64	9.76
Ceuta	470	36.27	35.64	35.96	9.20	6.36	5.94	8.36	7.07	6.78
Melilla	359	24.39	23.38	23.69	8.10	5.65	5.26	11.67	10.17	9.68

Table 6.4: Results on poverty gap for Spanish SILC data: province, sample size, direct, EBLUP and SEBLUP estimates (%), estimated variance of direct estimator, estimated MSEs of EBLUP and SEBLUP ($\times 10^4$), CVs of direct, EBLUP and SEBLUP (%).

Province	n_d	$\hat{F}^{w}_{1d}$	$\hat{F}^{EBLUP}_{1d}$	$\hat{F}^{SEBLUP}_{1d}$	$\mathrm{var}(\hat{F}^{w}_{1d})$	$mse(\hat{F}^{EB}_{1d})$	$mse(\hat{F}^{SEBLUP}_{1d})$	$\mathrm{cv}(\hat{F}^{w}_{1d})$	$\mathrm{cv}(\hat{F}^{EBLUP}_{1d})$	$\mathrm{cv}(\hat{F}^{SEBLUP}_{1d})$
Álava	191	1.99	1.95	1.97	0.40	0.38	0.38	31.93	31.80	31.33
Albacete	346	10.69	10.34	10.07	1.87	1.46	1.41	12.81	11.67	11.79
Alicante	1078	5.48	5.48	5.51	0.32	0.30	0.31	10.32	10.07	10.06
Almería	397	11.05	10.03	10.42	1.56	1.26	1.27	11.31	11.20	10.83
Ávila	116	11.64	10.20	10.74	4.69	2.68	2.72	18.61	16.07	15.36
Badajoz	987	12.38	12.06	11.98	0.79	0.70	0.71	7.16	6.95	7.03
Baleares	1269	2.88	2.91	2.91	0.18	0.17	0.17	14.67	14.35	14.31
Barcelona	2841	3.32	3.32	3.31	0.07	0.06	0.06	7.69	7.64	7.66
Burgos	335	3.89	4.03	4.20	1.04	0.89	0.89	26.17	23.33	22.39
Cáceres	563	8.42	8.39	8.60	0.77	0.69	0.69	10.43	9.87	9.63
Cádiz	795	9.07	9.15	9.09	0.70	0.63	0.63	9.24	8.69	8.73

Castellón	236	3.93	4.29	4.45	1.34	1.09	1.12	29.50	24.36	23.79
Ciudad Real	499	7.14	7.42	7.41	0.52	0.48	0.48	10.11	9.35	9.33
Cordoba	447	8.06	8.52	8.53	0.75	0.68	0.67	10.76	9.64	9.61
Coruña La	990	7.89	7.67	7.61	0.45	0.42	0.42	8.55	8.47	8.54
Cuenca	183	9.73	9.67	8.93	1.94	1.51	1.43	14.32	12.69	13.39
Gerona	283	2.00	2.06	2.09	0.33	0.32	0.32	28.91	27.44	27.02
Granada	417	15.12	12.59	12.68	2.95	1.94	1.89	11.36	11.05	10.85
Guadalajara	178	2.01	2.25	2.25	0.20	0.19	0.19	21.98	19.36	19.32
Guipúzcoa	570	3.54	3.34	3.31	0.61	0.56	0.55	22.05	22.35	22.41
Huelva	244	9.85	9.56	8.91	3.21	2.08	2.03	18.20	15.08	16.00
Huesca	230	2.79	3.01	3.10	0.79	0.70	0.71	31.87	27.78	27.10
Jaén	463	10.49	10.73	10.76	1.19	1.01	1.01	10.39	9.34	9.32
Leon	437	7.36	7.20	7.27	1.28	1.05	1.04	15.39	14.23	14.06
Lérida	260	9.33	8.13	7.70	2.75	1.88	1.90	17.76	16.89	17.90
Rioja La	1019	4.21	4.25	4.25	0.17	0.17	0.17	9.91	9.69	9.68
Lugo	346	6.45	6.26	6.42	2.40	1.75	1.75	24.04	21.13	20.59
Madrid	1889	3.45	3.44	3.43	0.20	0.19	0.19	12.88	12.70	12.73
Málaga	758	9.92	9.51	9.35	1.12	0.94	0.93	10.69	10.20	10.31
Murcia	1770	7.92	7.91	7.89	0.25	0.24	0.24	6.29	6.17	6.18
Navarra	1128	2.87	2.86	2.86	0.17	0.16	0.16	14.29	14.15	14.18
Orense	258	5.96	6.30	6.23	1.81	1.40	1.38	22.58	18.76	18.85
Oviedo	1606	2.83	2.87	2.87	0.12	0.12	0.12	12.38	12.11	12.10
Palencia	143	5.89	6.21	6.33	1.73	1.34	1.34	22.32	18.67	18.26
Palmas Las	943	8.02	8.02	8.04	0.94	0.83	0.84	12.12	11.34	11.39
Pontevedra	896	3.71	3.77	3.78	0.15	0.15	0.15	10.48	10.18	10.16
Salamanca	328	9.36	9.18	9.30	1.33	1.08	1.07	12.31	11.32	11.15
Tenerife	762	8.73	8.61	8.58	0.70	0.63	0.63	9.58	9.21	9.24
Santander	867	2.79	2.84	2.84	0.16	0.16	0.16	14.57	14.10	14.09
Segovia	115	8.97	7.88	8.11	3.47	2.18	2.15	20.77	18.73	18.09
Sevilla	963	3.99	4.11	4.09	0.13	0.13	0.13	9.10	8.76	8.77
Soria	41	19.10	7.37	7.48	70.78	5.10	5.00	44.04	30.64	29.90
Tarragona	268	2.36	2.50	2.51	0.19	0.18	0.18	18.44	17.19	17.08
Teruel	144	4.85	4.74	4.87	2.46	1.80	1.77	32.33	28.29	27.33
Toledo	550	6.75	6.97	6.84	0.63	0.57	0.57	11.75	10.85	11.05
Valencia	1428	4.66	4.60	4.51	0.35	0.33	0.34	12.70	12.54	12.87
Valladolid	598	6.84	6.57	6.55	0.72	0.64	0.65	12.41	12.22	12.29
Vizcaya	1047	2.42	2.44	2.45	0.12	0.12	0.12	14.22	13.96	13.89
Zamora	209	11.24	10.61	11.02	3.14	2.39	2.42	15.77	14.58	14.12
Zaragoza	1129	4.51	4.48	4.39	0.44	0.41	0.42	14.73	14.35	14.68
Ceuta	470	17.92	15.72	16.17	3.15	2.21	2.22	9.90	9.46	9.22
Melilla	359	12.20	10.43	11.28	3.42	2.27	2.22	15.16	14.45	13.22

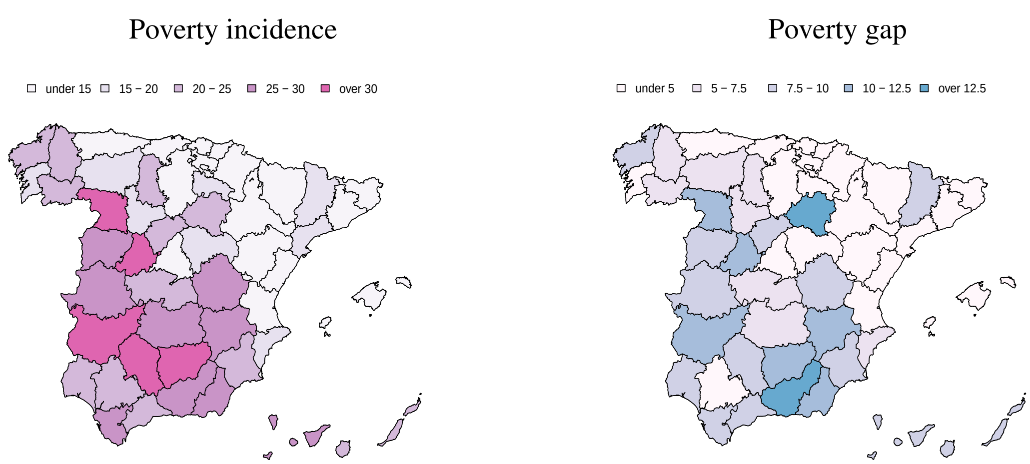


Figure 7.2: Cartograms of estimated percent poverty incidences (left) and poverty gaps (right) in Spanish provinces.

Chapter 8

Spatio-temporal Fay-Herriot models

8.1 Spatial Fay-Herriot model with uncorrelated time effects

8.1.1 The model

Let y_{dt} be a direct estimator of the target population parameter and $\mathbf{x}_{dt}$ a vector containing the aggregated values of p auxiliary variables for time instant t and domain d. We assume that y_{dt} follows the model (MODEL A)

$$y_{dt} = \mathbf{x}_{dt}\beta + u_{1d} + u_{2dt} + e_{dt}, \quad d = 1,\dots,D, \quad t = 1,\dots,T, \tag{8.1}$$

where $\{u_{1d}\}$, $\{u_{2dt}\}$ and $\{e_{dt}\}$ are independent, with $\{u_{1d}\}_{d=1}^{D}$ following a SAR(1) process with variance σ_1^2, spatial autocorrelation parameter ρ_1 and proximity matrix $\mathbf{W}$, $\{u_{2dt}\}$ are i.i.d. from $N(0,\sigma_2^2)$ and e_{dt} are independent from $N(0,\sigma_{dt}^2)$. Here, σ_1^2, ρ_1, σ_2^2 and ρ_2 are unknown parameters whereas the error variances σ_{dt}^2 are assumed to be known. Stacking the elements involved in model (8.1) in columns as

$$\begin{aligned}
&\mathbf{y} = \underset{1\le d\le D}{\text{col}}(\underset{1\le t\le T}{\text{col}}(y_{dt})), \quad \mathbf{e} = \underset{1\le d\le D}{\text{col}}(\underset{1\le t\le T}{\text{col}}(e_{dt})),\\
&\mathbf{u}_1 = \underset{1\le d\le D}{\text{col}}(u_{1d}), \quad \mathbf{u}_2 = \underset{1\le d\le D}{\text{col}}(\mathbf{u}_{2d}), \quad \mathbf{u}_{2d} = \underset{1\le t\le T}{\text{col}}(u_{2dt}),\\
&\mathbf{X} = \underset{1\le d\le D}{\text{col}}(\underset{1\le t\le T}{\text{col}}(\mathbf{x}_{dt})), \quad \mathbf{x}_{dt} = \underset{1\le j\le p}{\text{col}'}(x_{dtj}), \quad \beta = \underset{1\le j\le p}{\text{col}}(\beta_j),\\
&\mathbf{Z}_1 = \underset{1\le d\le D}{\text{diag}}(\mathbf{1}_T), \quad \mathbf{Z}_2 = \mathbf{I}_{M\times M}, \quad M = DT,
\end{aligned}$$

the model in vector-matrix notation is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{e}. \tag{8.2}$$

Then, it holds that $\mathbf{u}_1 \sim N(\mathbf{0},\mathbf{V}_{u_1})$, $\mathbf{u}_2 \sim N(\mathbf{0},\mathbf{V}_{u_2})$ and $\mathbf{e} \sim N(\mathbf{0},\mathbf{V}_e)$, with covariance matrices given by

$$\begin{aligned}
\mathbf{V}_{u_1} &= \sigma_1^2\Omega_1(\rho_1), \quad \Omega_1(\rho_1) = \left[(\mathbf{I}_D - \rho_1\mathbf{W})'(\mathbf{I}_D - \rho_1\mathbf{W})\right]^{-1},\\
\mathbf{V}_{u_2} &= \sigma_2^2\mathbf{I}_{DT},\\
\mathbf{V}_e &= \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_{ed}), \quad \mathbf{V}_{ed} = \underset{1\le t\le T}{\text{diag}}(\sigma_{dt}^2),
\end{aligned}$$

We assume that the rows of the proximity matrix $\mathbf{W}$ are stochastic vectors, i.e. with components summing up to one. The covariance matrix of $\mathbf{y}$ is

$$\text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}_1\mathbf{V}_{u_1}\mathbf{Z}_1' + \mathbf{Z}_2\mathbf{V}_{u_2}\mathbf{Z}_2' + \mathbf{V}_e = \mathbf{Z}_1\mathbf{V}_{u_1}\mathbf{Z}_1' + \underset{1\leq d\leq D}{\text{diag}}(\sigma_2^2\mathbf{I}_T + \mathbf{V}_{ed}).$$

The EBLUE of β and the EBLUP of $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2')'$ are given by

$$\widehat{\beta} = (\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{y} \qquad \text{and} \qquad \widehat{\mathbf{u}} = \widehat{\mathbf{V}}_u\mathbf{Z}'\widehat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\beta}),$$

where $\widehat{\mathbf{V}}_u = \text{diag}(\widehat{\mathbf{V}}_{u_1}, \widehat{\mathbf{V}}_{u_2})$ and $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$. The hats over the covariance matrices indicate that the model parameters σ_1^2, ρ_1 and σ_2^2 have been replaced by consistent estimators $\widehat{\sigma}_1^2$, $\widehat{\rho}_1$ and $\widehat{\sigma}_2^2$. Finally, the empirical best predictor of μ_{dt} under the spatial model with uncorrelated time effects (SEBA) is given by

$$\widehat{\mu}_{dt} = \mathbf{x}_{dt}\widehat{\beta} + \widehat{u}_{1d} + \widehat{u}_{2dt}.$$

8.1.2 Parametric bootstrap MSE estimation

Below we describe the steps of a parametric bootstrap procedure to estimate the MSE of the SEBA:

1. Fit model (8.1) to the original data $\{(\mathbf{x}_{dt}, y_{dt}),\ t = 1,\ldots,T,\ d = 1,\ldots,D\}$ and obtain model parameter estimates $\widehat{\beta}$, $\widehat{\sigma}_1^2$, $\widehat{\rho}_1$ and $\widehat{\sigma}_2^2$.

2. Repeat $B = 1000$ times ($b = 1,\ldots,B$)

   2.1. Generate, $\widehat{u}_{1d}^{(b)}$, $d = 1,\ldots,D$ from a SAR(1) process using as true variance $\widehat{\sigma}_1^2$, as spatial autocorrelation $\widehat{\rho}_1$ and proximity matrix $\mathbf{W}$. Independently, generate $u_{2dt}^{(b)} \sim N(0, \widehat{\sigma}_2^2)$, $t = 1,\ldots,T$, $d = 1,\ldots,D$. Calculate bootstrap parameters $\mu_{dt}^{(b)} = \widehat{\beta}x_{dt} + u_{1d}^{(b)} + u_{2dt}^{(b)}$, $t = 1,\ldots,T$, $d = 1,\ldots,D$.

   2.2. Generate $e_{dt}^{(b)} \sim N(0, \sigma_{dt}^2)$ and calculate $y_{dt}^{(b)} = \mu_{dt}^{(b)} + e_{dt}^{(b)}$, $t = 1,\ldots,T$, $d = 1,\ldots,D$.

   2.3. Fit the model to the bootstrap data $\{(\mathbf{x}_{dt}, y_{dt}^{(b)}),\ t = 1,\ldots,T,\ d = 1,\ldots,D\}$ and obtain the SEBA, $\widehat{\mu}_{dt}^{(b)} = \mathbf{x}_{dt}\widehat{\beta}^{(b)} + \widehat{u}_{1d}^{(b)} + \widehat{u}_{2dt}^{(b)}$.

4. Calculate the basic parametric bootstrap MSE estimator

$$mse_{dt}(\widehat{\mu}_{dt}) = \frac{1}{B}\sum_{b=1}^{B}(\widehat{\mu}_{dt}^{(b)} - \mu_{dt}^{(b)})^2.$$

8.2 Spatial Fay-Herriot model with correlated time effects

8.2.1 The model

Let y_{dt} be a direct estimator of the characteristic of interest and $\mathbf{x}_{dt}$ a vector containing the aggregated values of p of auxiliary variables for time instant t and domain d. Now we consider the following model for y_{dt} (MODEL B)

$$y_{dt} = \mathbf{x}_{dt}\beta + u_{1d} + u_{2dt} + e_{dt}, \quad d = 1,\ldots,D, \quad t = 1,\ldots,T. \tag{8.3}$$

Here, $\{u_{1d}\}$, $\{u_{2dt}\}$ and $\{e_{dt}\}$ are independent where $\{u_{1d}\}_{d=1}^{D}$ follow a SAR(1) process with variance σ_1^2, spatial autocorrelation parameter ρ_1 and row-standardized proximity matrix $\mathbf{W}$, $\{u_{2dt}\}_{t=1}^{T}$ for each d are i.i.d. following an AR(1) process with autocorrelation parameter ρ_2, and the errors e_{dt} are independent from $N(0,\sigma_{dt}^2)$, where the variances σ_{dt}^2 are assumed to be known. Again, stacking the elements of the model in columns as in

$$
\begin{aligned}
&\mathbf{y} = \underset{1\le d\le D}{\text{col}}(\underset{1\le t\le T}{\text{col}}(y_{dt})), \quad \mathbf{e} = \underset{1\le d\le D}{\text{col}}(\underset{1\le t\le T}{\text{col}}(e_{dt})),\\
&\mathbf{u}_1 = \underset{1\le d\le D}{\text{col}}(u_{1d}), \quad \mathbf{u}_2 = \underset{1\le d\le D}{\text{col}}(\mathbf{u}_{2d}), \quad \mathbf{u}_{2d} = \underset{1\le t\le T}{\text{col}}(u_{2dt}),\\
&\mathbf{X} = \underset{1\le d\le D}{\text{col}}(\underset{1\le t\le T}{\text{col}}(\mathbf{x}_{dt})), \quad \mathbf{x}_{dt} = \underset{1\le j\le p}{\text{col}'}(x_{dtj}), \quad \beta = \underset{1\le j\le p}{\text{col}}(\beta_j),\\
&\mathbf{Z}_1 = \underset{1\le d\le D}{\text{diag}}(\mathbf{1}_T), \quad \mathbf{Z}_2 = \mathbf{I}_{M\times M}, \quad M = DT,
\end{aligned}
$$

the model in vector-matrix notation is

$$
\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{e}, \tag{8.4}
$$

Then, it holds that $\mathbf{u}_1 \sim N(\mathbf{0}, \mathbf{V}_{u_1})$, $\mathbf{u}_2 \sim N(\mathbf{0}, \mathbf{V}_{u_2})$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{V}_e)$, where the covariance matrices are given by

$$
\begin{aligned}
\mathbf{V}_{u_1} &= \sigma_1^2\Omega_1(\rho_1), \quad \Omega_1(\rho_1) = \left[(\mathbf{I}_D - \rho_1\mathbf{W})'(\mathbf{I}_D - \rho_1\mathbf{W})\right]^{-1} \triangleq \mathbf{C}^{-1}(\rho_1),\\
\mathbf{V}_{u_2} &= \sigma_2^2\Omega_2(\rho_2), \quad \Omega_2(\rho_2) = \underset{1\le d\le D}{\text{diag}}(\Omega_{2d}(\rho_2)),\\
\mathbf{V}_e &= \underset{1\le d\le D}{\text{diag}}(\mathbf{V}_{ed}), \quad \mathbf{V}_{ed} = \underset{1\le t\le T}{\text{diag}}(\sigma_{dt}^2),\\
\Omega_{2d} &= \Omega_{2d}(\rho_2) = \frac{1}{1-\rho_2^2}\begin{pmatrix}
1 & \rho_2 & \dots & \rho_2^{T-2} & \rho_2^{T-1}\\
\rho_2 & 1 & \ddots & & \rho_2^{T-2}\\
\vdots & \ddots & \ddots & \ddots & \vdots\\
\rho_2^{T-2} & & \ddots & 1 & \rho_2\\
\rho_2^{T-1} & \rho_2^{T-2} & \dots & \rho_2 & 1
\end{pmatrix}_{T\times T}.
\end{aligned}
$$

Then, the variance of $\mathbf{y}$ is

$$
\text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}_1\mathbf{V}_{u_1}\mathbf{Z}_1' + \mathbf{Z}_2\mathbf{V}_{u_2}\mathbf{Z}_2' + \mathbf{V}_e = \mathbf{Z}_1\mathbf{V}_{u_1}\mathbf{Z}_1' + \underset{1\le d\le D}{\text{diag}}(\sigma_2^2\Omega_{2d}(\rho_2) + \mathbf{V}_{ed}).
$$

The EBLUE of β and the EBLUP of $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2')'$ are given by

$$
\widehat{\beta} = (\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{y} \qquad \text{and} \qquad \widehat{\mathbf{u}} = \widehat{\mathbf{V}}_u\mathbf{Z}'\widehat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\beta}),
$$

where $\widehat{\mathbf{V}}_u = \text{diag}(\widehat{\mathbf{V}}_{u_1}, \widehat{\mathbf{V}}_{u_2})$ and $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$, where the hats over the covariance matrices indicate that the model parameters σ_1^2, ρ_1, σ_2^2 and ρ_2, have been replaced by consistent estimators. Finally, the empirical best predictor of μ_{dt} under the spatial model with AR(1) time effects (SEBB) is given by

$$
\widehat{\mu}_{dt} = \mathbf{x}_{dt}\widehat{\beta} + \widehat{u}_{1d} + \widehat{u}_{2dt}.
$$

8.2.2 Parametric bootstrap estimation of the MSE

The parametric bootstrap procedure to estimate the MSE of the SEBB is obtained by replacing Step 2.1 in Section 8.1.2 by new Step 2.1B below:

2.1B. Generate, independently, $u_{1d}^{(b)}$, $d=1,\ldots,D$, from a SAR(1) process with variance $\widehat{\sigma}_1^2$ and spatial autocorrelation parameter $\widehat{\rho}_1$. Independently of $u_{1d}^{(b)}$ and independently for each d, generate $u_{2dt}^{(b)}$, $t=1,\ldots,T$, from an AR(1) process with variance $\widehat{\sigma}_2^2$ and autocorrelation parameter $\widehat{\rho}_2$. Calculate bootstrap parameters $\mu_{dt}^{(b)}=\widehat{\beta}x_{dt}+u_{1d}^{(b)}+u_{2dt}^{(b)}$, $t=1,\ldots,T$, $d=1,\ldots,D$.

8.3 Application with Spanish EU-SILC data

The two models, MODEL A and MODEL B, have been applied to estimate FGT poverty measures (specifically, poverty incidences and gaps) in Spanish provinces at year 2006, using SLCS data from years 2004-2006. Thus, here we have $T=3$ time instants and $D=52$ provinces, which are taken here as the provinces. Direct estimators of the FGT poverty measures for $\alpha=0,1$ are obtained similarly as described in Section 7.5, but for each instant $t=1,2,3$, that is,

$$\widehat{F}_{\alpha dt}^{w}=\frac{1}{N_d}\sum_{j\in s_d} w_{dj}F_{\alpha djt} \quad \text{and} \quad \text{var}(\widehat{F}_{\alpha dt}^{w})=\frac{1}{N_{dt}^2}w_{djt}(w_{djt}-1)F_{\alpha djt}^2, \quad \alpha=0,1,$$

where z_t is the poverty line at time t and

$$F_{\alpha djt}=\left(\frac{z_t-E_{djt}}{z}\right)^{\alpha} I(E_{djt}<z_t), \quad j=1,\ldots,n_d, \quad d=1,\ldots,D, \quad t=1,\ldots,T.$$

Thus, MODEL A given in (8.1) has been applied with $y_{dt}=\widehat{F}_{\alpha dt}^{w}$ and $\sigma_{dt}^2=\text{var}(\widehat{F}_{\alpha dt}^{w})$, separately for $\alpha=0$ (poverty incidence) and $\alpha=1$ (poverty gap). Similarly for MODEL B. The proximity matrix was the same as described in Section 7.5, which has value 1 for a pair of provinces that share a boundary and at the same time belong to the same Autonomous Community and 0 otherwise. As covariates, in principle the province proportions of people in the following categories at each year were considered:

- Five categories of variable age;
- Gender (Male/Female);
- Spanish nationality (yes/no);
- Four categories of Education level;
- Three categories of Employment (unemployed, employed, inactive).

These proportions were estimated from the Spanish EAPS survey of the corresponding year, see Section 1.2. These proportions have small variability due to the large sample size of this survey. However, in MODELS A and B for poverty incidence, the only proportions that were significant were the following:

- age4: Age in between 50-64;

- gender2: Female;
- educ2: Secondary school;
- lab1: Employed.

All these proportions get a negative coefficient in both models (see Table 8.1), which means that provinces with larger proportions of people in the above categories have smaller poverty incidence. See in Table 8.2 that the estimated spatial autocorrelation is similar in both models ($\widehat{\rho}_1 = 0.4617$ in MODEL A), which is quite significant. This indicates that direct estimators for neighbor provinces within the same region are correlated. For MODEL B, the autocorrelation parameter of the AR(1) process is $\widehat{\rho}_1 = 0.1817$, which is not very significant. See also in Table 8.3 that the loglikelihood of MODEL B is slightly larger but the other two goodness-of-fit measures, which penalize the log-likelihood with the number of unknown parameters, are better for MODEL A. This suggests that the AR(1) process might not be necessary, and this could be due to the fact that covariates explain sufficiently the time correlation.

	MODEL A				MODEL B			
Variable	Coef	Std. Error	t value	p value	Coef	Std. Error	t value	p value
constant	1.304	0.517	2.524	0	1.338	0.513	2.608	0
age4	-1.186	0.795	-1.492	0.003	-1.196	0.799	-1.497	0.003
gender2	-1.211	0.925	-1.31	0.01	-1.258	0.916	-1.372	0.007
edu2	-0.238	0.18	-1.326	0.009	-0.248	0.181	-1.372	0.007
lab1	-0.465	0.322	-1.446	0.005	-0.478	0.323	-1.478	0.004

Table 8.1: Estimated model coefficients obtained from fitting MODEL A and MODEL B to the poverty incidence.

	MODEL A		MODEL B	
Parameter	Estimate	Std. error	Estimate	Std. error
σ_1^2	0.0022	0.0012	0.0020	0.0013
ρ_1	0.4617	0.2566	0.4871	0.2664
σ_2^2	0.00065	0.0003	0.0007	0.0004
ρ_2			0.1817	0.7247

Table 8.2: Estimated variance components under MODEL A and MODEL B for the poverty incidence.

Model	loglike	AIC	BIC
MODEL A	247.146	-478.293	-453.894
MODEL B	247.210	-476.420	-448.971

Table 8.3: Goodness-of-fit measures of MODEL A and MODEL B for the poverty incidence.

Results of the fit of MODELS A and B for the poverty gap are given in Table 8.4. Observe that the same explanatory variables are significant and their coefficients get the same signs as in the model for the poverty incidence. Thus, the effect of these variables on the poverty gap is similar to the effect on the poverty incidence, only a little less significant. However, the spatial autocorrelation, ρ_1, is larger than in

	MODEL A				MODEL B			
Variable	Coef	Std. Error	t value	p value	Coef	Std. Error	t value	p value
constant	0.496	0.226	2.197	1.657e-05	0.495	0.226	2.186	1.829e-05
age4	-0.431	0.35	-1.231	0.016	-0.427	0.35	-1.222	0.017
gender2	-0.486	0.409	-1.19	0.02	-0.486	0.41	-1.185	0.02
edu2	-0.094	0.087	-1.092	0.032	-0.093	0.086	-1.076	0.035
lab1	-0.181	0.14	-1.299	0.011	-0.181	0.14	-1.3	0.011

Table 8.4: Estimated model coefficients obtained from fitting MODEL A and MODEL B to the poverty gap.

	MODEL A		MODEL B	
Parameter	Estimate	Std. error	Estimate	Std. error
σ_1^2	0.0003	0.0002	0.0003	0.0002
ρ_1	0.5430	0.2552	0.5316	0.2633
σ_2^2	0.0002	8.0320e-05	0.0002	0.0001
ρ_2			-0.04668	0.5182

Table 8.5: Estimated variance components under MODEL A and MODEL B for the poverty gap.

the model for the poverty incidence, but the autocorrelation of the AR(1) process, ρ_2, is now practically zero.

Table 7.7 lists direct, SEBA and SEBB small area estimates of poverty incidence obtained under MODELS A and B respectively. It shows also their estimated MSEs and their percent coefficients of variation. MSEs of SEBA and SEBB estimators have been obtained using the parametric bootstrap described in Section 7.3. Comparing with direct estimators, the SEBA and SEBB get a considerable reduction in CV for the provinces with smaller sample sizes such as Ávila, Castellón, Soria and Teruel, but there is gain in practically all provinces. Now comparing SEBA and SEBB, we cannot observe a clear gain of one with respect to the other, so it seems that the inclusion of the AR(1) process to explain the time correlation is not necessary in the Spatial model with the given covariates.

Table 7.8 lists the results for the poverty gap. Again, we can see that the SEBA and SEBB are better in CV than direct estimators, but the gain is smaller than in the models for the poverty incidence. Similarly as above, SEBA and SEBB are similar in efficiency and therefore probably the simpler model, MODEL A, is good enough.

Figure 8.1 (left) plots the values of the SEBA estimates of the poverty incidences, obtained under MODEL A, against the corresponding direct estimators for the Spanish provinces. Figure 8.1 (right) plots SEBB estimates, obtained under MODEL B, against direct estimates. Observe that the estimators obtained under both models are very similar, and in both cases lie along the line $y = x$, which means that they are approximately design-unbiased, except for three provinces, Soria, Avila and Cuenca. These provinces have a very small sample size and few observe people under the poverty line, and with these small sample sizes their large direct estimates are not very reliable. Thus, the two models are predicting a smaller poverty incidence for them.

The corresponding plots for the poverty gaps are given in Figure 8.2. Again, the model estimators seem to be design-unbiased in general, although the poverty gap of the province of Soria, with the

Model	loglike	AIC	BIC
MODEL A	365.209	-714.419	-690.020
MODEL B	365.222	-712.445	-684.995

Table 8.6: Goodness-of-fit measures of MODEL A and MODEL B for the poverty gap.

smallest sample size, is considerably decreased by the models.

Figure 8.3 displays the cartograms of estimated percent poverty incidences (left) and poverty gaps (right) for Spanish provinces, obtained using the SEBA and SEBB estimates (both estimators gave the same plots). The provinces with largest estimated poverty incidences, with more than 30% of people under the poverty line, are Badajoz, Cáceres, Córdoba, Jaen, Granada, Ávila and Cuenca. According to these estimates, the provinces with largest poverty gaps, whose average relative distance to the poverty line is over 10%, are Badajoz, Jaen, Granada, Albacete, Ávila and Zamora.

Table 7.7: Results on poverty incidence for Spanish SILC data: province, sample size, direct, SEBA and SEBB estimates (%), estimated variance of direct estimator, estimated MSEs of SEBA and SEBB ($\times 10^4$), CVs of direct, SEBA and SEBB (%).

Province	n_d	$\widehat{F}_{0d}^{w}$	$\widehat{F}_{0d}^{SEBA}$	$\widehat{F}_{0d}^{SEBB}$	var$(\widehat{F}_{0d}^{w})$	$mse(\widehat{F}_{0d}^{SEBA})$	$mse(\widehat{F}_{0d}^{SEBB})$	cv$(\widehat{F}_{0d}^{w})$	cv$(\widehat{F}_{0d}^{SEBA})$	cv$(\widehat{F}_{0d}^{SEBB})$
Álava	191	8.06	8.79	8.74	6.00	4.26	4.19	30.40	23.48	23.42
Albacete	346	26.15	27.36	27.25	8.19	4.92	4.84	10.94	8.11	8.07
Alicante	1078	17.54	17.66	17.65	1.86	1.56	1.52	7.77	7.06	7.00
Almería	397	33.73	30.83	30.73	9.24	5.56	5.56	9.01	7.65	7.68
Ávila	116	39.00	30.39	30.77	41.22	11.74	11.70	16.46	11.28	11.12
Badajoz	987	37.90	35.22	35.25	4.75	3.33	3.65	5.75	5.18	5.42
Baleares	1269	10.42	10.88	10.87	1.05	0.91	0.88	9.82	8.76	8.65
Barcelona	2841	9.62	9.62	9.62	0.34	0.31	0.30	6.03	5.81	5.68
Burgos	335	12.65	12.95	12.94	4.60	3.11	3.43	16.96	13.62	14.32
Cáceres	563	29.27	30.94	30.87	6.51	4.35	4.33	8.72	6.74	6.74
Cádiz	795	28.48	28.81	28.73	4.14	2.94	2.85	7.14	5.95	5.88
Castellón	236	12.22	15.33	15.29	7.16	4.57	4.99	21.90	13.95	14.61
Ciudad Real	499	29.42	29.69	29.70	6.72	4.64	4.31	8.81	7.26	6.99
Cordoba	447	30.78	31.30	31.17	7.83	4.65	4.38	9.09	6.89	6.72
Coruña La	990	22.75	22.30	22.30	2.45	1.97	1.95	6.88	6.30	6.26
Cuenca	183	41.21	32.69	32.50	31.34	10.07	10.39	13.58	9.71	9.92
Gerona	283	5.79	6.95	6.88	2.13	1.85	1.81	25.19	19.56	19.56
Granada	417	32.11	30.83	30.79	9.10	5.89	4.93	9.40	7.87	7.21
Guadalajara	178	11.89	14.23	14.28	6.50	4.12	4.30	21.46	14.26	14.52
Guipúzcoa	570	8.24	8.45	8.47	2.03	1.64	1.85	17.28	15.13	16.05
Huelva	244	22.29	23.37	23.26	8.62	5.09	5.39	13.17	9.65	9.98
Huesca	230	8.34	9.45	9.42	4.88	3.65	3.34	26.47	20.20	19.41
Jaén	463	31.01	29.37	29.55	7.00	4.19	4.83	8.53	6.97	7.44
Leon	437	19.22	18.84	18.89	5.55	4.11	4.02	12.26	10.75	10.61
Lérida	260	20.91	18.87	18.97	10.39	5.57	5.32	15.41	12.51	12.16
Rioja La	1019	18.97	18.73	18.78	2.71	2.11	2.34	8.68	7.76	8.15
Lugo	346	21.10	21.69	21.52	7.97	4.75	5.41	13.38	10.04	10.81
Madrid	1889	11.74	11.52	11.57	1.03	0.91	0.94	8.62	8.29	8.36
Málaga	758	24.09	23.20	23.27	3.61	3.00	2.72	7.89	7.46	7.09

Murcia	1770	23.72	23.55	23.57	1.85	1.54	1.54	5.74	5.26	5.26
Navarra	1128	9.24	9.22	9.22	1.08	1.12	0.93	11.23	11.49	10.43
Orense	258	23.72	20.19	20.13	15.82	6.95	6.69	16.77	13.06	12.85
Oviedo	1606	11.54	11.56	11.59	1.02	0.93	0.96	8.75	8.34	8.46
Palencia	143	25.13	23.34	23.19	18.48	8.07	8.34	17.11	12.17	12.45
Palmas Las	943	23.41	23.77	23.70	4.20	3.12	2.68	8.76	7.43	6.91
Pontevedra	896	19.40	18.62	18.70	2.97	2.31	2.42	8.89	8.17	8.32
Salamanca	328	31.87	29.01	29.05	12.23	6.19	6.10	10.98	8.58	8.51
Tenerife	762	27.49	25.88	26.23	4.76	3.53	3.40	7.94	7.26	7.03
Santander	867	11.24	11.28	11.36	1.89	1.55	1.62	12.24	11.04	11.20
Segovia	115	33.67	29.97	29.71	36.42	12.08	11.83	17.92	11.60	11.58
Sevilla	963	21.87	21.61	21.59	2.55	2.23	2.20	7.31	6.91	6.87
Soria	41	36.88	26.53	26.53	147.78	20.08	19.85	32.97	16.89	16.79
Tarragona	268	14.87	15.49	15.48	5.51	3.85	3.94	15.79	12.67	12.81
Teruel	144	12.03	12.96	13.05	9.04	5.72	5.36	25.00	18.46	17.75
Toledo	550	23.44	23.01	23.13	5.01	3.33	3.36	9.55	7.93	7.92
Valencia	1428	13.79	14.15	14.16	1.49	1.25	1.36	8.84	7.90	8.23
Valladolid	598	18.65	18.81	18.90	3.89	2.76	3.30	10.57	8.84	9.61
Vizcaya	1047	9.54	9.24	9.25	1.08	0.97	0.96	10.89	10.64	10.59
Zamora	209	30.06	29.94	29.88	14.90	7.57	7.32	12.84	9.19	9.05
Zaragoza	1129	11.67	11.62	11.66	1.38	1.22	1.15	10.05	9.51	9.19
Ceuta	470	36.27	36.59	36.46	9.20	5.82	5.24	8.36	6.59	6.28
Melilla	359	24.39	25.64	25.57	8.10	4.98	5.17	11.67	8.70	8.89

Table 7.8: Results on poverty gap for Spanish SILC data: province, sample size, direct, SEBA and SEBB estimates (%), estimated variance of direct estimator, estimated MSEs of SEBA and SEBB ($\times 10^4$), CVs of direct, SEBA and SEBB (%).

Province	n_d	$\widehat{F}^{w}_{1d}$	$\widehat{F}^{SEBA}_{1d}$	$\widehat{F}^{SEBB}_{1d}$	var($\widehat{F}^{w}_{1d}$)	$mse(\widehat{F}^{SEBA}_{1d})$	$mse(\widehat{F}^{SEBB}_{1d})$	cv($\widehat{F}^{w}_{1d}$)	cv($\widehat{F}^{SEBA}_{1d}$)	cv($\widehat{F}^{SEBB}_{1d}$)
Álava	191	1.99	2.12	2.12	0.40	0.39	0.40	31.93	29.67	29.82
Albacete	346	10.69	10.03	10.03	1.87	1.19	1.18	12.81	10.89	10.85
Alicante	1078	5.48	5.49	5.49	0.32	0.28	0.28	10.32	9.67	9.61
Almería	397	11.05	10.33	10.34	1.56	1.09	1.11	11.31	10.09	10.18
Ávila	116	11.64	10.07	10.02	4.69	2.03	2.06	18.61	14.14	14.32
Badajoz	987	12.38	11.69	11.69	0.79	0.65	0.66	7.16	6.87	6.96
Baleares	1269	2.88	3.02	3.03	0.18	0.16	0.16	14.67	13.29	13.26
Barcelona	2841	3.32	3.31	3.31	0.07	0.06	0.06	7.69	7.60	7.44
Burgos	335	3.89	4.11	4.10	1.04	0.75	0.78	26.17	21.13	21.53
Cáceres	563	8.42	8.80	8.80	0.77	0.62	0.63	10.43	8.97	9.00
Cádiz	795	9.07	9.23	9.24	0.70	0.55	0.55	9.24	8.06	8.03
Castellón	236	3.93	4.57	4.56	1.34	0.97	0.98	29.50	21.54	21.71
Ciudad Real	499	7.14	7.34	7.33	0.52	0.46	0.46	10.11	9.27	9.27
Cordoba	447	8.06	8.41	8.42	0.75	0.59	0.58	10.76	9.11	9.08
Coruña La	990	7.89	7.66	7.66	0.45	0.38	0.39	8.55	8.08	8.13
Cuenca	183	9.73	9.20	9.22	1.94	1.29	1.34	14.32	12.33	12.55
Gerona	283	2.00	2.16	2.17	0.33	0.32	0.32	28.91	26.04	26.07
Granada	417	15.12	11.94	11.93	2.95	1.65	1.62	11.36	10.77	10.69
Guadalajara	178	2.01	2.24	2.24	0.20	0.19	0.19	21.98	19.25	19.29
Guipúzcoa	570	3.54	3.36	3.35	0.61	0.50	0.51	22.05	21.16	21.29

Huelva	244	9.85	8.62	8.63	3.21	1.65	1.60	18.20	14.91	14.67
Huesca	230	2.79	2.96	2.96	0.79	0.66	0.63	31.87	27.49	26.77
Jaén	463	10.49	9.75	9.72	1.19	0.83	0.83	10.39	9.33	9.39
Leon	437	7.36	7.11	7.10	1.28	0.98	1.00	15.39	13.95	14.12
Lérida	260	9.33	7.35	7.32	2.75	1.54	1.50	17.76	16.90	16.70
Rioja La	1019	4.21	4.29	4.29	0.17	0.16	0.16	9.91	9.46	9.42
Lugo	346	6.45	6.32	6.36	2.40	1.36	1.35	24.04	18.43	18.27
Madrid	1889	3.45	3.40	3.40	0.20	0.18	0.18	12.88	12.49	12.51
Málaga	758	9.92	8.85	8.83	1.12	0.89	0.88	10.69	10.66	10.65
Murcia	1770	7.92	7.90	7.90	0.25	0.22	0.22	6.29	5.98	5.97
Navarra	1128	2.87	2.92	2.92	0.17	0.18	0.18	14.29	14.59	14.67
Orense	258	5.96	5.66	5.68	1.81	1.20	1.23	22.58	19.35	19.51
Oviedo	1606	2.83	2.87	2.87	0.12	0.12	0.12	12.38	11.84	11.91
Palencia	143	5.89	6.14	6.14	1.73	1.17	1.10	22.32	17.60	17.08
Palmas Las	943	8.02	7.90	7.90	0.94	0.74	0.73	12.12	10.89	10.82
Pontevedra	896	3.71	3.77	3.76	0.15	0.14	0.14	10.48	10.06	10.03
Salamanca	328	9.36	8.86	8.85	1.33	0.89	0.93	12.31	10.68	10.92
Tenerife	762	8.73	8.49	8.46	0.70	0.62	0.61	9.58	9.28	9.24
Santander	867	2.79	2.81	2.81	0.16	0.15	0.15	14.57	13.84	13.95
Segovia	115	8.97	8.79	8.82	3.47	1.84	1.80	20.77	15.44	15.24
Sevilla	963	3.99	4.07	4.07	0.13	0.14	0.14	9.10	9.17	9.16
Soria	41	19.10	8.42	8.43	70.78	3.84	4.27	44.04	23.29	24.50
Tarragona	268	2.36	2.54	2.54	0.19	0.18	0.18	18.44	16.57	16.72
Teruel	144	4.85	4.38	4.35	2.46	1.50	1.52	32.33	27.98	28.34
Toledo	550	6.75	6.81	6.81	0.63	0.49	0.50	11.75	10.32	10.40
Valencia	1428	4.66	4.66	4.66	0.35	0.30	0.30	12.70	11.81	11.72
Valladolid	598	6.84	6.65	6.65	0.72	0.56	0.59	12.41	11.30	11.51
Vizcaya	1047	2.42	2.40	2.40	0.12	0.11	0.11	14.22	14.01	13.77
Zamora	209	11.24	10.61	10.61	3.14	1.79	1.77	15.77	12.60	12.53
Zaragoza	1129	4.51	4.35	4.34	0.44	0.38	0.38	14.73	14.18	14.25
Ceuta	470	17.92	16.91	16.93	3.15	1.80	1.83	9.90	7.94	7.99
Melilla	359	12.20	12.74	12.76	3.42	1.73	1.74	15.16	10.33	10.34

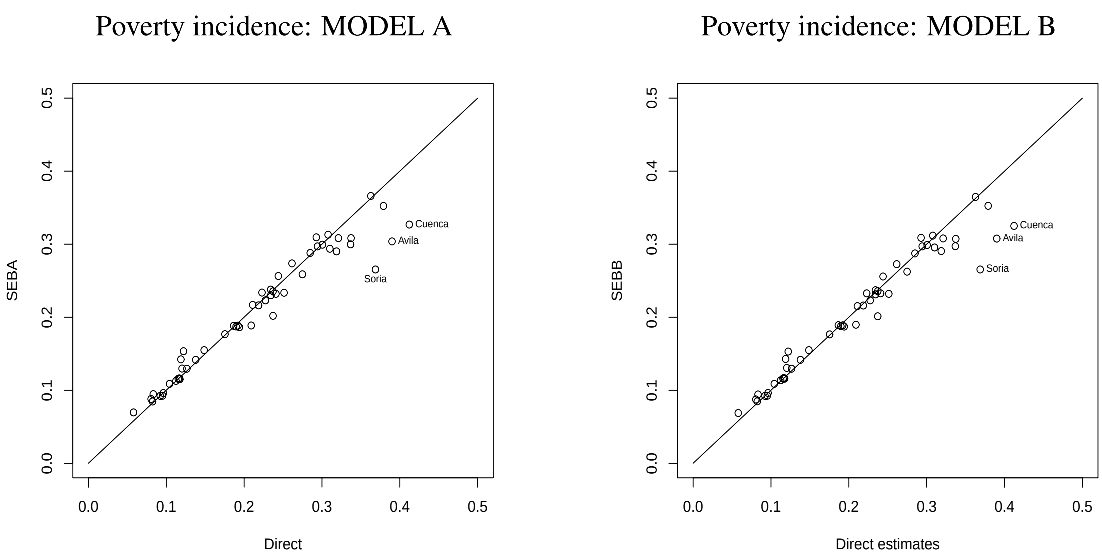


Figure 8.1: SEBA (left) and SEBB (right) versus direct estimators of poverty incidences in Spanish provinces.

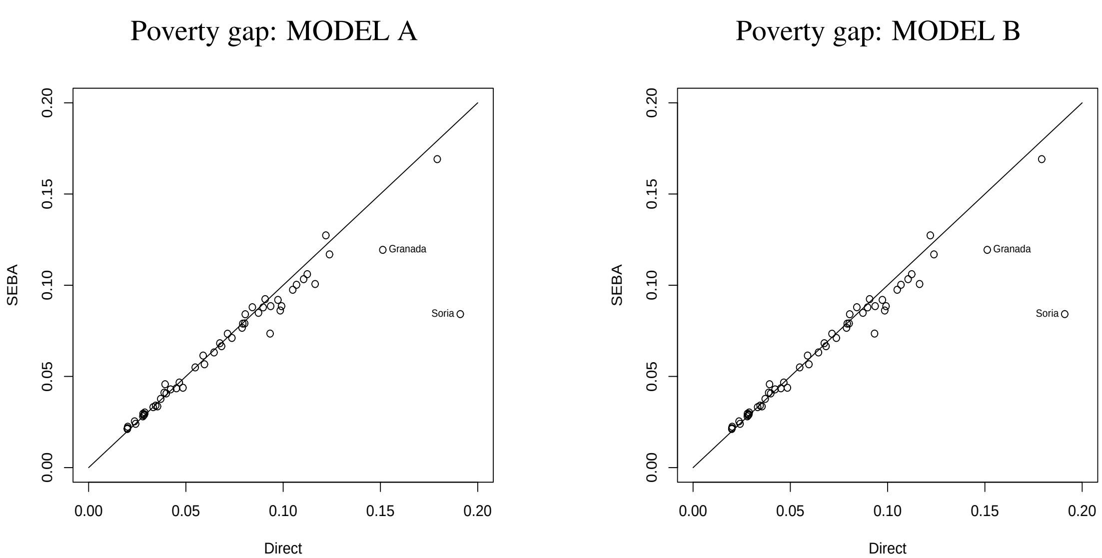


Figure 8.2: SEBA (left) and SEBB (right) versus direct estimators of poverty gap in Spanish provinces.

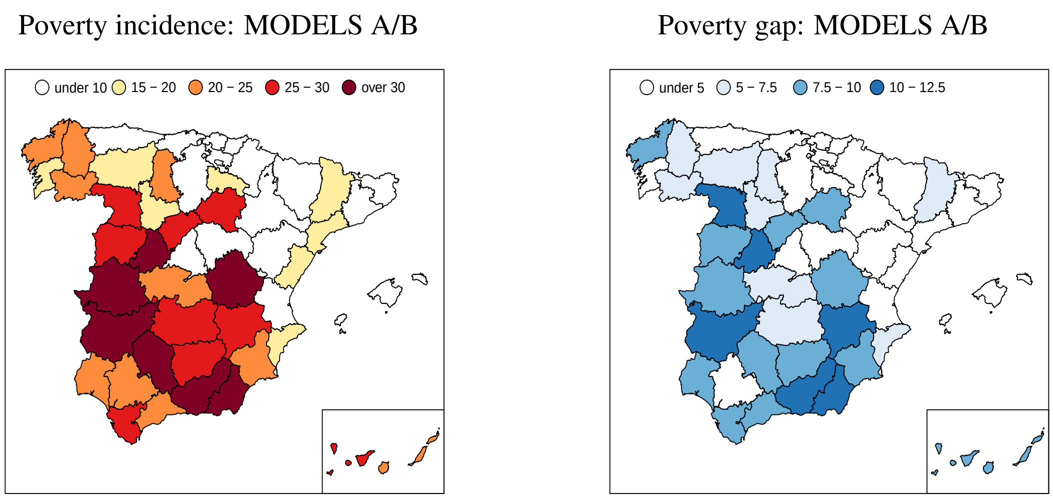


Figure 8.3: Cartograms of estimated percent poverty incidences (left) and poverty gaps (right) in Spanish provinces, obtained using MODEL A or MODEL B.

Chapter 9

Unit level time models

9.1 Unit-level model with correlated time effects

9.1.1 The model

Let us consider a with two nested random factors, where the first factor has D levels and, for each level d ($d=1,\dots,D$) of this factor, the second factor has m_d levels. More concretely, let us consider the model

$$\mathbf{y}=\mathbf{X}\beta+\mathbf{Z}_1\mathbf{u}_1+\mathbf{Z}_2\mathbf{u}_2+\mathbf{W}^{-1/2}\mathbf{e}, \tag{9.1}$$

where $\mathbf{u}_1=\mathbf{u}_{1,D\times 1}\sim N(0,\sigma_1^2\mathbf{I}_D)$, $\mathbf{u}_2=\mathbf{u}_{2,M\times 1}\sim N(0,\sigma_2^2\Omega(\rho))$ and $\mathbf{e}=\mathbf{e}_{n\times 1}\sim N(0,\sigma_0^2\mathbf{I}_n)$ are independent, $\mathbf{y}=\mathbf{y}_{n\times 1}$, $\mathbf{X}=\mathbf{X}_{n\times p}$ with $\mathrm{r}(\mathbf{X})=p$, $\beta=\beta_{p\times 1}$, $\mathbf{Z}_1=\underset{1\le d\le D}{\mathrm{diag}}\,(\mathbf{1}_{n_d})_{n\times D}$, $\mathbf{Z}_2=\underset{1\le d\le D}{\mathrm{diag}}\,(\underset{1\le t\le m_d}{\mathrm{diag}}\,(\mathbf{1}_{n_{dt}}))_{n\times M}$, $M=\sum_{d=1}^D m_d$, $n=\sum_{d=1}^D n_d$, $n_d=\sum_{t=1}^{m_d} n_{dt}$, $\mathbf{I}_a$ is the $a\times a$ identity matrix, $\mathbf{1}_a$ is the $a\times 1$ vector with all its elements equal to 1, $\mathbf{W}=\underset{1\le d\le D}{\mathrm{diag}}\,(\mathbf{W}_d)$, $\mathbf{W}_d=\underset{1\le t\le m_d}{\mathrm{diag}}\,(\mathbf{W}_{dt})$, $\mathbf{W}_{dt}=\underset{1\le j\le n_{dt}}{\mathrm{diag}}\,(w_{dtj})_{n\times n}$ with known $w_{dtj}>0$, $d=1,\dots,D$, $t=1,\dots,m_d$, $j=1,\dots,n_{dt}$, $\Omega(\rho)=\underset{1\le d\le D}{\mathrm{diag}}\,(\Omega_d)$ and

$$\Omega_d=\Omega_d(\rho)=\frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \dots & \rho^{m_d-2} & \rho^{m_d-1}\\ \rho & 1 & \ddots & & \rho^{m_d-2}\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ \rho^{m_d-2} & & \ddots & 1 & \rho\\ \rho^{m_d-1} & \rho^{m_d-2} & \dots & \rho & 1 \end{pmatrix}_{m_d\times m_d}.$$

Model (9.1) can alternatively be written in the form

$$y_{dtj}=\mathbf{x}_{dtj}\beta+u_{1,d}+u_{2,dt}+w_{dtj}^{-1/2}e_{dtj},\quad d=1,\dots,D, t=1,\dots,m_d, j=1,\dots,n_{dt}, \tag{9.2}$$

where y_{dtj} is the target variable for the sample unit j, time t and domain d, and $\mathbf{x}_{dtj}$ is the row (d,t,j) of matrix $\mathbf{X}$. The random vectors $(u_{2d1},\dots,u_{2dm_d})$, $d=1,\dots,D$, are i.i.d. AR(1).

In what follows we use the alternative parameters

$$\sigma^2=\sigma_0^2,\quad \varphi_1=\frac{\sigma_1^2}{\sigma_0^2},\quad \varphi_2=\frac{\sigma_2^2}{\sigma_0^2},\quad \rho=\rho.$$

Let $\sigma = (\sigma^2, \varphi_1, \varphi_2, \rho)$ be the vector of variance components, with $\sigma^2 > 0$, $\varphi_1 > 0$, $\varphi_2 > 0$ and $-1 < \rho < 1$. If σ is known, the BLUE of $\beta = (\beta_1, \ldots, \beta_p)'$ and the BLUP of $\mathbf{u} = (\mathbf{u}_1', \mathbf{u}_2')'$ are

$$\widehat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \qquad \text{and} \qquad \widehat{\mathbf{u}} = \mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\widehat{\beta}\right). \tag{9.3}$$

The empirical versions (EBLUP) of $\widehat{\beta}$ and $\widehat{\mathbf{u}}$ are obtained by substituting the variance componentes by their estimates. Here, REML estimators are considered.

9.1.2 The REML estimators

The restricted log-likelihood is

$$l_{reml}(\sigma) = -\frac{1}{2}(n-p)\log 2\pi - \frac{1}{2}(n-p)\log\sigma^2 - \frac{1}{2}\log|\mathbf{K}'\Sigma\mathbf{K}| - \frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{y},$$

where

$$\mathbf{P} = \mathbf{K}(\mathbf{K}'\Sigma\mathbf{K})^{-1}\mathbf{K}' = \Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1},\ \mathbf{K} = \mathbf{W} - \mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}.$$

Let us denote the derivatives of $\Omega(\rho)$ by $\Omega'(\rho) = \frac{\partial\Omega(\rho)}{\partial\rho}$ and $\Omega''(\rho) = \frac{\partial^2\Omega(\rho)}{\partial\rho^2}$. By taking partial derivatives with respect to σ^2, φ_1^2, φ_2^2 and ρ we obtain the components of the score vector $S(\sigma)$.

$$\begin{aligned}
S_{\sigma^2} &= -\frac{n-p}{2\sigma^2} + \frac{1}{2\sigma^4}\mathbf{y}'\mathbf{P}\mathbf{y},\\
S_{\varphi_1} &= -\frac{1}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\} + \frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{y},\\
S_{\varphi_2} &= -\frac{1}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\} + \frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{y},\\
S_{\rho} &= -\frac{\varphi_2}{2}\operatorname{tr}\left\{\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\right\} + \frac{\varphi_2}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{y},
\end{aligned}$$

The elements of the Fisher information matrix,

$$\begin{aligned}
F_{\sigma^2\sigma^2} &= -\frac{n-p}{2\sigma^4} + \frac{1}{\sigma^4}\operatorname{tr}\{\mathbf{P}\Sigma\} = \frac{n-p}{2\sigma^4}, \quad F_{\sigma^2\varphi_1} = \frac{1}{2\sigma^2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\},\\
F_{\sigma^2\varphi_2} &= \frac{1}{2\sigma^2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\}, \quad F_{\sigma^2\rho} = \frac{\varphi_2}{2\sigma^2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\},\\
F_{\varphi_1\varphi_1} &= \frac{1}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\}, \quad F_{\varphi_1\varphi_2} = \frac{1}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\}\\
F_{\varphi_1\rho} &= \frac{\varphi_2}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\}, \quad F_{\varphi_2\varphi_2} = \frac{1}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\},\\
F_{\varphi_2\rho} &= \frac{\varphi_2}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\Omega(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\}, \quad F_{\rho\rho} = \frac{\varphi_2^2}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\Omega'(\rho)\mathbf{Z}_2'\}.
\end{aligned}$$

The updating formula of the Fisher-scoring algorithm is

$$\sigma^{k+1} = \sigma^k + \mathbf{F}^{-1}(\sigma^k)\mathbf{S}(\sigma^k).$$

9.1.3 The EBLUP and its mean squared error

The EBLUP of the linear parameter $\eta = \mathbf{a}'\mathbf{y} = \mathbf{a}_s'\mathbf{y}_s + \mathbf{a}_r'\mathbf{y}_r$ is

$$\widehat{\eta} = \mathbf{a}_s'\mathbf{y}_s + \mathbf{a}_r'\left[\mathbf{X}_r\widehat{\beta} + \widehat{\mathbf{V}}_{rs}\widehat{\mathbf{V}}_{ss}^{-1}(\mathbf{y}_s - \mathbf{X}_s\widehat{\beta})\right]$$

As $\mathbf{V}_{ers} = \mathbf{0}$, $\mathbf{V}_{rs} = \mathbf{Z}_r\mathbf{V}_u\mathbf{Z}_s' + \mathbf{V}_{ers} = \mathbf{Z}_r\mathbf{V}_u\mathbf{Z}_s'$ and $\widehat{\mathbf{u}} = \Sigma_u\mathbf{Z}_s'\mathbf{V}_{ss}^{-1}(\mathbf{y}_s - \mathbf{X}_s\widehat{\beta})$, we get

$$\begin{aligned}\widehat{\eta} &= \mathbf{a}_s'\mathbf{y}_s + \mathbf{a}_r'\left[\mathbf{X}_r\widehat{\beta} + \mathbf{Z}_r\Sigma_u\mathbf{Z}_s'\widehat{\mathbf{V}}_{ss}^{-1}(\mathbf{y}_s - \mathbf{X}_s\widehat{\beta})\right] = \mathbf{a}_s'\mathbf{y}_s + \mathbf{a}_r'\left[\mathbf{X}_r\widehat{\beta} + \mathbf{Z}_r\widehat{\mathbf{u}}\right] \\ &= \mathbf{a}'\left[\mathbf{X}\widehat{\beta} + \mathbf{Z}_1\widehat{\mathbf{u}}_1 + \mathbf{Z}_2\widehat{\mathbf{u}}_2\right] + \mathbf{a}_s'\left[\mathbf{y}_s - \mathbf{X}_s\widehat{\beta} - \mathbf{Z}_{s1}\widehat{\mathbf{u}}_1 - \mathbf{Z}_{s2}\widehat{\mathbf{u}}_2\right].\end{aligned}$$

Under model (9.2), $\overline{Y}_{dt} = \frac{1}{N_{dt}}\sum_{j=1}^{N_{dt}} y_{dtj}$ can be written as a linear parameter $\eta = \mathbf{a}'\mathbf{y}$, where

$$\begin{aligned}\mathbf{a}' &= \frac{1}{N_{dt}}(\mathbf{0}_{N_1}', \ldots, \mathbf{0}_{N_{d-1}}', \mathbf{0}_{N_{d1}}', \ldots, \mathbf{0}_{N_{d(i-1)}}', \mathbf{1}_{N_{dt}}', \mathbf{0}_{N_{d(i+1)}}', \ldots, \mathbf{0}_{N_{dm_d}}', \mathbf{0}_{N_{d+1}}', \ldots, \mathbf{0}_{N_D}') \\ &= \frac{1}{N_{dt}}(\mathbf{0}_{N_1}', \ldots, \mathbf{0}_{N_{d-1}}', \underset{1\le k\le m_d}{\text{col}'}\left[\delta_{tk}\mathbf{1}_{N_{dk}}'\right], \mathbf{0}_{N_{d+1}}', \ldots, \mathbf{0}_{N_D}') = \frac{1}{N_{dt}}\underset{1\le \ell\le D}{\text{col}'}\{\delta_{d\ell}\underset{1\le k\le m_\ell}{\text{col}'}[\delta_{tk}\mathbf{1}_{N_{\ell k}}']\}\end{aligned}$$

with $\delta_{ab} = 1$ si $a = b$ and $\delta_{ab} = 0$ si $a \neq b$. It holds that $\mathbf{a}'\mathbf{X} = \overline{X}_{dt}$,

$$\begin{aligned}\mathbf{a}'\mathbf{Z}_1 &= \frac{1}{N_{dt}}\underset{1\le \ell\le D}{\text{col}'}\{\delta_{d\ell}\underset{1\le k\le m_\ell}{\text{col}'}[\delta_{tk}\mathbf{1}_{N_{\ell k}}']\}\underset{1\le \ell\le D}{\text{diag}}(\mathbf{1}_{N_\ell}) = \underset{1\le \ell\le D}{\text{col}'}\{\delta_{d\ell}\} = \overline{\mathbf{Z}}_{1,dt}, \\ \mathbf{a}'\mathbf{Z}_2 &= \frac{1}{N_{dt}}\underset{1\le \ell\le D}{\text{col}'}\{\delta_{d\ell}\underset{1\le k\le m_\ell}{\text{col}'}[\delta_{tk}\mathbf{1}_{N_{\ell k}}']\}\underset{1\le \ell\le D}{\text{diag}}(\underset{1\le k\le m_\ell}{\text{diag}}(\mathbf{1}_{N_{\ell k}})) = \underset{1\le \ell\le D}{\text{col}'}\{\underset{1\le k\le m_\ell}{\text{col}'}\{\delta_{d\ell}\delta_{tk}\}\} = \overline{\mathbf{Z}}_{2,dt}.\end{aligned}$$

If $n_{dt} > 0$, the EBLUP of $\overline{Y}_{dt}$ is

$$\widehat{\overline{Y}}_{dt}^{eblup} = \overline{\mathbf{X}}_{dt}\widehat{\beta} + \overline{\mathbf{Z}}_{1,dt}\widehat{\mathbf{u}}_1 + \overline{\mathbf{Z}}_{2,dt}\widehat{\mathbf{u}}_2 + f_{dt}\left[\overline{\mathbf{y}}_{s,dt} - \overline{\mathbf{X}}_{s,dt}\widehat{\beta} - \overline{\mathbf{Z}}_{1,dt}\widehat{\mathbf{u}}_1 - \overline{\mathbf{Z}}_{2,dt}\widehat{\mathbf{u}}_2\right],$$

where $\overline{\mathbf{y}}_{s,dt} = \frac{1}{n_{dt}}\sum_{j=1}^{n_{dt}} y_{dtj}$, $\overline{\mathbf{X}}_{s,dt} = \frac{1}{n_{dt}}\sum_{j=1}^{n_{dt}}\mathbf{x}_{dtj}$ and $f_{dt} = \frac{n_{dt}}{N_{dt}}$. If $n_{dt} = 0$, the EBLUP of $\overline{Y}_{dt}$ is the synthetic part

$$\widehat{\overline{Y}}_{dt}^{eblup} = \overline{\mathbf{X}}_{dt}\widehat{\beta} + \overline{\mathbf{Z}}_{1,dt}\widehat{\mathbf{u}}_1 + \overline{\mathbf{Z}}_{2,dt}\widehat{\mathbf{u}}_2.$$

Let $\theta = (\sigma_0^2, \varphi_1, \varphi_2, \rho)$ be the vector of variance components. A second order approximation to the mean squared error of the EBLUP is

$$MSE(\widehat{\overline{Y}}_{dt}^{eblup}) = g_1(\theta) + g_2(\theta) + g_3(\theta) + g_4(\theta),$$

where

$$\begin{aligned}g_1(\theta) &= \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r, \\ g_2(\theta) &= [\mathbf{a}_r'\mathbf{X}_r - \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_s'\mathbf{V}_{es}^{-1}\mathbf{X}_s]\mathbf{Q}_s[\mathbf{X}_r'\mathbf{a}_r - \mathbf{X}_s'\mathbf{V}_{es}^{-1}\mathbf{Z}_s\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r], \\ g_3(\theta) &\approx \text{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}_s(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\}, \\ g_4(\theta) &= \mathbf{a}_r'\mathbf{V}_{er}\mathbf{a}_r,\end{aligned}$$

and

$$
\begin{aligned}
\mathbf{a}_r' &= \frac{1}{N_{dt}}\left(\mathbf{0}_{N_1-n_1}',\ldots,\mathbf{0}_{N_{d-1}-n_{d-1}}', \underset{1\le k\le m_d}{\text{col}'}\left[\delta_{tk}\mathbf{1}_{N_{dk}-n_{dk}}'\right], \mathbf{0}_{N_{d+1}-n_{d+1}}',\ldots,\mathbf{0}_{N_D-n_D}'\right),\\
\mathbf{Z}_r &= \left[\mathbf{Z}_{1r}\,\mathbf{Z}_{2r}\right],\ \mathbf{T}_s=\mathbf{V}_u-\mathbf{V}_u\mathbf{Z}_s'\mathbf{V}_s^{-1}\mathbf{Z}_s\mathbf{V}_u,\ \mathbf{V}_u=\begin{pmatrix}\sigma_1^2\mathbf{I}_D & \mathbf{0}\\ \mathbf{0} & \sigma_2^2\Omega(\rho)\end{pmatrix},\ \mathbf{V}_{er}^{-1}=\sigma^{-2}\mathbf{W}_r,\\
\mathbf{Z}_s &= \left[\mathbf{Z}_{1s}\,\mathbf{Z}_{2s}\right],\quad \mathbf{V}_s^{-1}=\underset{1\le d\le D}{\text{diag}}\{\mathbf{V}_{ds}^{-1}\},\quad \mathbf{Q}_s=(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1},\quad \mathbf{b}'=\mathbf{a}_r'\mathbf{Z}_r\mathbf{V}_u\mathbf{Z}_s'\mathbf{V}_s^{-1}.
\end{aligned}
$$

9.2 Unit-level model with independent time effects

9.2.1 The model

Let us consider the model

$$\mathbf{y}=\mathbf{X}\beta+\mathbf{Z}_1\mathbf{u}_1+\mathbf{Z}_2\mathbf{u}_2+\mathbf{W}^{-1/2}\mathbf{e}, \tag{9.4}$$

where $\mathbf{u}_1=\mathbf{u}_{1,D\times 1}\sim N(0,\sigma_1^2\mathbf{I}_D)$, $\mathbf{u}_2=\mathbf{u}_{2,M\times 1}\sim N(0,\sigma_2^2\mathbf{I}_M)$ and $\mathbf{e}=\mathbf{e}_{n\times 1}\sim N(0,\sigma_0^2\mathbf{I}_n)$ are independent, $\mathbf{y}=\mathbf{y}_{n\times 1}$, $\mathbf{X}=\mathbf{X}_{n\times p}$ with $\text{r}(\mathbf{X})=p$, $\beta=\beta_{p\times 1}$, $\mathbf{Z}_1=\underset{1\le d\le D}{\text{diag}}(\mathbf{1}_{n_d})_{n\times D}$, $\mathbf{Z}_2=\underset{1\le d\le D}{\text{diag}}(\underset{1\le t\le m_d}{\text{diag}}(\mathbf{1}_{n_{dt}}))_{n\times M}$, $M=\sum_{d=1}^D m_d$, $n=\sum_{d=1}^D n_d$, $n_d=\sum_{t=1}^{m_d} n_{dt}$, $\mathbf{I}_a$ is the $a\times a$ identity matrix, $\mathbf{1}_a$ is the $a\times 1$ vector with all its elements equal to 1, $\mathbf{W}=\underset{1\le d\le D}{\text{diag}}(\mathbf{W}_d)$, $\mathbf{W}_d=\underset{1\le t\le m_d}{\text{diag}}(\mathbf{W}_{dt})$, $\mathbf{W}_{dt}=\underset{1\le j\le n_{dt}}{\text{diag}}(w_{dtj})_{n\times n}$ with known $w_{dtj}>0$, $d=1,\ldots,D$, $t=1,\ldots,m_d$, $j=1,\ldots,n_{dt}$. Model (9.4) can alternatively be written in the form

$$y_{dtj}=\mathbf{x}_{dtj}\beta+u_{1,d}+u_{2,dt}+w_{dtj}^{-1/2}e_{dtj},\quad d=1,\ldots,D,t=1,\ldots,m_d,j=1,\ldots,n_{dt}, \tag{9.5}$$

where y_{dtj} is the target variable for the sample unit j, time t and domain d, and $\mathbf{x}_{dtj}$ is the row (d,t,j) of matrix $\mathbf{X}$. In what follows we use the alternative parameters

$$\sigma^2=\sigma_0^2,\quad \varphi_1=\frac{\sigma_1^2}{\sigma_0^2},\quad \varphi_2=\frac{\sigma_2^2}{\sigma_0^2}.$$

Let $\sigma=(\sigma^2,\varphi_1,\varphi_2)$ be the vector of variance components, with $\sigma^2>0$, $\varphi_1>0$ and $\varphi_2>0$. If σ is known, the BLUE of $\beta=(\beta_1,\ldots,\beta_p)'$ and the BLUP of $\mathbf{u}=(\mathbf{u}_1',\mathbf{u}_2')'$ are

$$\widehat{\beta}=(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\qquad \text{y}\qquad \widehat{\mathbf{u}}=\mathbf{V}_u\mathbf{Z}'\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\widehat{\beta}\right). \tag{9.6}$$

The empirical versions (EBLUP) of $\widehat{\beta}$ and $\widehat{\mathbf{u}}$ are obtained by substituting the variance componentes by their estimates. Here, REML estimators are considered.

9.2.2 The REML estimators

The REML log-likelihood function is

$$l_{reml}(\sigma)=-\frac{1}{2}(n-p)\log 2\pi-\frac{1}{2}(n-p)\log\sigma^2-\frac{1}{2}\log|\mathbf{K}'\Sigma\mathbf{K}|-\frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{y},$$

where

$$\mathbf{P}=\mathbf{K}(\mathbf{K}'\Sigma\mathbf{K})^{-1}\mathbf{K}'=\Sigma^{-1}-\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1},\ \mathbf{K}=\mathbf{W}-\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}.$$

By taking partial derivatives with respect to σ^2, φ_1^2 and φ_2^2 we get the components of the vectors of scores $S(\sigma)$.

$$\begin{aligned}
S_{\sigma^2} &= -\frac{n-p}{2\sigma^2}+\frac{1}{2\sigma^4}\mathbf{y}'\mathbf{P}\mathbf{y},\\
S_{\varphi_1} &= -\frac{1}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\}+\frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{y},\\
S_{\varphi_2} &= -\frac{1}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\}+\frac{1}{2\sigma^2}\mathbf{y}'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\mathbf{P}\mathbf{y},
\end{aligned}$$

The elements of the Fisher information matrix are

$$\begin{aligned}
F_{\sigma^2\sigma^2} &= -\frac{n-p}{2\sigma^4}+\frac{1}{\sigma^4}\operatorname{tr}\{\mathbf{P}\Sigma\}=\frac{n-p}{2\sigma^4},\quad F_{\sigma^2\varphi_1}=\frac{1}{2\sigma^2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\},\\
F_{\sigma^2\varphi_2} &= \frac{1}{2\sigma^2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\},\quad F_{\varphi_1\varphi_1}=\frac{1}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\},\\
F_{\varphi_1\varphi_2} &= \frac{1}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_1\mathbf{Z}_1'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\},\quad F_{\varphi_2\varphi_2}=\frac{1}{2}\operatorname{tr}\{\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\mathbf{P}\mathbf{Z}_2\mathbf{Z}_2'\}.
\end{aligned}$$

The updating formula of the Fisher-scoring algorithm is

$$\sigma^{k+1}=\sigma^k+\mathbf{F}^{-1}(\sigma^k)\mathbf{S}(\sigma^k).$$

9.2.3 The EBLUP and its mean squared error

The EBLUP of the linear parameter $\eta=\mathbf{a}'\mathbf{y}=\mathbf{a}_s'\mathbf{y}_s+\mathbf{a}_r'\mathbf{y}_r$ is

$$\widehat{\eta}=\mathbf{a}_s'\mathbf{y}_s+\mathbf{a}_r'\left[\mathbf{X}_r\widehat{\beta}+\widehat{\mathbf{V}}_{rs}\widehat{\mathbf{V}}_{ss}^{-1}(\mathbf{y}_s-\mathbf{X}_s\widehat{\beta})\right]$$

As $\mathbf{V}_{ers}=\mathbf{0}$, $\mathbf{V}_{rs}=\mathbf{Z}_r\mathbf{V}_u\mathbf{Z}_s'+\mathbf{V}_{ers}=\mathbf{Z}_r\mathbf{V}_u\mathbf{Z}_s'$ and $\widehat{\mathbf{u}}=\mathbf{V}_u\mathbf{Z}_s'\mathbf{V}_{ss}^{-1}(\mathbf{y}_s-\mathbf{X}_s\widehat{\beta})$, we have

$$\begin{aligned}
\widehat{\eta} &= \mathbf{a}_s'\mathbf{y}_s+\mathbf{a}_r'\left[\mathbf{X}_r\widehat{\beta}+\mathbf{Z}_r\widehat{\mathbf{V}}_u\mathbf{Z}_s'\widehat{\mathbf{V}}_{ss}^{-1}(\mathbf{y}_s-\mathbf{X}_s\widehat{\beta})\right]=\mathbf{a}_s'\mathbf{y}_s+\mathbf{a}_r'\left[\mathbf{X}_r\widehat{\beta}+\mathbf{Z}_r\widehat{\mathbf{u}}\right]\\
&= \mathbf{a}'\left[\mathbf{X}\widehat{\beta}+\mathbf{Z}_1\widehat{\mathbf{u}}_1+\mathbf{Z}_2\widehat{\mathbf{u}}_2\right]+\mathbf{a}_s'\left[\mathbf{y}_s-\mathbf{X}_s\widehat{\beta}-\mathbf{Z}_{s1}\widehat{\mathbf{u}}_1-\mathbf{Z}_{s2}\widehat{\mathbf{u}}_2\right].
\end{aligned}$$

Under model (9.5), $\overline{Y}_{dt}=\frac{1}{N_{dt}}\sum_{j=1}^{N_{dt}}y_{dtj}$ can be written as a linear parameter $\eta=\mathbf{a}'\mathbf{y}$, where

$$\begin{aligned}
\mathbf{a}' &= \frac{1}{N_{dt}}(\mathbf{0}_{N_1}',\ldots,\mathbf{0}_{N_{d-1}}',\mathbf{0}_{N_{d1}}',\ldots,\mathbf{0}_{N_{d(i-1)}}',\mathbf{1}_{N_{dt}}',\mathbf{0}_{N_{d(i+1)}}',\ldots,\mathbf{0}_{N_{dm_d}}',\mathbf{0}_{N_{d+1}}',\ldots,\mathbf{0}_{N_D}')\\
&= \frac{1}{N_{dt}}(\mathbf{0}_{N_1}',\ldots,\mathbf{0}_{N_{d-1}}',\underset{1\le k\le m_d}{\operatorname{col}'}\left[\delta_{tk}\mathbf{1}_{N_{dk}}'\right],\mathbf{0}_{N_{d+1}}',\ldots,\mathbf{0}_{N_D}')=\frac{1}{N_{dt}}\underset{1\le\ell\le D}{\operatorname{col}'}\{\delta_{d\ell}\underset{1\le k\le m_\ell}{\operatorname{col}'}[\delta_{tk}\mathbf{1}_{N_{\ell k}}']\},
\end{aligned}$$

with $\delta_{ab}=1$ if $a=b$ and $\delta_{ab}=0$ if $a\neq b$. It hold that $\mathbf{a}'\mathbf{X}=\overline{X}_{dt}$,

$$\begin{aligned}
\mathbf{a}'\mathbf{Z}_1 &= \frac{1}{N_{dt}}\underset{1\le\ell\le D}{\text{col}'}\{\delta_{d\ell}\underset{1\le k\le m_\ell}{\text{col}'}[\delta_{tk}\mathbf{1}'_{N_{\ell k}}]\}\underset{1\le\ell\le D}{\text{diag}}(\mathbf{1}_{N_\ell})=\underset{1\le\ell\le D}{\text{col}'}\{\delta_{d\ell}\}=\overline{\mathbf{Z}}_{1,dt},\\
\mathbf{a}'\mathbf{Z}_2 &= \frac{1}{N_{dt}}\underset{1\le\ell\le D}{\text{col}'}\{\delta_{d\ell}\underset{1\le k\le m_\ell}{\text{col}'}[\delta_{tk}\mathbf{1}'_{N_{\ell k}}]\}\underset{1\le\ell\le D}{\text{diag}}(\underset{1\le k\le m_\ell}{\text{diag}}(\mathbf{1}_{N_{\ell k}}))=\underset{1\le\ell\le D}{\text{col}'}\{\underset{1\le k\le m_\ell}{\text{col}'}\{\delta_{d\ell}\delta_{tk}\}\}=\overline{\mathbf{Z}}_{2,dt}.
\end{aligned}$$

If $n_{dt}>0$, the EBLUP of $\overline{Y}_{dt}$ is

$$\widehat{\overline{Y}}_{dt}^{eblup}=\overline{\mathbf{X}}_{dt}\widehat{\beta}+\overline{\mathbf{Z}}_{1,dt}\widehat{\mathbf{u}}_1+\overline{\mathbf{Z}}_{2,dt}\widehat{\mathbf{u}}_2+f_{dt}\left[\overline{\mathbf{y}}_{s,dt}-\overline{\mathbf{X}}_{s,dt}\widehat{\beta}-\overline{\mathbf{Z}}_{1,dt}\widehat{\mathbf{u}}_1-\overline{\mathbf{Z}}_{2,dt}\widehat{\mathbf{u}}_2\right],$$

where $\overline{\mathbf{y}}_{s,dt}=\frac{1}{n_{dt}}\sum_{j=1}^{n_{dt}}y_{dtj}$, $\overline{\mathbf{X}}_{s,dt}=\frac{1}{n_{dt}}\sum_{j=1}^{n_{dt}}\mathbf{x}_{dtj}$ and $f_{dt}=\frac{n_{dt}}{N_{dt}}$.

If $n_{dt}=0$, the EBLUP of $\overline{Y}_{dt}$ is

$$\widehat{\overline{Y}}_{dt}^{eblup}=\overline{\mathbf{X}}_{dt}\widehat{\beta}+\overline{\mathbf{Z}}_{1,dt}\widehat{\mathbf{u}}_1+\overline{\mathbf{Z}}_{2,dt}\widehat{\mathbf{u}}_2.$$

Let $\theta=(\sigma_0^2,\varphi_1,\varphi_2)$ be the vector of variance components. The mean squared error of the EBLUP of $\overline{Y}_{dt}$ is

$$MSE(\widehat{\overline{Y}}_{dt}^{eblup})=g_1(\theta)+g_2(\theta)+g_3(\theta)+g_4(\theta),$$

where

$$\begin{aligned}
g_1(\theta) &= \mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r,\\
g_2(\theta) &= [\mathbf{a}_r'\mathbf{X}_r-\mathbf{a}_r'\mathbf{Z}_r\mathbf{T}_s\mathbf{Z}_s'\mathbf{V}_{es}^{-1}\mathbf{X}_s]\mathbf{Q}_s[\mathbf{X}_r'\mathbf{a}_r-\mathbf{X}_s'\mathbf{V}_{es}^{-1}\mathbf{Z}_s\mathbf{T}_s\mathbf{Z}_r'\mathbf{a}_r],\\
g_3(\theta) &\approx \text{tr}\left\{(\nabla\mathbf{b}')\mathbf{V}_s(\nabla\mathbf{b}')'E\left[(\widehat{\theta}-\theta)(\widehat{\theta}-\theta)'\right]\right\},\\
g_4(\theta) &= \mathbf{a}_r'\mathbf{V}_{er}\mathbf{a}_r,
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{a}_r' &= \frac{1}{N_{dt}}\left(\mathbf{0}'_{N_1-n_1},\ldots,\mathbf{0}'_{N_{d-1}-n_{d-1}},\underset{1\le k\le m_d}{\text{col}'}\left[\delta_{tk}\mathbf{1}'_{N_{dk}-n_{dk}}\right],\mathbf{0}'_{N_{d+1}-n_{d+1}},\ldots,\mathbf{0}'_{N_D-n_D}\right),\\
\mathbf{Z}_r &= [\mathbf{Z}_{1r}\,\mathbf{Z}_{2r}],\ \mathbf{T}_s=\mathbf{V}_u-\mathbf{V}_u\mathbf{Z}_s'\mathbf{V}_s^{-1}\mathbf{Z}_s\Sigma_u,\ \mathbf{V}_u=\begin{pmatrix}\sigma_1^2\mathbf{I}_D & \mathbf{0}\\ \mathbf{0} & \sigma_2^2\mathbf{I}_M\end{pmatrix},{}^{\text{o}},\mathbf{V}_{er}=\sigma^2\mathbf{W}_r^{-1},\\
\mathbf{Z}_s &= [\mathbf{Z}_{1s}\,\mathbf{Z}_{2s}],\quad \mathbf{V}_s^{-1}=\underset{1\le d\le D}{\text{diag}}\{\mathbf{V}_{ds}^{-1}\},\quad \mathbf{Q}_s=(\mathbf{X}_s'\mathbf{V}^{-1}\mathbf{X}_s)^{-1},\quad \mathbf{b}'=\mathbf{a}_r'\mathbf{Z}_r\mathbf{V}_u\mathbf{Z}_s'\mathbf{V}_s^{-1}.
\end{aligned}$$

9.3 Estimation of poverty indicators

Let us consider a finite population P_t partitioned into D domains P_{dt} at time period t, and denote their sizes by N_t and N_{dt}, $d=1,\ldots,D$. Let z_{dtj} be an income variable measured in all the units of the population and let z_t be the poverty line, so that units with $z_{dtj}<z_t$ are considered as poor at time period t. The main goal of this section is to estimate the poverty incidence (proportion of individuals under poverty) and the

poverty gap in Spanish domains. These two measures belongs to the FGT family proposed by Foster et al. (1984), given by

$$Y_{\alpha;dt} = \frac{1}{N_{dt}} \sum_{j=1}^{N_{dt}} y_{\alpha;dtj}, \quad \text{where } y_{\alpha;dtj} = \left(\frac{z_t - z_{dtj}}{z_t} \right)^{\alpha} I(z_{dtj} < z_t), \tag{9.7}$$

$I(z_{dtj} < z_t) = 1$ if $z_{dtj} < z_t$ and $I(z_{dtj} < z_t) = 0$ otherwise. The proportion of units under poverty in the domain d and period t is thus $Y_{0;dt}$ and the poverty gap is $Y_{1;dt}$.

We use the data from the Spanish Living Conditions Survey (SLCS) corresponding to years 2004-2006 and described in Tables 1.1. We consider $D = 104$ domains obtained by crossing 52 provinces with 2 sexes. Domains d (province - sex) are partitioned in subdomains i (census sections - sex), so that we construct new data files aggregating the data at the census section level. As the target variables $y_{\alpha;dti}$, $\alpha = 0,1$, are not continuous at the unit-level, we can obtain avoid this problem by doing aggregations at the subdomain level. The idea is thus treating the census section as basic units and to fit unit-level linear mixed models to the new generated data.

The new unit-level target variable will be $\bar{y}_{dti} = \hat{y}_{dti}^{dir} / \hat{N}_{dti}^{dir}$, where

$$\hat{y}_{dti}^{dir} = \sum_{j \in S_{dti}} w_{dtij} y_{dtij}, \quad \hat{N}_{dti}^{dir} = \sum_{j \in S_{dti}} w_{dtij},$$

S_{dti} is the sample and w_{dtij} is the official calibrated sampling weight of individual j in subdomain i, domain d and time instant t. These weights take into account for non response. Similar formulas are applied for the calculation of the new unit-level auxiliary variables $\bar{\mathbf{x}}_{dti}$.

Therefore, we consider the unit-level (units are census sections) linear mixed model

$$\bar{y}_{dti} = \bar{\mathbf{x}}_{dti}\beta + u_{1,d} + u_{2,dt} + e_{dti}, \quad d = 1,\ldots,D,\ t = 1,\ldots,m_d,\ i = 1,\ldots,n_{dt},$$

where $\bar{y}_{dti}$ is de direct estimate of the average of variables $y_{\alpha;dti}$ at the census section level, n_{dt} is the total number of sampled census sections within domain d and time instant t, $\mathbf{x}_{dti}$ is the $1 \times p$ vector containing the values of the selected explanatory variables, which are the direct estimates of the census section category proportions of the auxiliary variables for domain d, time instant t and census section i. The considered auxiliary variables are the domain means of the category indicators of the variables AGE, EDUCATION, CITIZENSHIP and LABOR described in Section 1.2.

We first consider $1 \times p$ vector $\bar{\mathbf{x}}_{dti}$ containing the values of all the categories (except the last one) of the explanatory variables. The first position of $\bar{\mathbf{x}}_{dti}$ contains a "1", so that $p = 1+4+3+1+3 = 12$. Random effects errors are assumed to follow the distributional assumptions of model (3.1) either restricted to $\rho = 0$ (model 0) or without this restriction (model 1). As some of the explanatory variables where not significative, the starting models where simplified to include only the auxiliary variables appearing in Tables 9.3.1. As the 90% confidence interval for ρ is $-0.5492 \pm 1,3435$ for $\alpha = 0$, we recommend model 1. Regression parameters and their corresponding p-values are presented in Table 9.3.1.

model 1	*constant*	*age2*	*age3*	*age4*	*age5*	*edu1*	*edu2*	*cit1*	*lab1*
β	0.5044	-0.2328	-0.2551	-0.2891	-0.2475	0.3156	0.1577	-0.1670	-0.1757
inf	0.3210	-0.4145	-0.4207	-0.4429	-0.3993	0.2221	0.0648	-0.3101	-0.2682
sup	0.6879	-0.0511	-0.0895	-0.1353	-0.0957	0.4091	0.2506	-0.0239	-0.0832
p	0.0000	0.0350	0.0113	0.0020	0.0073	0.0000	0.0052	0.0550	0.0018
model 0	*constant*	*age2*	*age3*	*age4*	*age5*	*edu1*	*edu2*	*cit1*	*lab1*
β	0.5047	-0.2327	-0.2547	-0.2888	-0.2472	0.3154	0.1576	-0.1668	-0.1758
inf	0.3213	-0.4142	-0.4201	-0.4426	-0.3990	0.2218	0.0647	-0.3100	-0.2682
sup	0.6882	-0.0511	-0.0892	-0.1350	-0.0954	0.4089	0.2505	-0.0237	-0.0835
p	0.0000	0.0350	0.0113	0.0020	0.0074	0.0000	0.0053	0.0552	0.0017

Table 9.3.1. Regression parameters, confidence intervals and p-values for $\alpha = 0$.

By observing the signs of the regression parameters 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 greater than 16, Spanish citizenship and with higher proportion of employed people. The education coefficients for elementary, secondary and university education are 0.3156, 0.1577 and 0 respectively, which means that poverty proportion is, as expected, inversely related with the level of education.

Table 9.3.2 presents the variance components estimates and their asymptotic confidence intervals for $\alpha = 0$ under models 1 and 0. We observe that the confidence intervals for σ_0^2, under both models, do not contain the 0. This is something somewhat expected as the global sample size is large enough to separate the obtained estimates (0.03795 or 0.03794) from 0 with a 90% confidentiality. This fact does not happens with the other two variances, maybe because the number of domains ($D = 2 \times 52 = 104$) and the number of time instants $T = 3$ are not large enough. The correlation coefficient ρ is estimated with a very low precision. Further, its obtained sign is not intuitive. This is because of the low number of time instants in the data file. Therefore, in this case we can not recommend the use of model 1.

	model 1			model 0		
	estimate	inf	sup	estimate	inf	sup
σ_0^2	0.03795	0.03713	0.03876	0.03794	0.03712	0.03876
σ_1^2	0.00481	-0.02789	0.03751	0.00482	-0.02785	0.03749
σ_2^2	0.00100	-0.00282	0.00482	0.00100	-0.00141	0.00341
ρ	-0.54915	-1.89269	0.79438			

Table 9.3.2. Variance parameters and confidence intervals for $\alpha = 0$.

Residuals $\hat{e}_{dt} = \bar{y}_{dt} - \bar{\mathbf{X}}_{dt}\hat{\beta} - \hat{u}_{dt}$ of fitted model 1 are plotted against the observed values $\bar{y}_{dt}$ in the Figure 9.1 for $\alpha = 0$. The dispersion graph shows that EBLUP1 estimates are over and below direct estimates, so that the design unbiased property of the direct estimator is not completely lost by using the model 1. On the right part of the figure we observe that residuals tend to be positive, which means that the model is smoothing the value of the direct estimator larger values. We find that this is an interesting property because it protects us from the presence of outliers in the collection of direct domain estimates.

The two considered estimators of the poverty proportion (EBLUP0 and EBLUP1) are plotted in the Figure 9.2 (left). Their root mean squared error estimates are plotted in the Figure 9.2 (right). We observe that the EBLUP1 is the one presenting the best results and it is thus the one we recommend. Finally full numerical information is presented in the Table B.1 for the poverty proportion. In these tables EBLUP0 and EBLUP1 estimates are labeled by eblup0 and eblup1 respectively.

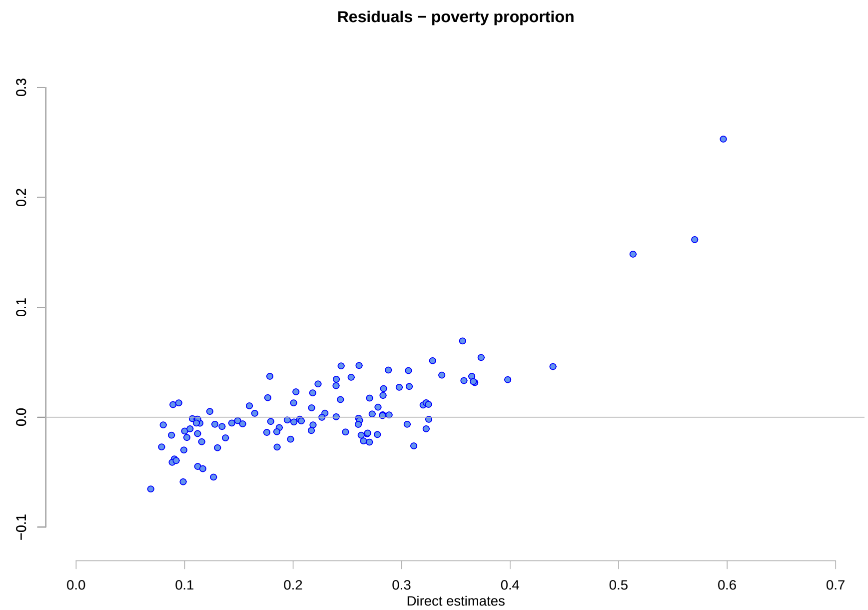


Figure 9.1: Residuals versus direct estimates.

In the Figure 9.3 the Spanish provinces are plotted in 4 colored categories depending on the values of the EBLUP1 estimates in % of the poverty proportions, i.e. $p_d = 100 \cdot \hat{Y}^{eblup1}_{0;d,2006}$. We observe that the Spanish regions where the proportion of the population under the poverty line is smallest are those situated in the north and east, like Cataluña, Aragón, Navarra, País Vasco, Cantabria and Baleares. On the other hand the Spanish regions with higher poverty proportion are those situated in the center-south, like Andalucía, Extremadura, Murcia, Castilla La Mancha, Canarias, Canarias, Ceuta and Melilla. In an intermediate position we can find regions that are in the center-north of Spain, like Galicia, La Rioja, Castilla León, Asturias, Comunidad Valenciana and Madrid. If we investigate how far the annual net incomes of population under the poverty line z_{2006} are from z_{2006}, we observe that in the Spanish regions situated in the center-north there exist a distance that is generally lower than the 6% of z_{2006}. However, the cited distance is in general greater than 6% of z_{2006} in the center-south.

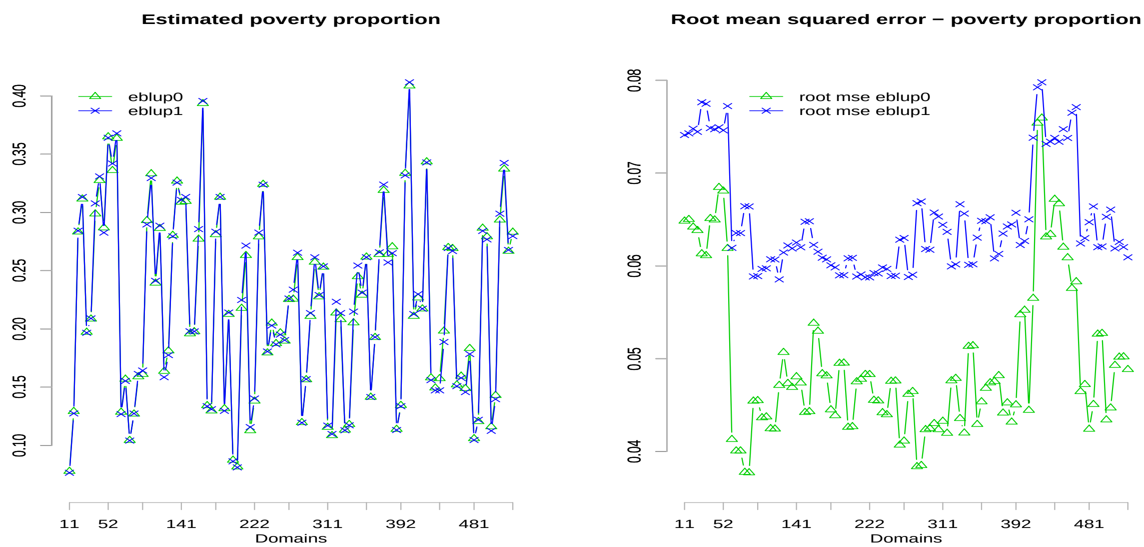
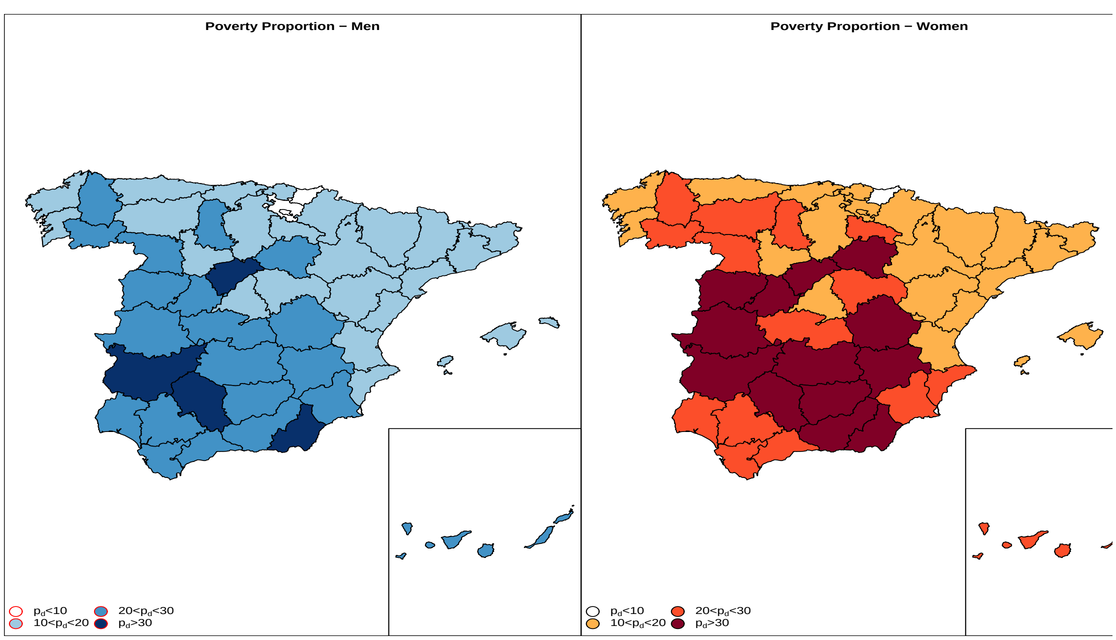


Figure 9.2: Estimates of poverty proportions (left) and squared roots of their estimated MSEs (right).



,

Figure 9.3: Estimates of Spanish poverty proportions for men (left) and women (right).

	men / poverty proportions / women				men / sqrt.mse / women			
d	eblup0	eblup1	eblup0	eblup1	eblup0	eblup1	eblup0	eblup1
1	0.0778	0.0765	0.1291	0.1275	0.0648	0.0741	0.0650	0.0743
2	0.2832	0.2842	0.3115	0.3132	0.0642	0.0748	0.0638	0.0744
3	0.1971	0.1967	0.2085	0.2090	0.0613	0.0776	0.0611	0.0775
4	0.2987	0.3077	0.3274	0.3309	0.0651	0.0748	0.0650	0.0747
5	0.2868	0.2826	0.3649	0.3640	0.0685	0.0749	0.0681	0.0746
6	0.3360	0.3419	0.3636	0.3679	0.0619	0.0772	0.0413	0.0619
7	0.1285	0.1269	0.1569	0.1553	0.0401	0.0635	0.0401	0.0635
8	0.1043	0.1042	0.1269	0.1274	0.0378	0.0664	0.0377	0.0664
9	0.1590	0.1613	0.1610	0.1643	0.0454	0.0589	0.0455	0.0589
10	0.2931	0.2897	0.3332	0.3295	0.0436	0.0597	0.0437	0.0597
11	0.2393	0.2413	0.2863	0.2886	0.0425	0.0607	0.0424	0.0607
12	0.1636	0.1585	0.1812	0.1775	0.0471	0.0585	0.0507	0.0614
13	0.2803	0.2795	0.3269	0.3256	0.0473	0.0622	0.0469	0.0619
14	0.3088	0.3107	0.3095	0.3133	0.0481	0.0625	0.0474	0.0620
15	0.1958	0.1979	0.1976	0.1979	0.0442	0.0648	0.0443	0.0648
16	0.2771	0.2857	0.3934	0.3956	0.0538	0.0623	0.0530	0.0615
17	0.1342	0.1337	0.1296	0.1311	0.0484	0.0609	0.0482	0.0607
18	0.2809	0.2834	0.3131	0.3132	0.0445	0.0601	0.0439	0.0598
19	0.1316	0.1297	0.2125	0.2140	0.0495	0.0590	0.0495	0.0590
20	0.0874	0.0863	0.0816	0.0811	0.0426	0.0608	0.0427	0.0609
21	0.2177	0.2248	0.2631	0.2715	0.0475	0.0589	0.0478	0.0591
22	0.1127	0.1157	0.1381	0.1403	0.0483	0.0588	0.0483	0.0588
23	0.2791	0.2827	0.3241	0.3237	0.0455	0.0592	0.0455	0.0592
24	0.1795	0.1806	0.2050	0.2029	0.0442	0.0598	0.0440	0.0597
25	0.1874	0.1868	0.1966	0.1952	0.0476	0.0589	0.0476	0.0589
26	0.1896	0.1904	0.2253	0.2261	0.0407	0.0628	0.0411	0.0630
27	0.2256	0.2337	0.2613	0.2654	0.0462	0.0588	0.0465	0.0590
28	0.1196	0.1195	0.1564	0.1569	0.0384	0.0668	0.0385	0.0669
29	0.2108	0.2136	0.2573	0.2615	0.0424	0.0618	0.0425	0.0617
30	0.2276	0.2291	0.2530	0.2537	0.0430	0.0657	0.0424	0.0653
31	0.1156	0.1170	0.1084	0.1098	0.0433	0.0644	0.0420	0.0637
32	0.2137	0.2233	0.2080	0.2136	0.0476	0.0600	0.0479	0.0602
33	0.1132	0.1130	0.1180	0.1173	0.0436	0.0666	0.0420	0.0656
34	0.2053	0.2149	0.2449	0.2544	0.0513	0.0601	0.0514	0.0602
35	0.2290	0.2311	0.2617	0.2615	0.0429	0.0630	0.0454	0.0649
36	0.1414	0.1415	0.1927	0.1930	0.0468	0.0648	0.0474	0.0652
37	0.2639	0.2657	0.3189	0.3239	0.0475	0.0608	0.0482	0.0613
38	0.2640	0.2571	0.2705	0.2650	0.0441	0.0635	0.0452	0.0642
39	0.1138	0.1132	0.1344	0.1337	0.0432	0.0645	0.0450	0.0657
40	0.3335	0.3318	0.4086	0.4117	0.0547	0.0623	0.0552	0.0627
41	0.2108	0.2127	0.2266	0.2298	0.0444	0.0650	0.0565	0.0738
42	0.2172	0.2173	0.3434	0.3429	0.0754	0.0792	0.0759	0.0798
43	0.1581	0.1561	0.1494	0.1473	0.0631	0.0731	0.0634	0.0734
44	0.1573	0.1473	0.1982	0.1888	0.0672	0.0738	0.0667	0.0734
45	0.2699	0.2695	0.2691	0.2668	0.0620	0.0747	0.0609	0.0738
46	0.1520	0.1509	0.1595	0.1579	0.0576	0.0765	0.0583	0.0771
47	0.1488	0.1458	0.1832	0.1784	0.0465	0.0624	0.0472	0.0630
48	0.1058	0.1046	0.1205	0.1220	0.0424	0.0647	0.0451	0.0664
49	0.2864	0.2841	0.2791	0.2768	0.0526	0.0620	0.0527	0.0621
50	0.1161	0.1125	0.1429	0.1397	0.0434	0.0653	0.0447	0.0661
51	0.2935	0.2989	0.3372	0.3425	0.0493	0.0619	0.0502	0.0626
52	0.2667	0.2678	0.2830	0.2797	0.0502	0.0620	0.0488	0.0609

Table 9.3.3. Estimated domain poverty proportions and RMSE's by sex.

Chapter 10

Review of M-quantile methods for small area estimation

In this chapter we resume the main theory on the M-quantile models that has been applied to small area estimation of poverty indicators, quantiles and averages in three Italian regions, namely Lombardia, Toscana and Campania. A full description of the methodology used in the application can be found in Deliverable 12 and 16, chapter 10. Here, we devote section 10.1 is devoted to the presentation of the M-quantile model and to the estimation of small area means and quantiles under the M-quantile model. We then focus on the estimation of the Mean Squared Error (MSE) of the small area estimators; in Section 10.2 we move to the estimation of poverty indicators which present a special case of estimating small area quantiles. We consider estimation for two popular poverty indicators namely, the Head Count Ratio (HCR) and the Poverty Gap. In addition, we also consider estimation for fuzzy set indicators that have more recently attracted interest in poverty studies.

In Section 10.3 we recall the fundamental ideas and formulas to be used under a nonparametric M-quantile approach. In fact, when nonparametric models are more appropriate, as when the functional form of the relationship between the response variable and the covariates is unknown or has a complicated functional form, an approach based on penalized splines can offer significant advantages compared with one based on a linear model. Pratesi et al. (2008) and Pratesi et al. (2009) have extended the p-spline regression model to the M-quantile method for the estimation of the small area parameters using a nonparametric specification of the conditional M-quantile of the response variable given the covariates.

Finally we devote section 10.4 to M-quantile geographically based regression. In fact, in some applications the assumption of independence of the small area effects can not be suitable. In other words observations that are spatially close may be more related than observations that are further apart. This spatial correlation can be accounted for by assuming that the regression coefficients vary spatially across the geography of interest. In a recent paper Salvati et al. (2008) proposed an M-quantile Geographically Weighted Regression (GWR) small area model extending the traditional M-quantile regression model by allowing local rather than global parameters to be estimated.

10.1 Linear M-quantile regression models

In the application, for each study variable y, the M-quantile of order q for the conditional density of y given $\mathbf{X}$ is defined as the solution $Q_q(x;\psi)$ of the estimating equation $\int \psi_q(y-Q)f(y|\mathbf{X})dy = 0$, where ψ denotes the influence function associated with the M-quantile.

In a linear M-quantile regression model the q-th M-quantile $Q_q(x,\psi)$ of the conditional distribution of y given $\mathbf{X}$ is such that

$$Q_q(x;\psi) = \mathbf{X}\beta_\psi(q) \tag{10.1}$$

where $\psi_q(r_{iq\psi}) = 2\psi\{s^{-1}r_{iq\psi}\}\ \left\{qI(r_{jq\psi} > 0) + (1-q)I(r_{jq\psi} \leq 0)\right\}$ and s is a suitable robust estimate of scale, e.g. the MAD estimate $s = median\left|r_{jq\psi}\right|/0.6745$. The influence function is specified as in the Huber Proposal 2, $\psi(u) = uI(-c \leq u \leq c) + c\,\mathtt{sgn}(u)$. For specified q and continuous ψ, an estimate $\hat{\beta}_\psi(q)$ of $\beta_\psi(q)$ is obtained via iterative weighted least squares. Note that there is a different set of regression parameters for each q.

10.1.1 Estimation of small area means and quantiles

The small area mean and the cumulative distribution function of a given variable of interest are estimated using M-quantile models under the unified estimation framework for estimating any small area target parameter that was defined by Tzavidis et al. (2010).

Let $\Omega_d = \{1,\ldots,N_d\}$ be the population of area d. Let $\mathbf{y}_d = (y_1,\ldots,y_{N_d})'$ denote the variable values for the N_d small area population elements. We consider a sample $s_d \subset \Omega_d$, of $n_d \leq N_d$ units, and we denote with $r_d = \Omega_d - s_d$ the set of non sampled units. For each population unit j, let $\mathbf{x}_j = (x_{1j},\ldots,x_{pj})$ denote a vector of p known auxiliary variables. The small area specific empirical distribution function of y for area d is

$$F_d = N_d^{-1}\Big[\sum_{j\in s_d}\mathrm{I}(y_j \leqslant t) + \sum_{j\in r_d}\mathrm{I}(y_j \leqslant t)\Big]. \tag{10.2}$$

Given the sample data, the prediction of the values y_j for the non-sampled units in small area d can be done simply replacing the unknown non-sample values of y by their predicted values $\hat{y}_j$ under an appropriate model, leading to a plug-in estimator of (10.2) of the form

$$\hat{F}_d = N_d^{-1}\Big[\sum_{j\in s_d}\mathrm{I}(y_j \leqslant t) + \sum_{j\in r_d}\mathrm{I}(\hat{y}_j \leqslant t)\Big]. \tag{10.3}$$

An estimator of the mean $\overline{Y}_d$ of y in area d is then defined by the value of the mean functional defined by (10.3). This leads to the usual plug-in estimator of the mean,

$$\hat{\overline{Y}}_d = \int_{-\infty}^{\infty} t d\hat{F}_d(t) = N_d^{-1}\left(\sum_{\mathbf{x}\in s_d} y_j + \sum_{j\in r_d}\hat{y}_j\right).$$

The predicted value of a non-sample unit j in area d corresponds to an estimate $\hat{\mu}_j$ of its expected value given that it is located in area d.

Giving that for unit j with values y_j and $\mathbf{x}_j$, the M-quantile coefficient is the value θ_j such that $Q_{\theta_j}(\mathbf{x}_j;\psi) = y_j$. If a hierarchical structure does explain part of the variability in the population data, units within clusters (areas) defined by this hierarchy are expected to have similar M-quantile coefficients.

When the conditional M-quantiles are assumed to follow a linear model, with $\beta_\psi(q)$ a sufficiently smooth function of q, this suggests an estimator of the distribution function:

$$\hat{F}_d^{MQ}(t) = N_d^{-1}\left\{\sum_{j\in s_d} I(y_j \le t) + \sum_{j\in r_d} I(\mathbf{x}_j\hat{\beta}_\psi(\hat{\theta}_d) \le t)\right\} \tag{10.4}$$

where $\mathbf{x}_j\hat{\beta}_\psi(\hat{\theta}_d)$ is used to predict the unobserved value y_j for population unit $j \in r_d$. When there are no sampled observations in area d then $\hat{\theta}_d = 0.5$.

Using the empirical distribution function and the linear M-quantile small area models one can define the estimator of the small area mean as:

$$\hat{\bar{Y}}_d^{MQ}(t) = \int_{-\infty}^{\infty} t d\hat{F}_d^{MQ}(t) = N_d^{-1}\left\{\sum_{j\in s_d} y_j + \sum_{j\in r_d} \mathbf{x}_j\hat{\beta}_\psi(\hat{\theta}_d)\right\}. \tag{10.5}$$

The naive M-quantile mean estimator (10.5) can be biased. A more robust estimator can be derived following Chambers and Dunstan (1986), and their estimator of the finite population distribution function that is

$$\hat{F}_d^{CD}(t) = N_d^{-1}\left\{\sum_{j\in s_d} I(y_j \le t) + \sum_{k\in r_d} n_d^{-1} \sum_{j\in s_d} I(\hat{y}_k + (y_j - \hat{y}_j) \le t)\right\}. \tag{10.6}$$

It can be shown that under the CD estimator of the small area distribution function the mean functional defined by (10.6) takes the value

$$\hat{\bar{Y}}_d^{CD} = \int_{-\infty}^{\infty} t d\hat{F}_d^{CD}(t) = N_d^{-1}\left\{\sum_{j\in s_d} y_j + \sum_{j\in r_d} \hat{y}_j + (f_d^{-1} - 1)\sum_{j\in s_d}(y_j - \hat{y}_j)\right\} \tag{10.7}$$

where $f_d = n_d N_d^{-1}$ is the sampling fraction in area d, $\hat{y}_j = \mathbf{x}_j\hat{\beta}_\psi(\hat{\theta}_d)$ and $\hat{y}_j$ can be obtained under the linear M-quantile small area model. We refer to (10.7) as the bias adjusted M-quantile mean predictor. An useful approach is to limit the variability of the bias correction term in (10.7) by using robust (huberized) residuals instead of raw residuals. In particular,

$$\hat{F}_d^{CDRob}(t) = N_d^{-1}\left\{\sum_{j\in s_d} I(y_j \le t) + \sum_{k\in r_d} n_d^{-1} \sum_{j\in s_d} I\left(\hat{y}_k + \nu_j \psi\{y_j - \hat{y}_j\} \leqslant t\right)\right\} \tag{10.8}$$

where ν_j is a robust estimate of scale for area individual j in area d. For details on properties of (10.6) and (10.8) see Deliverable 12 and 16 Chapter 10.

An alternative to this estimator that is both design-consistent and model-consistent has been proposed by Rao et al. (1990) (hereafter referred to as RKM). Under simple random sampling within the small areas the RKM estimator of the finite population distribution function is

$$\hat{F}_d^{RKM}(t) = n_d^{-1}\left\{\sum_{j\in s_d} I(y_j \le t) + N_d^{-1}\sum_{k\in r_d} n^{-1}\sum_{j\in s_d} I(y_j - \hat{y}_j \le t - \hat{y}_k) \right.$$
$$\left. -(n_d^{-1} - N_d^{-1})\sum_{k\in s_d} n_d^{-1}\sum_{j\in s_d} I(y_j - \hat{y}_j \le t - \hat{y}_k)\right\}. \tag{10.9}$$

Here we just note that the RKM estimator can be used to define an estimator of a small area characteristic that can be represented as a functional of the small area distribution function in exactly the same way as the CD-type estimator (10.7) can be used for this purpose. In general, the resulting estimators will not be the same. An exception is the RKM-based estimator of the area d mean, which is the same as the CD-based estimator of this mean under simple random sampling.

Turning now to the small area quantiles we note that an estimator of the qth quantile of the distribution of y in area d is straightforwardly defined as the solution to the estimating equation

$$\int_{-\infty}^{\hat{\mu}_{qd}} d\hat{F}_d(t) = q, \tag{10.10}$$

where $\hat{F}_d(t)$ is suitable estimator of the area d distribution of y such as the CD or the RKM estimators and $\hat{\mu}_{qd}$ is the estimated qth quantile in small area d. For quantiles other than the median the estimation can be done using either the CD-type specifications or the RKM specification for $\hat{F}_d(t)$, with $\hat{y}_j$ defined by an M-quantile linear small area model.

10.1.2 Mean Squared Error (MSE) estimation for estimators of small area means and quantiles

In the application we use the robust mean squared error estimation method for the naive M-quantile estimator (10.5) proposed by Chambers and Tzavidis (2006). We use also an extension of their argument to define an estimator that is a first order approximation to the mean squared error of the estimator (10.6) of the small area mean when this is based on an M-quantile regression fit. Details are in Chapter 10 of deliverable 12 and 16, here we recall the basic steps of the estimation procedure. Recall that (10.5) can be written

$$\hat{\bar{Y}}_d^{MQ} = \mathbf{w}'_{s_d}\mathbf{y}_s, \tag{10.11}$$

where $\mathbf{w}_{s_d} = (w_{jd}) = n_d^{-1}\Delta_{s_d} + (1 - N_d^{-1}n_d)\mathbf{W}_d\mathbf{X}_s(\mathbf{X}_s'\mathbf{W}_d\mathbf{X}_s)^{-1}\{\bar{\mathbf{x}}_{r_d} - \bar{\mathbf{x}}_{s_d}\}$ with Δ_{sd} denoting the n-vector that "picks out" the sample units from area d. Here $\bar{\mathbf{x}}_{s_d}$ and $\bar{\mathbf{x}}_{r_d}$ denote the sample and non-sample means of x in area d. Also, these weights are 'locally calibrated' on x since

$$\sum_{j\in s} w_{jd}\mathbf{x}_j = \bar{\mathbf{x}}_{s_d} + (1 - f_d)(\bar{\mathbf{x}}_{r_d} - \bar{\mathbf{x}}_{s_d}) = \bar{\mathbf{x}}_d.$$

A first order approximation to the mean squared error of (10.11) then treats the weights as fixed and applies standard methods of robust mean squared error estimation for linear estimators of population quantities (see Royall and Cumberland (1978)). With this approach, the prediction variance of $\hat{\bar{Y}}_d^{CD}$ is estimated by

$$\widehat{Var}(\hat{\bar{Y}}_d^{CD}) = \sum_{g=1}^{d}\sum_{j\in s_g}\lambda_{jdg}\left(y_j - \mathbf{x}_j\hat{\beta}_\psi(\hat{\theta}_g)\right)^2, \tag{10.12}$$

where $\lambda_{jdg} = \{(w_{jd} - 1)^2 + (n_d - 1)^{-1}(N_d - n_d)\}\mathrm{I}(g = d) + w_{jg}^2\mathrm{I}(g \neq d)$. This prediction variance estimator implicitly assumes a model where the regression of y on x varies between areas, and that this variation is consistently estimated by the fit of the M-quantile regression model in each area. Furthermore, since the weights defining $\hat{\bar{Y}}_d^{CD}$ are locally calibrated on x, it immediately follows that (10.6) is unbiased under the same model and hence no correction for its bias is necessary when estimating its mean squared error. This can be compared with the estimator of the mean squared error of the naive M-quantile estimator $\hat{\bar{Y}}_d^{MQ}$ described in Chambers and Tzavidis (2006), which includes a squared bias term.

The linearization-based prediction variance estimator (10.12) is defined only when the estimator of interest can be written as a weighted sum of sample values, as in case of the mean. It cannot be used with quantile estimators defined by solving (10.10). The mean squared error of quantiles has been estimated via a bootstrap procedure.

We define two bootstrap schemes that resample residuals from an M-quantile model fit. The first scheme draws samples from the empirical distribution of suitably recentered residuals. The second scheme draws samples from a smoothed version of this empirical distribution. Using these two schemes, we generate a bootstrap population, from which we then draw bootstrap small area samples. In order to define the bootstrap population, we first calculate the M-quantile small area model residuals $\varepsilon_{jd} = y_{jd} - \hat{\beta}_\psi(\hat{\theta}_d)$.

A bootstrap finite population $U^* = (y^*_{jd}, \mathbf{x}_{jd}), j \in U, d = 1, \cdots, D$ with

$$y^*_{jd} = \mathbf{x}_{jd}\hat{\beta}_\psi(\hat{\theta}_d) + \varepsilon^*_{jd}$$

is then generated, where the bootstrap residuals ε^*_{jd} are obtained by sampling from an estimator of the distribution function $\hat{G}(u)$ of the model residuals ε_{jd}. In order to define $\hat{G}(u)$ we consider two approaches: (i) sampling from the empirical distribution function of the model residuals and (ii) sampling from a smoothed distribution function of these residuals. In each case sampling of the residuals can be done in two ways, (i) by sampling from the distribution of all residuals without conditioning on the small area - we refer to this as the unconditional approach; (ii) by sampling from the conditional distribution of residuals within small area d - we refer to this as the conditional approach.

In either case, bootstrap samples s^*_d are then drawn using simple random sampling within the small areas and without replacement. In what follows we denote by $F_{N,d}(t)$ the unknown true distribution function of the finite population values in area d, by $\hat{F}_d^{CD}(t)$ the CD estimator of $F_{N,d}(t)$ based on sample s_d, by $F^*_{N,d}(t)$ the known true distribution function of the bootstrap population U^*_d in area d, and by $\hat{F}_d^{*,CD}(t)$ the CD estimator of $F^*_{N,d}(t)$ based on bootstrap sample s^*_d. We then estimate the mean squared error of the CD estimator (10.6) as follows. Starting from sample s, selected from a finite population U without replacement, we generate B bootstrap populations, U^{*b}, using one of the four above mentioned methods for estimating the distribution of the residuals. From each bootstrap population, U^{*b}, we select L samples using simple random sampling within the small areas and without replacement in a way such that $n^*_d = n_d$. Finally, bootstrap estimators of the bias and variance of the CD estimator of the distribution function in area j are defined respectively by

$$\widehat{Bias}_d = B^{-1}L^{-1}\sum_{b=1}^{B}\sum_{l=1}^{L}\left(\hat{F}_d^{bl,CD}(t) - F^{*b}_{N,d}(t)\right)$$

and

$$\widehat{Var}_d = B^{-1}L^{-1}\sum_{b=1}^{B}\sum_{l=1}^{L}\left(\hat{F}_d^{*bl,CD}(t) - \hat{\bar{F}}_d^{*bl,CD}(t)\right)^2,$$

where

$$\hat{\bar{F}}_d^{*bl,CD}(t) = L^{-1}\sum \hat{F}_d^{*bl,CD}(t)$$

is the distribution function of the bth bootstrap population and $\hat{F}_d^{*bl,CD}(t)$ is the CD estimator of $F_{N,d}^{*,b}(t)$ computed from the lth sample of the bth bootstrap population, $(b = 1,\cdots,B, l = 1,\cdots,L)$. The bootstrap estimator of the mean squared error of the CD-based small area estimate is finally calculated as

$$\widehat{MSE}_d\left(\hat{F}_d^{CD}(t)\right) = \widehat{Var}_d + \widehat{Bias}_d^2. \tag{10.13}$$

The above bootstrap procedure can also be used to construct confidence intervals for the value of $F_{N,d}(t)$ by "reading off" appropriate quantiles of the bootstrap distribution of $F_d^{CD}(t)$.

10.2 Small Area Models for Poverty Estimation

The M-quantile model can be also utilized for estimating more complex statistics such as small area poverty indicators (Foster et al., 1984). In this section we review the approach to poverty estimation based on the M-quantile small area model which does not impose strong distributional assumptions and is outlier robust. Hence, the use of the M-quantile model for poverty estimation may protect us against departures from assumptions of the Molina and Rao (2009) approach, namely the Empirical Best Prediction (EBP) approach, which is based on the nested error regression model. A full description of these methods can be find in the deliverable 12 and 16, Chapter 2 and Chapter 10 section 2 respectively for the EBP approach and the M-quantile approach.

10.2.1 Definitions of poverty indicators

In this section we look over the estimation of two poverty indicators i.e. the incidence of poverty or *Head Count Ratio* F_0 and the *Poverty Gap* F_1 (see Foster et al. (1984)). Denoting by t the poverty line, the FGT poverty measures for a small area d are defined as

$$F_{\alpha d} = (\frac{t - y_{jd}}{t})^{\alpha}\mathrm{I}(y_{jd} \leqslant t). \tag{10.14}$$

Setting $\alpha = 0$ defines the *Head Count Ratio* whereas setting $\alpha = 1$ defines the *Poverty Gap*.

10.2.2 The M-quantile approach for poverty estimation

In this section we discuss estimation of the poverty indicators of interest under the M-quantile model. To start with, the target is to estimate F_{ad} using the M-quantile small area model

$$F_{\alpha d} = N_d^{-1}\Big[\sum_{j\in s_d} F_{\alpha d} + \sum_{j\in r_d} F_{\alpha d}\Big], \tag{10.15}$$

To estimate the (10.15) we have to predict the out of sample component. This can be achieved using the ideas we described in Section 10.1.1 for estimating the small area distribution function under the M-quantile small area model. The head count ratio estimation is a special case of quantile estimation since we are interested in estimating the number of individuals/households below a threshold. As a result one approach to estimating $F_{\alpha=0,d}$ is by using a smearing-type estimator of the distribution function such as the Chambers-Dunstan estimator. In this case, an estimator $\hat{F}^{MQ}_{\alpha=0,d}$ of $F^{MQ}_{\alpha=0,d}$ is

$$\hat{F}_{ad} = N_d^{-1}\left\{\sum_{j\in s_d} I(y_j \leq t) + \sum_{k\in r_d} n_d^{-1} \sum_{j\in s_d} I(\hat{y}_k + (y_j - \hat{y}_j) \leq t)\right\}$$

The above can be evaluated using the following procedure.

1 Fit the M-quantile small area model (10.1) using the raw $\mathbf{y}_s$ sample values and obtain estimates of β and q_d;

2 draw an out of sample vector using

$$y^*_{jdr} = \mathbf{x}_{jdr}\hat{\beta}(\hat{\theta}_d) + e^*_{jdr},$$

where e^*_{jdr} is a vector of size $N_d - n_d$ drawn from the Empirical Distribution Function (EDF) of the estimated M-quantile regression residuals or from a smooth version of this distribution and $\hat{\beta}$, $\hat{\theta}_d$ are obtained from the previous step;

3 repeat the process H times. Each time combine the sample data and out of sample data for estimating the target using

$$\hat{F}^{MQ}_{ad} = N_d^{-1}\Big[\sum_{j\in s_d} \mathrm{I}(y_j \leqslant t) + \sum_{j\in r_d} \mathrm{I}(y^*_j \leqslant t)\Big];$$

4 average the results over H Monte Carlo replications.

With the same procedure we can also estimate the poverty gap indicator. Moreover, different approaches can be used for drawing e^*_{jdr}. One can draw conditional (upon the small area) or unconditional residuals from the EDF or from a smoothed version of the EDF. A mean squared error of the M-quantile estimates of the incidence of poverty can be obtained using the non-parametric bootstrap approach described in Tzavidis et al. (2010).

10.2.3 Alternative measures for poverty: fuzzy indicators at a small area level with M-quantile models

In this section we review estimators for the fuzzy monetary and fuzzy supplementary indicators at a small area level based on the M-quantile model. Fuzzy approach considers poverty as a matter of degree rather than an attribute that is simply present or absent for individuals in the population of Betti et al.(2009).

Fuzzy indicators have been made such that they vary from 0 to 1, where 0 indicates the richest person in the population while 1 indicates the poorest. They belong to the class of generalized Gini inequality indicators.

The fuzzy monetary indicator is based on the equivalised income, E. For small area d is defined as:

$$FM_d = N_d^{-1}\sum_{j=1}^{N_d} FM_j$$

where FM_j is the fuzzy monetary index for the jth unit in the population:

$$FM_j = \left[(N_d - 1)^{-1} \sum_{k=1}^{N_d} I(E_k > E_j) \right]^{\alpha - 1} \left[\frac{\sum_{k=1}^{N_d} E_k I(E_k > E_j)}{\sum_{k=1}^{N_d} E_k} \right].$$

The parameter α is arbitrary, but Cheli and Betti (1999) have chosen α so that the mean of FM_j for the whole population is equal to the head count ratio computed for the official poverty line.

An estimator of FM_d under the M-quantile model is given by

$$\widehat{FM}_d^{MQ} = N_d^{-1} \left[\sum_{j \in s_d} FM_j + \sum_{j \in r_d} \widehat{FM}_j^{MQ} \right]. \tag{10.16}$$

Under the M-quantile model an empirical approach for estimating equation (10.16) is implemented using the same Monte-Carlo approximation described in section 10.2.2.

1. Fit the M-quantile model (10.1) using the raw E sample values and obtain estimates of $\beta(\theta_d)$;

2. draw an out of sample vector using

   $$E_j^* = \mathbf{x}_j \hat{\beta}(\hat{\theta}_d) + e_j^* \quad j \in r_d,$$

   where $e_j^*, j \in r_d$ is drawn from the EDF of the estimated M-quantile regression residuals and $\hat{\beta}(\hat{\theta}_d)$ is obtained from the previous step;

3. repeat the process H times. Each time combine the sample data and the out of sample data for estimating the target using

   $$\widehat{FM}_d^{MQ} = N_d^{-1} \left[\sum_{j \in s_d} FM_j + \sum_{j \in r_d} \widehat{FM}_j^{MQ} \right],$$

   where $\widehat{FM}_j^{MQ}$ is estimated using the observed and the predicted equivalised incomes $E_d = \{E_j, j \in s_d \cup E_j^*, j \in r_d\}$.

4. average the results over L replications.

We remand to the final WP1 deliverable for a deepened discussion about fuzzy sets. The used approach is similar in spirit to the EBP approach proposed by Molina and Rao (2009).

10.3 Nonparametric M-quantile regression models in small area estimation

M-quantile models do not depend on strong distributional assumptions, but they assume that the quantiles of the distribution are some known parametric function of the covariates. When the functional form of the relationship between the q-th M-quantile and the covariates deviates from the assumed one, the traditional M-quantile regression can lead to biased estimates of the β coefficients. Pratesi et al. (2008)

and Salvati et al. (2010b) have extended this approach to the M-quantile method for the estimation of the small area parameters using a nonparametric specification of the conditional M-quantile of the response variable given the covariates.

Let us consider only smoothing with one covariate x_1, a nonparametric model for the qth quantile can be written as $Q_q(x_1,\psi) = \tilde{m}_{\psi,q}(x_1)$, where the function $\tilde{m}_{\psi,q}(\cdot)$ is unknown and, in the smoothing context, usually assumed to be continuous and differentiable. Here, we will assume that it can be approximated sufficiently well by the following function

$$m_{\psi,q}[x_1;\beta_\psi(q),\gamma_\psi(q)] = \beta_{0\psi}(q)+\beta_{1\psi}(q)x_1+\ldots+\beta_{p\psi}(q)x_1^p+\sum_{k=1}^{K}\gamma_{k\psi}(q)(x_1-\kappa_k)_+^p, \qquad (10.17)$$

where p is the degree of the spline, $(t)_+^p = t^p$ if $t>0$ and 0 otherwise, κ_k for $k=1,\ldots,K$ is a set of fixed knots, $\beta_\psi(q) = (\beta_{0\psi}(q),\beta_{1\psi}(q),\ldots,\beta_{p\psi}(q))^t$ is the coefficient vector of the parametric portion of the model and $\beta\gamma_\psi(q) = (\gamma_{1\psi}(q),\ldots,\gamma_{K\psi}(q))^t$ is the coefficient vector for the spline one. The latter portion of the model allows for handling nonlinearities in the structure of the relationship. The spline model (10.17) uses a truncated polynomial spline basis to approximate the function $\tilde{m}_{\psi,q}(\cdot)$.

The influence of the knots is limited by putting a constraint on the size of the spline coefficients: typically $\sum_{k=1}^{K}\gamma_{k\psi}^2(q)$ is bounded by some constant, while the parametric coefficients $\beta_\psi(q)$ are left unconstrained. Therefore estimation can be accommodated by mimicking penalization of an objective function and solving the following set of estimating equations

$$\sum_{j=1}^{n}\psi_q(y_j-\mathbf{x}_j\beta_\psi(q)-\mathbf{z}_j\gamma_\psi(q))(\mathbf{x}_j,\mathbf{z}_j)\text{trace}+\lambda\begin{bmatrix}\mathbf{0}_{(1+p)}\\ \gamma_\psi(q)\end{bmatrix} = \mathbf{0}_{(1+p+K)}, \qquad (10.18)$$

assuming that

$$\psi_q(r_{jq\psi}) = 2\psi\{s^{-1}r_{jq\psi}\}\{(1-q)I(r_{jq\psi}\le 0)+qI(r_{jq\psi}>0)\}$$

where $r_{jq\psi} = y_j-\mathbf{x}_j\beta_\psi(q)-\mathbf{z}_j\gamma_\psi(q)$, s is a robust estimate of scale, e.g. the MAD estimate $s = \text{median}|r_{jq\psi}|/0.6745$, $\mathbf{x}_j$ here is the j-th row of the $n\times(1+p)$ matrix

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{11}^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{1n}^p \end{bmatrix},$$

while $\mathbf{z}_j$ is the j-th row of the $n\times K$ matrix

$$\mathbf{Z} = \begin{bmatrix} (x_{11}-\kappa_1)_+^p & \cdots & (x_{11}-\kappa_K)_+^p \\ \vdots & \ddots & \vdots \\ (x_{1n}-\kappa_1)_+^p & \cdots & (x_{1n}-\kappa_K)_+^p \end{bmatrix},$$

and λ is a Lagrange multiplier that controls the level of smoothness of the resulting fit. An iterative solution is needed here to obtain estimates $\hat{\beta}_\psi(q)$ and $\hat{\gamma}_\psi(q)$. Consider the Huber proposal 2 influence function (see Huber(1981)) an algorithm based on iteratively reweighted penalized least squares is proposed in Pratesi et al. (2009) to effectively compute the parameter estimates.

Extension to bivariate smoothing can be handled by assuming $Q_q(x_1,x_2,\psi) = \tilde{m}_{\psi,q}(x_1,x_2)$. This is of central interest in a number of application areas when referenced responses need to be converted to maps, as in environment and poverty mapping. In particular, the following model is assumed at quantile q for unit i:

$$m_{\psi,q}[x_{1j}, x_{2j}; \beta_\psi(q), \gamma_\psi(q)] = \beta_{0\psi}(q) + \beta_{1\psi}(q)x_{1j} + \beta_{2\psi}(q)x_{2j} + \mathbf{z}_i\gamma_\psi(q). \tag{10.19}$$

Here $\mathbf{z}_j$ is the j-th row of the following $n \times K$ matrix

$$\mathbf{Z} = [C(\tilde{\mathbf{x}}_j - \mathbf{k}_k)]_{\substack{1\leqslant j\leqslant n\\ 1\leqslant k\leqslant K}} [C(\mathbf{k}_k - \mathbf{k}_{k'})]^{-1/2}_{1\leqslant k,k'\leqslant K}, \tag{10.20}$$

where $C(\beta t) = ||\beta t||^2 \log ||\beta t||$, $\tilde{\mathbf{x}}_j = (x_{1j}, x_{2j})$ and $\mathbf{k}_k$, $k = 1, \ldots, K$ are knots. See Pratesi et al. (2009) for details on this. Here, it is enough to note that the estimation procedure can again be pursued with (10.18) where $\mathbf{x}_j = (1, \tilde{\mathbf{x}}_j)$.

10.3.1 Small area estimator of the mean and of the quantiles

Salvati et al. (2010b) have applied the P-splines M-quantile regression to the estimation of a small area mean as follows. The first step is to estimate the M-quantile coefficients q_{jd} as illustrated in paragraph 10.1 for the linear case treated in Chambers and Tzavidis (2006). Recall that the M-quantile coefficient q_{jd} of unit j in area d is the value q_{jd} such that $Q_{q_{jd}}(x_{1jd}, \psi) = y_{jd}$. The unit level coefficients are estimated by defining a fine grid of values on the interval $(0,1)$ and using the sample data to fit the p-splines M-quantile regression functions at each value q on this grid. If a data point lies exactly on the q-th fitted curve, then the coefficient of the corresponding sample unit is equal to q. Otherwise, to obtain q_{jd}, a linear interpolation over the grid is used. An estimate of the mean quantile for area d is obtained by taking the corresponding average value of the sample M-quantile coefficient of each unit in area d. The small area estimator of the mean may be taken as:

$$\hat{\bar{Y}}_d = \frac{1}{N_d}\Big\{\sum_{j\in s_d} y_{jd} + \sum_{j\in r_d} \hat{y}_{jd}\Big\}, \tag{10.21}$$

where the unobserved value for population unit $j \in r_d$ is predicted using

$$\hat{y}_{jd} = \mathbf{x}_{jd}\hat{\beta}_\psi(\hat{\theta}_d) + \mathbf{z}_{jd}\hat{\gamma}_\psi(\hat{\theta}_d),$$

where $\hat{\beta}_\psi(\hat{\theta}_d)$ and $\hat{\gamma}_\psi(\hat{\theta}_d)$ are the coefficient vectors of the parametric and spline portion, respectively, of the fitted p-splines M-quantile regression function at $\hat{\theta}_d$.

However, the estimator of the small area mean can be biased for small areas containing outliers. The bias-adjusted estimator for the mean is given by

$$\hat{\bar{Y}}_d^{NPMQ/CD} = \frac{1}{N_d}\Big\{\sum_{j\in U_d} \hat{y}_{jd} + \frac{N_d}{n_d}\sum_{j\in s_d} (y_{jd} - \hat{y}_{jd})\Big\}, \tag{10.22}$$

where $\hat{y}_{jd}$ denotes the predicted values for the population units in s_d and in U_d. Due to the bias correction in (10.22), this predictor will have higher variability and so should only be used when the estimator (10.21) is expected to have substantial bias, e.g. when there are large outlying data points. An alternative approach to dealing with the bias-variance trade off in (10.22) in such a situation is to limit the variability of the bias correction term in (10.22) by using robust (huberized) residuals instead of raw residuals. In particular,

$$\hat{\hat{Y}}_d^{NPMQ/Rob} = \frac{1}{N_d}\Big\{\sum_{j\in s_d} y_{jd} + \sum_{j\in r_d}\hat{y}_{jd} + \frac{N_d - n_d}{n_d}\sum_{j\in s_d}\nu_d\psi\Big(\frac{y_{jd}-\hat{y}_{jd}}{\nu_d}\Big)\Big\} \tag{10.23}$$

where ν_d is a robust estimate of scale for area d (see Tzavidis and Chambers (2007)).

Using the nonparametric M-quantile predictor for the non sampled units we can define a model unbiased estimator of the small area distribution function (10.6):

$$\hat{F}_d^{NPQM/CD}(t) = N_d^{-1}\left\{\sum_{j\in s_d} I(y_j \le t) + \sum_{k\in r_d} n_d^{-1}\sum_{j\in s_d} I(\mathbf{x}_{kd}\hat{\beta}_\psi(\hat{\theta}_d) - \mathbf{z}_{jd}\hat{\gamma}_\psi(\hat{\theta}_d) + (y_j - \mathbf{x}_{jd}\hat{\beta}_\psi(\hat{\theta}_d) + \mathbf{z}_{jd}\hat{\gamma}_\psi(\hat{\theta}_d)) \le t)\right\}. \tag{10.24}$$

Similarly to M-quantile small area models, the qth quantile $\hat{\mu}_{qd}$ of the distribution of y in area d is straightforwardly estimated by the solution to

$$\int_{-\infty}^{\hat{\mu}_{qd}} d\hat{F}_d^{NPQM}(t) = q. \tag{10.25}$$

10.3.2 Mean squared error estimation

Salvati et al. (2010b) also propose an estimator of the MSE of the small area mean. For fixed q and λ, the $\hat{\hat{Y}}_j$ in (10.23) can be written as the following linear combination of the observed y_{jd} plus an additional part due to the huberized residuals. In particular,

$$\hat{\hat{Y}}_d^{NPMQ/Rob} = \frac{1}{N_d}\sum_{j\in s} w_{jd}y_{jd}, \tag{10.26}$$

where the weights $\mathbf{w}_d = (w_{1d},\dots,w_{nd})^T$ are given by

$$\mathbf{w}_d = \Big\{1 + \frac{N_d - n_d}{n_d}b_{jd}\Big\}\mathbf{1}_{s_d} +$$

$$+\mathbf{W}(\hat{\theta}_d)[\mathbf{X},\mathbf{Z}]\Big([\mathbf{X},\mathbf{Z}]\text{trace}\mathbf{W}(\hat{\theta}_d)[\mathbf{X},\mathbf{Z}] + \lambda\mathbf{G}\Big)^{-1}\left(\mathbf{T}_{r_d} - \frac{N_d - n_d}{n_d}\mathbf{T}_{s_d}\right) \tag{10.27}$$

with $b_{jd} = \psi\Big(\frac{y_{jd}-\hat{y}_{jd}}{\nu_d}\Big)/\Big(\frac{y_{jd}-\hat{y}_{jd}}{\nu_d}\Big)$, $\mathbf{1}_{s_d}$ the n-vector with j^{th} component equal to one whenever the corresponding sample unit is in area j and to zero otherwise, $\mathbf{W}(\hat{\theta}_d)$ a diagonal $n\times n$ matrix that contains the final set of weights produced by the iteratively reweighted penalized least squares algorithm used to estimate the regression coefficients, $\mathbf{G} = \text{diag}\{\mathbf{0}_{1+p}, \mathbf{1}_K\}$ with $1+p$ the number of columns of $\mathbf{X}$ and K the number of columns of $\mathbf{Z}$, and with $\mathbf{T}_{r_d}$ and $\mathbf{T}_{s_d}$ the totals of the covariates for the non-sampled and the sampled units in area d, respectively. Note that $\mathbf{T}_{s_d} = \sum_{j\in s_d}[\mathbf{x}_{jd}\ \mathbf{z}_{jd}]^T b_{jd}$.

The weights derived from (10.27) are treated as fixed and a "plug in" estimator of the mean squared error of estimator (10.26) can be proposed by using standard methods for robust estimation of the variance of unbiased weighted linear estimators (see Royall and Cumberland (1978)) and by following the results due to Chambers and Tzavidis (2006). The prediction variance of (10.26) can be approximated by

$$\mathrm{var}(\hat{\bar{Y}}_d^{NPMQ/Rob} - \bar{Y}_d) \approx \frac{1}{N_d^2}\Big[\sum_{j\in s_d}\Big\{d_{jd}^2 + \frac{N_d - n_d}{n_d - 1}\Big\}\mathrm{var}(y_{jd}) + \sum_{j\in s\backslash s_d} b_{jd}^2 \mathrm{var}(y_{jd})\Big] \qquad (10.28)$$

with $b_{jd} = w_{jd} - 1$ if $j \in s_d$ and $b_{jd} = w_{jd}$ otherwise, and $s\backslash s_d$ the set of sampled units outside area d. Following the area level residual approach of Tzavidis and Chambers (2006), we can interpret $\mathrm{var}(y_{jd})$ conditionally to the specific area d from which y_d is drawn and hence replace $\mathrm{var}(y_{jd})$ in (10.28) by $(y_{jd} - \hat{y}_{jd})^2$. Salvati et al. (2010b) develop a robust estimator of the mean squared error of (10.26) that is given by

$$\widehat{\mathrm{var}}(\hat{\bar{Y}}_d^{NPMQ/Rob}) = \frac{1}{N_d^2}\left[\sum_{j\in s_d}\left\{b_{jd}^2 + \frac{N_d - n_d}{n_d - 1}\right\}(y_{jd} - \hat{y}_{jd})^2 + \sum_{j\in s\backslash s_d} b_{jd}^2 (y_{jd} - \hat{y}_{jd})^2\right]. \qquad (10.29)$$

Since the bias-adjusted nonparametric M-quantile estimator is an approximately unbiased estimator of the small area mean, the squared bias will not impact significantly the mean squared error estimator. The main limitation of the MSE estimator is that it does not account for the variability introduced in estimating the area specific q's and λ. We note also that we can obtain an estimate only for areas where there are at least two sampled units. Details on the property of the MSE estimator can be found in Tzavidis et al. (2010) and Salvati et al. (2010a).

10.4 M-quantile GWR models

An approach to incorporate the spatial information in the regression model is by assuming that the regression coefficients vary spatially across the geography of interest. Geographically Weighted Regression (GWR) (see Brundson et al. (1996)) extends the traditional regression model by allowing local rather than global parameters to be estimated. In a recent paper Salvati et al. (2008) proposed an M-quantile GWR small area model. The authors proposed an extension to the GWR model, the M-quantile GWR model, i.e. a locally robust model for the M-quantiles of the conditional distribution of the outcome variable given the covariates. Here we report a brief description of the M-quantile GWR model.

10.4.1 M-quantile geographically weighted regression

In this Section we define a spatial extension to linear M-quantile regression based on GWR. Since M-quantile models do not depend on how areas are specified, we also drop the area subscript d from our notation in this Section.

Given n observations at a set of L locations $\{u_l; l = 1,\dots,L; L \leqslant n\}$ with n_l data values $\{(y_{jl}, \mathbf{x}_{jl}); i = 1,\dots,n_l\}$ observed at location u_l, a linear GWR model is a special case of a locally linear approximation to a spatially non-linear regression model and is defined as follows

$$y_{jl} = \mathbf{x}_{jl}^T \beta(u_l) + \varepsilon_{jl}, \qquad (10.30)$$

where $\beta(u_l)$ is a vector of p regression parameters that are specific to the location u_l and the ε_{il} are independently and identically distributed random errors with zero expected value and finite variance. The value of the regression parameter 'function' $\beta(u)$ at an arbitrary location u is estimated using weighted least squares

$$\hat{\beta}(u) = \Big\{ \sum_{l=1}^{L} w(u_l, u) \sum_{i=1}^{n_l} \mathbf{x}_{jl}\mathbf{x}_{jl}^T \Big\}^{-1} \Big\{ \sum_{l=1}^{L} w(u_l, u) \sum_{i=1}^{n_l} \mathbf{x}_{jl} y_{jl} \Big\},$$

where $w(u_l, u)$ is a spatial weighting function whose value depends on the distance from sample location u_l to u in the sense that sample observations with locations close to u receive more weight than those further away. In this application we use a Gaussian specification for this weighting function

$$w(u_l, u) = \exp\Big\{ -d_{u_l,u}^2 / 2b^2 \Big\}, \tag{10.31}$$

where $d_{u_l,u}$ denotes the Euclidean distance between u_l and u and $b > 0$ is the bandwidth. As the distance between u_l and u increases the spatial weight decreases exponentially. A global bandwidth is defined by minimising the cross-validation criterion proposed by Fotheringham et al. (2002):

$$CV = \sum_{l=1}^{L} \sum_{j=1}^{n_l} [y_{jl} - \hat{y}_{(j)l}(b)]^2,$$

where $\hat{y}_{(j)l}(b)$ is the predicted value of y_{jl}, using bandwidth b, with the observation y_{jl} omitted from the model fitting process. The value of b that minimises CV is then selected.

The GWR model (10.30) is a linear model for the conditional expectation of y given $\mathbf{X}$ at location u. That is, this model characterises the local behaviour of the conditional expectation of y given $\mathbf{X}$ as a linear function of $\mathbf{X}$. However, a more complete picture of the relationship between y and $\mathbf{X}$ at location u can be constructed by specifying a model for the conditional distribution of y given $\mathbf{X}$ at this location. Since the M-quantiles serve to characterise this conditional distribution, such a model can be defined by extending

$$Q_q(\mathbf{x}_j; \psi) = \mathbf{x}_j^T \beta_\psi(q). \tag{10.32}$$

to specify a linear model for the M-quantile of order q of the conditional distribution of y given $\mathbf{X}$ at location u, writing

$$Q_q(\mathbf{x}_{jl}; \psi, u) = \mathbf{x}_{jl}^T \beta_\psi(u; q), \tag{10.33}$$

where now $\beta_\psi(u;q)$ varies with u as well as with q. Like (10.30), we can interpret (10.33) as a local linear approximation, in this case to the (typically) non-linear order q M-quantile regression function of y on $\mathbf{X}$, thus allowing the entire conditional distribution (not just the mean) of y given $\mathbf{X}$ to vary from location to location. The parameter $\beta_\psi(u;q)$ in (10.33) at an arbitrary location u can be estimated by solving

$$\sum_{l=1}^{L} w(u_l, u) \sum_{j=1}^{n_l} \psi_q\{y_{jl} - \mathbf{x}_{jl}^T \beta_\psi(u;q)\} \mathbf{x}_{jl} = \mathbf{0}, \tag{10.34}$$

where $\psi_q(\varepsilon) = 2\psi(s^{-1}\varepsilon)\{qI(\varepsilon > 0) + (1-q)I(\varepsilon \leqslant 0)\}$, s is a suitable robust estimate of the scale of the residuals $y_{jl} - \mathbf{x}_{jl}^T \beta_\psi(u;q)$, e.g. $s = median|y_{jl} - \mathbf{x}_{jl}^T \beta_\psi(u;q)|/0.6745$, and we will typically assume a Huber Proposal 2 influence function, $\psi(\varepsilon) = \varepsilon I(-c \leqslant \varepsilon \leqslant c) + sgn(\varepsilon) I(|\varepsilon| > c)$. Provided c is bounded away from zero, we can solve (10.34) by combining the iteratively re-weighted least squares algorithm used to fit the 'spatially stationary' M-quantile model (10.32) and the weighted least squares algorithm used to fit a GWR model. The weighted least squares estimates of $\beta_{\psi}(u;q)$

$$\hat{\beta}_{\psi}(u;q) = \left\{\mathbf{X}^T W^*(u;q)\mathbf{X}\right\}^{-1}\mathbf{X}^T W^*(u;q)\mathbf{y}, \tag{10.35}$$

where $\mathbf{y}$ is the vector of n sample values and $\mathbf{X}$ is the corresponding matrix of order $n \times p$ of sample x values. The matrix $W^*(u;q)$ is a diagonal matrix of order n with entry, corresponding to a particular sample observation, set equal to the product of this observation's spatial weight, which depends on its distance from location u, and the weight that this observation has when the sample data are used to calculate the 'spatially stationary' M-quantile estimate $\hat{\beta}_{\psi}(q)$. An R function that implements an iterative re-weighted least squares algorithm for for fitting (10.33) is available from Salvati and Tzavidis (2010).

The fitted regression surface $\hat{Q}_q(\mathbf{x}_{jl};\psi,u) = \mathbf{x}_{jl}^T\hat{\beta}_{\psi}(u;q)$ then defines the fit of the M-quantile GWR model for the regression M-quantile of order q of y given $\mathbf{X}$ at location u.

One may argue that (10.33) is over-parameterised as it allows for both local intercepts and local slopes. An alternative spatial extension of the M-quantile regression model (10.32) that has a smaller number of parameters is one that combines local intercepts with global slopes and is defined as

$$Q_q(\mathbf{x}_{jl};\psi,u) = \mathbf{x}_{jl}^T\beta_{\psi}(q) + \delta_{\psi}(u;q), \tag{10.36}$$

where $\delta_{\psi}(u;q)$ is a real valued spatial process with zero mean function over the space defined by locations of interest. Model (10.36) is fitted in two steps. At the first step we ignore the spatial structure in the data and estimate $\beta_{\psi}(q)$ directly via the iterative re-weighted least squares algorithm used to fit the standard linear M-quantile regression model (10.32). Denote this estimate by $\hat{\beta}_{\psi}(q)$. At the second step we use geographic weighting to estimate $\delta_{\psi}(u;q)$ via

$$\hat{\delta}_{\psi}(u;q) = n^{-1}\sum_{l=1}^{L} w(u_l,u)\sum_{j=1}^{n_l}\psi_q\{y_{jl} - \mathbf{x}_{jl}^T\hat{\beta}_{\psi}(q)\}. \tag{10.37}$$

Choosing between (10.33) and (10.36) will depend on the particular situation and whether it is reasonable to believe that the slope coefficients in the M-quantile regression model vary significantly between locations.

10.4.2 Using M-quantile GWR models in small area estimation

SAR models allow for spatial correlation in the model error structure to be used to improve SAE. Alternatively, this spatial information can be incorporated directly into the model regression structure via an M-quantile GWR model for the same purpose. In this Section we describe how this can be achieved. We now assume that we have only one population value per location, allowing us to drop the index l. We also assume that the geographical coordinates of every unit in the population are known, which is the case with geo-coded data. The aim is to use these data to predict the area d mean of y using the M-quantile GWR models (10.33) and (10.36).

Following Chambers and Tzavidis (2006), we first estimate the M-quantile GWR coefficients $q_j; j \in s$ of the sampled population units without reference to the small areas of interest. A grid-based interpolation procedure for doing this under (10.32) is described by Chambers and Tzavidis (2006) and can be used directly with (10.36). We adapt this approach to the GWR M-quantile model (10.33) by first defining a fine grid of q values in the interval $(0,1)$. Chambers and Tzavidis (2006) use a grid that ranges between 0.01 and 0.99 with step 0.01. We employ the same grid definition and then use the sample data to fit (10.33) for each distinct value of q on this grid and at each sample location. The M-quantile GWR coefficient for unit j with values y_j and $\mathbf{x}_j$ at location u_j is finally calculated by using linear interpolation over this grid to find the unique value q_j such that $\hat{Q}_{q_j}(\mathbf{x}_j;\psi,u_j) \approx y_j$.

Provided there are sample observations in area d, an area d specific M-quantile GWR coefficient, $\hat{\theta}_d$ can be defined as the average value of the sample M-quantile GWR coefficients in area d, otherwise we set $\hat{\theta}_d = 0.5$. Following Tzavidis et al. (2010), the bias-adjusted M-quantile GWR predictor of the mean $\bar{Y}_d$ in small area d is then

$$\hat{\bar{Y}}_d^{MQGWR/CD} = N_d^{-1}\Big[\sum_{j\in U_d}\hat{Q}_{\hat{\theta}_d}(\mathbf{x}_j;\psi,u_j) + \frac{N_d}{n_d}\sum_{j\in s_d}\{y_j - \hat{Q}_{\hat{\theta}_d}(\mathbf{x}_j;\psi,u_j)\}\Big], \qquad (10.38)$$

where $\hat{Q}_{\hat{\theta}_d}(\mathbf{x}_j;\psi,u_j)$ is defined either via the MQGWR model (10.33) or via the MQGWR-LI model (10.36).

10.4.3 Mean squared error estimation

A "pseudo-linearization" MSE estimator for M-quantile small area estimators was recommended by Chambers and Tzavidis (2006) and it has now been used successfully in empirical studies reported in a number of published papers on SAE, including the recent publications by Tzavidis et al. (2010) and Salvati et al. (2010a). Below we extend the argument of these papers to defining an estimator of a first order approximation to the mean squared error of (10.38). This extension is based on (i) a model where the regression of y on $\mathbf{X}$ for a particular population unit depends on its location, with this regression specified by the locally linear GWR model (10.30), and (ii) the fact that estimators derived under the MQGWR model (10.33) or the MQGWR-LI model (10.36) can be written as linear combinations of the sample values of y. For example, from (10.35) we see that (10.38) can be expressed as a weighted sum of the sample y-values

$$\hat{\bar{Y}}_d^{MQGWR/CD} = N_d^{-1} w_d^T \mathbf{y}, \qquad (10.39)$$

where

$$w_d = \frac{N_d}{n_d}\mathbf{1}_d + \sum_{j\in r_d}\mathbf{H}_{jd}^T\mathbf{x}_j - \frac{N_d - n_d}{n_d}\sum_{j\in s_d}\mathbf{H}_{jd}^T\mathbf{x}_j. \qquad (10.40)$$

Here $\mathbf{1}_d$ is the n-vector with j-th component equal to one whenever the corresponding sample unit is in area d and is zero otherwise and

$$\mathbf{H}_{jd} = \Big\{\mathbf{X}^T W^*(u_j;\hat{\theta}_d)\mathbf{X}\Big\}^{-1}\mathbf{X}^T W^*(u_j;\hat{\theta}_d),$$

where $W^*(u;q)$ is the limit of the weighting matrices $W^{*(t-1)}(u;q)$ defined following (10.35).

If we assume that the weights defining the linear representation (10.39) are fixed, and that the values of y follow a location specific linear model, e.g. (10.30), then an estimator of the prediction variance of (10.39) can be computed following standard methods of heteroskedasticity-robust prediction variance estimation for linear predictors of population quantities (see Royall and Cumberland(1978)). Put $w_d = (w_{jd})$. This estimator is of the form

$$mse(\hat{\bar{Y}}_d^{MQGWR/CD}) = N_d^{-2}\sum_{k:n_k>0}\sum_{j\in s_k}\lambda_{jdk}\Big\{y_j - \hat{Q}_{\hat{\theta}_k}(\mathbf{x}_j;\psi,u_j)\Big\}^2, \qquad (10.41)$$

where $\lambda_{jdk} = \left\{(w_{jd} - 1)^2 + (n_d - 1)^{-1}(N_d - n_d)\right\}I(k = d) + w_{jk}^2 I(k \neq d)$ and $\hat{Q}_{\hat{\theta}_k}(\mathbf{x}_j;\psi,u_j)$ is assumed to define an unbiased estimator of the expected value of y_j given $\mathbf{x}_j$ at location u_j. Since the weights defining (10.40) reproduce the small area mean of $\mathbf{X}$, it also follows that (10.39) is unbiased for this mean in the special case where this expectation does not vary with location within the small area of interest, and so (10.41) then estimates the mean squared error of (10.39) in this special case. More generally, when the expectation of y_i given $\mathbf{x}_i$ varies from location to location within the small area, this unbiasedness holds on average provided sampling within the small area is independent of location, in which case (10.41) is an estimator of a first order approximation to the mean squared error of (10.39).

10.5 Semiparametric Fay-Herriot model using penalized splines

Traditional Fay-Herriot small area estimation models (Rao, 2003) are based on the hypothesis of a linear relationship between the variable of interest and the covariates, hypothesis that can represent a serious restriction in many real data applications. Furthermore, when detailed geo-referenced information for the units of analysis is available, as it is usually the case in environmental and poverty studies, traditional linear mixed models are not able to handle spatial proximity effects between the areas.

A semiparametric version of the basic Fay-Herriot model that is based on P-splines can also handle situations where the functional form of the relationship between the variable of interest and the covariates cannot be specified a priori (Giusti et al., in preparation). Furthermore, P-spline bivariate smoothing can easily introduce spatial effects in the area level model. This model is an extension with respect to other recently proposed models. In particular, extensions to random effects models have been proposed to allow for spatially correlated random area effects taking into account the information provided by neighboring areas (see Petrucci and Salvati (2006) and Pratesi and Salvati (2009)), but these models still rely on the linearity assumption. Opsomer et al. (2008) proposed a small area model based on P-splines but under the assumption that all the data are available at the unit level, and this can be a restriction in some situations.

10.5.1 Estimation of small area means

In this section we briefly recall the traditional model proposed by Fay and Herriot (1979) and we present the proposed extended model with a P-spline component. Let θ be the $d \times 1$ vector of the parameter of inferential interest (small area total y_d, small area mean $\bar{y}_d$ with $d = 1,...,D$) and assume that the $d \times 1$ vector of the direct estimator $\hat{\theta}$ is available and design unbiased. Denote the corresponding $d \times p$ matrix of the area level auxiliary variables by $\mathbf{X} = (\mathbf{x}_1,\ldots,\mathbf{x}_p)$. The Fay-Herriot model can be expressed as:

$$\hat{\theta} = \mathbf{X}\alpha + \mathbf{Z}\mathbf{u} + \varepsilon. \tag{10.42}$$

Here $\mathbf{u}$ is $m \times 1$ vector of independent and identically distributed random variables with mean $\mathbf{0}$ and $m \times m$ variance matrix $\Sigma_u = \sigma_u^2 \mathbf{I}_m$, $\mathbf{Z}$ is a $m \times m$ matrix of known positive constants, ε is the $m \times 1$ vector of independent sampling errors with mean $\mathbf{0}$ and known diagonal variance matrix $\mathbf{R} = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$ and α is the $q \times 1$ vector of regression parameters. The Fay-Herriot model is a general linear mixed model with diagonal covariance structure $\Sigma(\sigma_u^2) = \mathbf{Z}\Sigma_u\mathbf{Z}^T + \mathbf{R}$.

Assuming that the direct survey estimators are linear function of the covariates, the Fay-Herriot model produces reliable small area estimates by combining the design model and the regression model and then borrowing strength from other domains. In many situations, however, this hypothesis of linear relationship may fail down, leading to biased estimators of the small area parameters. In these cases significant advantages can be obtained with the proposed semiparametric specification of the Fay-Herriot model, which allows non linearities in the relationship between $\hat{\theta}$ and the auxiliary variables $\mathbf{X}$, can be obtained introducing a penalized-splines component.

A semiparametric model with one covariate x_1 can be written as $\tilde{m}(\mathbf{x}_1)$, where the function $\tilde{m}(\cdot)$ is unknown, but assumed to be sufficiently well approximated by the function

$$m(\mathbf{x}_1;\eta,\gamma)=\eta_0+\eta_1\mathbf{x}_1+\ldots+\eta_p\mathbf{x}_1^p+\sum_{k=1}^{K}\gamma_k(\mathbf{x}_1-\kappa_k)_+^p \tag{10.43}$$

where $\eta=(\eta_0,\eta_1,\ldots,\eta_p)^T$ is the $(p+1)$ vector of the coefficients of the polynomial function, $\gamma=(\gamma_1,\gamma_2,\ldots,\gamma_K)^T$ is the coefficient vector of the truncated polynomial spline basis (P-spline) and p is the degree of the spline $(t)_+^p=t^p$ if $t>0$ and 0 otherwise. The latter portion of the model allows for handling departures from a p-polynomial t in the structure of the relationship. In this portion κ_k for $k=1,\ldots,K$ is a set of fixed knots and if K is sufficiently large, the class of functions in (10.43) is very large and can approximate most smooth functions. Details on bases and knots choice can be found in Chapters 3 and 5 of Ruppert et al. (2003). Since a P-spline model can be viewed as a random-effects model (see Ruppert et al. (2003) and Opsomer et al. (2008)), it can be combined with the Fay-Herriot model for obtaining a semiparametric small area estimation framework based on linear mixed model regression.

Given the η and γ vectors, define

$$\mathbf{X}_1=\begin{bmatrix}1 & x_{11} & \cdots & x_{11}^p\\ \vdots & \vdots & \ddots & \vdots\\ 1 & x_{1m} & \cdots & x_{1m}^p\end{bmatrix},$$

and

$$\mathbf{S}=\begin{bmatrix}(x_{11}-\kappa_1)_+^p & \cdots & (x_{11}-\kappa_K)_+^p\\ \vdots & \ddots & \vdots\\ (x_{1m}-\kappa_1)_+^p & \cdots & (x_{1m}-\kappa_K)_+^p\end{bmatrix}.$$

Using only one covariate, x_1, the semiparametric Fay-Herriot can be written as

$$\hat{\beta}=\mathbf{X}\beta+\mathbf{S}\gamma+\mathbf{Zu}+\varepsilon, \tag{10.44}$$

where $\mathbf{X}=\mathbf{X}_1$, β is a $(p+1)$ vector of regression coefficients, the γ component can be treated as a $K\times 1$ vector of independent and identically distributed random variables with mean $\mathbf{0}$ and $K\times K$ variance matrix $\Sigma_\gamma=\sigma_\gamma^2\mathbf{I}_K$. The variance-covariance matrix of the model (10.44) is $\Sigma(\psi)=\mathbf{S}\Sigma_\gamma\mathbf{S}^T+\mathbf{Z}\Sigma_u\mathbf{Z}^T+\mathbf{R}$ where $\psi=(\sigma_\gamma^2,\sigma_u^2)^T$.

Model-based estimation of the small area parameters can be obtained by using the best linear unbiased prediction (see Henderson (1975):

$$\tilde{\theta}^B(\psi)=\mathbf{X}\tilde{\beta}(\psi)+\Lambda(\psi)[\hat{\theta}-\mathbf{X}\tilde{\beta}(\psi)] \tag{10.45}$$

with $\Lambda(\psi) = (\mathbf{S}\Sigma_\gamma \mathbf{S}^T + \mathbf{Z}\Sigma_u \mathbf{Z}^T)\Sigma^{-1}(\psi)$ and $\tilde{\beta}(\psi) = (\mathbf{X}^T\Sigma^{-1}(\psi)\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}(\psi)\hat{\theta}$.

Extension to bivariate smoothing can be handled by assuming $\tilde{m}(\mathbf{x}_1, \mathbf{x}_2) = m(\mathbf{x}_1, \mathbf{x}_2; \eta, \gamma)$. See details in Opsomer et al. (2008). This is of central interest in a number of application areas as environment, public health and poverty mapping. It has particular relevance when referenced responses need to be converted to maps.

10.5.2 Estimation of the MSE

The Mean Squared Error estimator (MSE) of $\tilde{\theta}^B(\psi)$, depending on the variance components $\psi = (\sigma^2_\gamma, \sigma^2_u)^T$, can be expressed as in Rao (2003):

$$MSE[\tilde{\theta}^B(\psi)] = g_1(\psi) + g_2(\psi) \tag{10.46}$$

where the first term

$$g_1(\psi) = \Lambda(\psi)\mathbf{R} = \mathbf{R} - \mathbf{R}\Sigma^{-1}(\psi)\mathbf{R} \tag{10.47}$$

is due to the estimation of random effects and it is of order $O(1)$, while the second term

$$g_2(\psi) = \mathbf{R}\Sigma^{-1}(\psi)\mathbf{X}(\mathbf{X}^T\Sigma^{-1}(\psi)\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}(\psi)\mathbf{R} \tag{10.48}$$

is due to the estimation of β and it is of order $O(m^{-1})$ for large m.

The estimator $\tilde{\theta}^B(\psi)$ depends on the unknown variance components σ^2_γ and σ^2_u. Replacing the parameters with estimators $\hat{\sigma}^2_\gamma$, $\hat{\sigma}^2_u$, a two stage estimator $\tilde{\theta}^E(\hat{\psi})$ is

$$\tilde{\theta}^E(\hat{\psi}) = \mathbf{X}\hat{\beta}(\hat{\psi}) + \hat{\Lambda}(\hat{\psi})[\hat{\theta} - \mathbf{X}\hat{\beta}(\hat{\psi})] \tag{10.49}$$

where $\hat{\beta}(\hat{\psi}) = (\mathbf{X}^T\hat{\Sigma}^{-1}(\hat{\psi})\mathbf{X})^{-1}\mathbf{X}^T\hat{\Sigma}^{-1}(\hat{\psi})\hat{\theta}$. Assuming normality of the random effects, σ^2_γ and σ^2_u can be estimated both by Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML) procedures (see Prasad and Rao (1990)).

For any $\hat{\psi}$ satisfying (ii) and (iii), the MSE of $\tilde{\theta}^E(\hat{\psi})$ can be decomposed as

$$MSE[\tilde{\theta}^E(\hat{\psi})] = g_1(\psi) + g_2(\psi) + E\left\{[\tilde{\theta}^E(\hat{\psi}) - \tilde{\theta}^B(\psi)]^2\right\} = g_1(\psi) + g_2(\psi) + g_3(\psi). \tag{10.50}$$

The results of Opsomer et al. (2008) can be used for deriving a second order approximation to the $g_3(\psi)$ term. It can be given by

$$g_3(\psi) = \mathbf{L}^T(\psi)\left[I^{-1}(\psi) \otimes \Sigma(\psi)\right]\mathbf{L}(\psi) + o(\delta_m/m) \tag{10.51}$$

where

$$\mathbf{L}(\psi) = [\mathbf{L}_{\sigma^2_\gamma}(\psi), \mathbf{L}_{\sigma^2_u}(\psi)]^T, \mathbf{L}_i(\psi) = \frac{\partial\Lambda(\psi)}{\partial\psi_i},\ i = 1,2.$$

Here $\otimes$ represents Kronecker product, $I^{-1}(\psi)$ is the inverse of the information matrix with $I^{-1}_{ij}(\psi) = 0.5tr[\mathbf{P}(\psi)\mathbf{B}_i\mathbf{P}(\psi)\mathbf{B}_j]$, $i,j = 1,2$, $\mathbf{P}(\psi) = \Sigma^{-1}(\psi) - \Sigma^{-1}(\psi)\mathbf{X}(\mathbf{X}^T\Sigma^{-1}(\psi)\mathbf{X})^{-1}\mathbf{X}^T\Sigma^{-1}(\psi)$, $\mathbf{B}_1 = \mathbf{S}\mathbf{S}^T$ and $\mathbf{B}_2 = \mathbf{Z}\mathbf{Z}^T$ and $\delta_m = o(\sqrt{m})$.

In practical applications, the EBLUP $\tilde{\theta}^{E}(\hat{\psi})$ should be accompanied with an estimate of the MSE. Following the results of Prasad and Rao (1990) and Das et al. (2004), Opsomer et al. (2008) extended the Prasad-Rao MSE estimator to models with more general covariance structure. An approximately unbiased estimator of the MSE is

$$mse[\tilde{\theta}^{E}(\hat{\psi})] = g_1(\hat{\psi}) + g_2(\hat{\psi}) + 2g_3(\hat{\psi}). \tag{10.52}$$

which is the same estimator derived by Prasad and Rao (1990). In formula (10.52), the term $g_3(\hat{\psi})$ appears twice due to a bias correction of $g_1(\hat{\psi})$.

An alternative procedure for estimating the MSE of the EBLUP $\tilde{\theta}^{E}(\hat{\psi})$ could be based on bootstrapping according to the bootstrap procedure proposed by González-Manteiga et al. (2007), Opsomer et al. (2008) and Molina et al. (2009).

Chapter 11

Estimating Small Area Averages

In this chapter we describe the application of small area methodologies for estimating small area averages in the regions of Lombardia, Toscana and Campania. In particular, the methodologies for estimating small area averages and the target geographies are listed below.

- M-quantile model for estimating averages at the province level
- Nonparametric M-quantile model for estimating averages at the province level
- MQGWR model applied only in the region of Toscana for estimating municipal averages
- Nonparametric Fay and Herriot model for estimating municipal averages

In the rest of the chapter we describe the specifications of the small area working models, the methodologies used for estimating small area averages and we present the results from the application of each model.

11.1 M-quantile Models

The working M-quantile small area model uses the equivalised household income as the outcome variable. The explanatory variables, common in the survey and in the Census micro-data, that we used in the model are the ownership status, the age of the head of the household, the employment status of the head of the household, the gender of the head of the household, the years of education of the head of the household and the household size. Figure 11.1 illustrates the relationship between the province M-quantile coefficients, estimated with the M-quantile small area model, and the province random effects estimated with a random effects model. This plot indicates that there is a positive correlation between M-quantile coefficients and random effects with negative random effects corresponding to M-quantile coefficients below 0.46 and positive random effects corresponding to M-quantile coefficients above 0.46. This provides strong evidence that the M-quantile coefficients capture well the between province variability.

The M-quantile small area methodology was fully described in Chapter 10 of Deliverables 12 and 16 and the properties of point and MSE estimators were evaluated using model-based and design-based simulations. A brief recall of the method can be find in Chapter 10. Using the M-quantile small area

Figure 11.1: Plot of the province M-quantile coefficients vs the province random effects

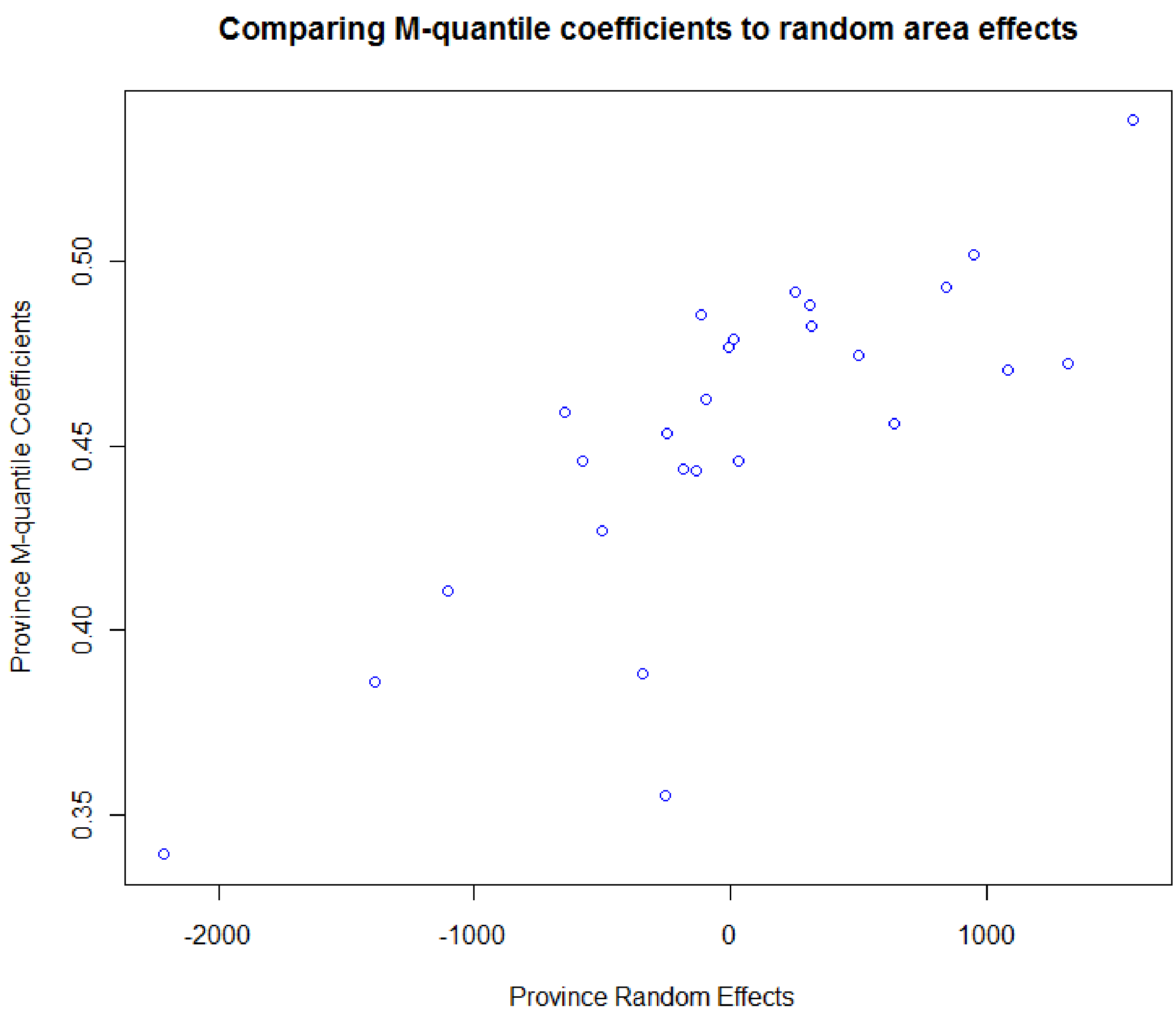


Table 11.1: Estimated Mean of Household Equivalised Income (MEAN) and estimated Root Mean Squared Error of the Mean estimator (RMSE) for the Lombardia Provinces. M-quantile model.

	PROVINCE	MEAN	RMSE
1	VARESE	21091.49	1305.98
2	COMO	18578.33	1137.01
3	SONDRIO	16307.16	1668.92
4	MILANO	20798.63	497.68
5	BERGAMO	18323.07	820.61
6	BRESCIA	16326.21	581.47
7	PAVIA	21081.25	4080.17
8	CREMONA	16774.18	883.69
9	MANTOVA	17774.90	677.24
10	LECCO	19497.61	1131.62
11	LODI	17052.58	965.49

model, estimates of the average income for provinces in the three Italian regions (Lombardia, Toscana and Campania) are obtained using estimator (10.7). Corresponding estimates of the Mean Squared Error (MSE) are obtained by applying estimator (10.12). The expressions of the point and MSE estimators can be found in Chapter 10 and in Chapter 10 of Deliverable 12 and 16.

The results (point and MSE estimation) from the application of the M-quantile model are presented in Figures 11.2, 11.3 and 11.4 and also in Tables 11.1, 11.2, 11.3. The estimates of the average income for each province show that there are intra-regional differences. In Lombardia, the provinces of Milano, Pavia and Varese have the highest mean equivalised household income while the provinces of Sondrio, Cremona and Brescia have lower average income. Such intra-regional variability is also present in Toscana. The provinces of Siena and Firenze appear to be as wealthy as the wealthier provinces of Lombardia whereas the provinces of Lucca and Massa-Cararra have lower average income. These results indicate that Toscana and Lombardia have similar levels of average equivalised household income although, one may say that Lombardia is somewhat wealthier. Looking now at the results of the southern region of Campania, it is clear that provinces in this region have smaller average equivalised household income than provinces in Lombardia and Toscana. Compared to Caserta and Benevento, the provinces of Avellino, Salerno and Napoli have higher average income although the intra-regional differences in Campania are not so pronounced.

11.2 Nonparametric M-quantile Model

Extending the linear M-quantile model, we further investigate the use of a Nonparametric M-quantile model for estimating small area averages. The use of the nonparametric M-quantile model is justified when the functional form of the relationship between the q-th M-quantile and the covariates deviates from linearity. In this case the traditional M-quantile small area model can lead to biased estimators of the small area parameters. The P-spline M-quantile small area model preserves the properties of the

Figure 11.2: Estimated Mean (Root Mean Squared Error) of Household Equivalised Income for Lombardia Provinces. M-quantile model.

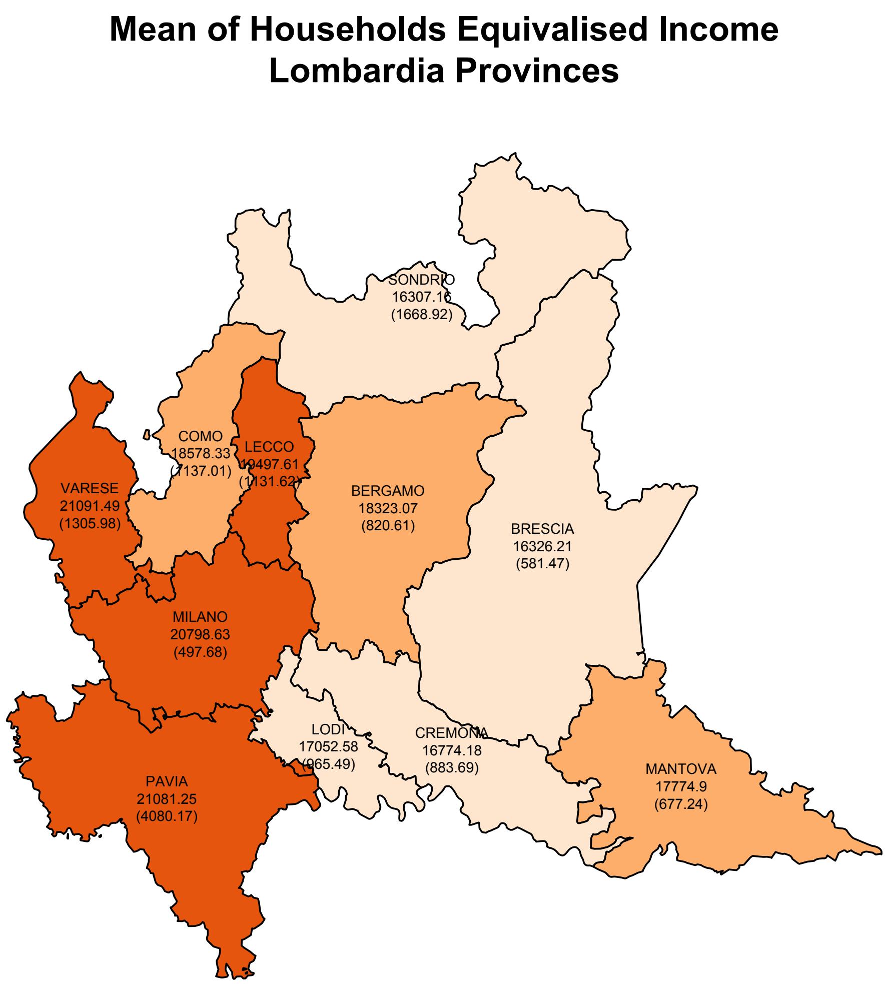


Table 11.2: Estimated Mean of Household Equivalised Income (MEAN) and estimated Root Mean Squared Error of the Mean estimator (RMSE) for the Toscana Provinces. M-quantile model.

	PROVINCE	MEAN	RMSE
1	MASSA CARRARA	14128.26	664.84
2	LUCCA	15867.69	766.80
3	PISTOIA	18980.76	1119.33
4	FIRENZE	19184.92	498.35
5	LIVORNO	17875.01	919.41
6	PISA	18550.16	876.37
7	AREZZO	18665.97	1014.42
8	SIENA	20228.98	1113.91
9	GROSSETO	16152.47	1151.84
10	PRATO	17702.87	632.74

Table 11.3: Estimated Mean of Household Equivalised Income (MEAN) and estimated Root Mean Squared Error of the Mean estimator (RMSE) for the Campania Provinces. M-quantile model.

	PROVINCE	MEAN	RMSE
1	CASERTA	11685.74	574.89
2	BENEVENTO	11312.89	1033.79
3	NAPOLI	12661.84	291.73
4	AVELLINO	12873.13	979.46
5	SALERNO	12715.91	502.22

Figure 11.3: Estimated Mean (Root Mean Squared Error) of Household Equivalised Income for Toscana Provinces. M-quantile model.

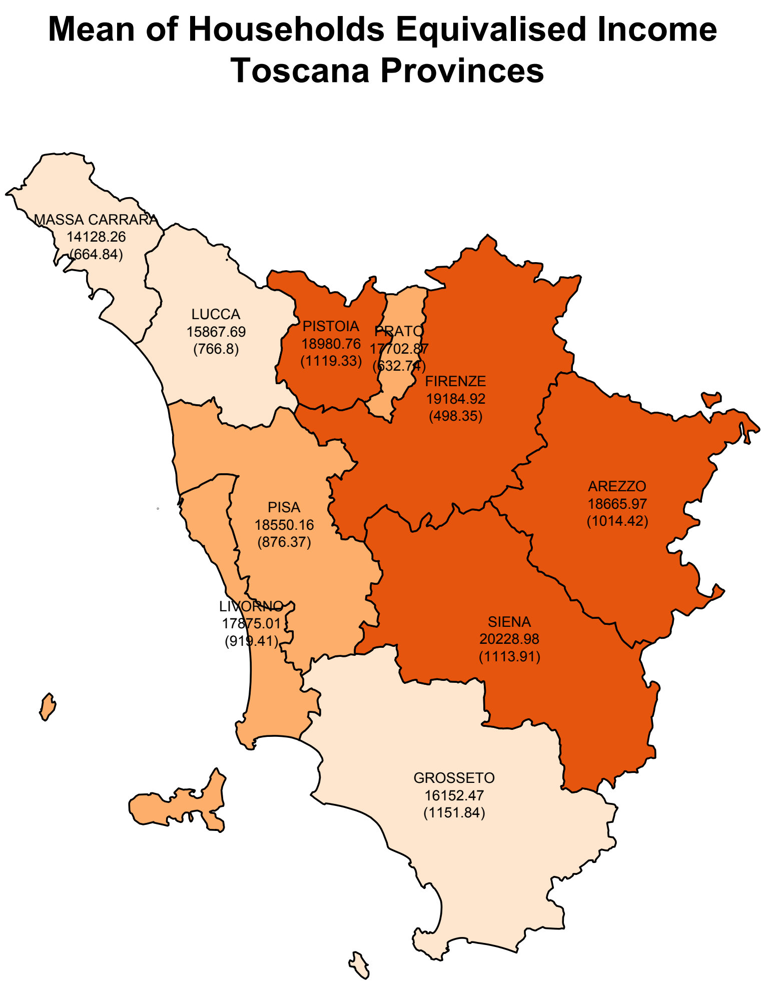


Figure 11.4: Estimated Mean (Root Mean Squared Error) of Household Equivalised Income for Campania Provinces. M-quantile model.

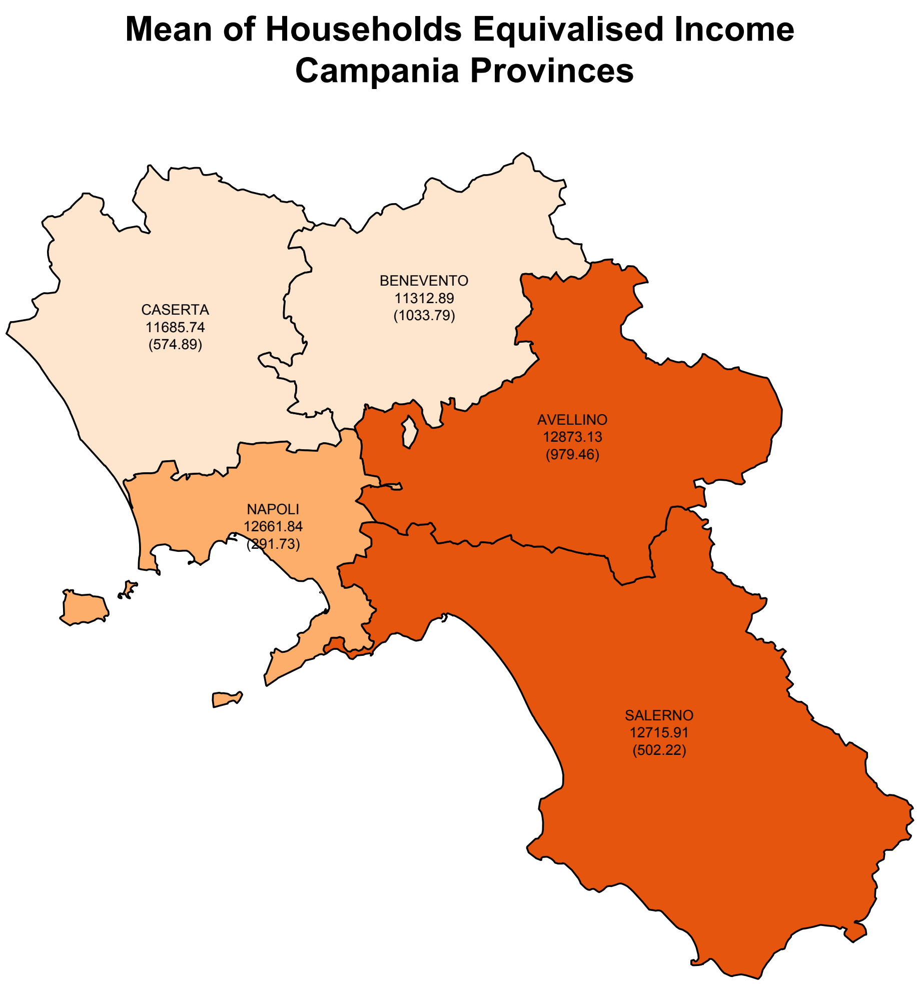
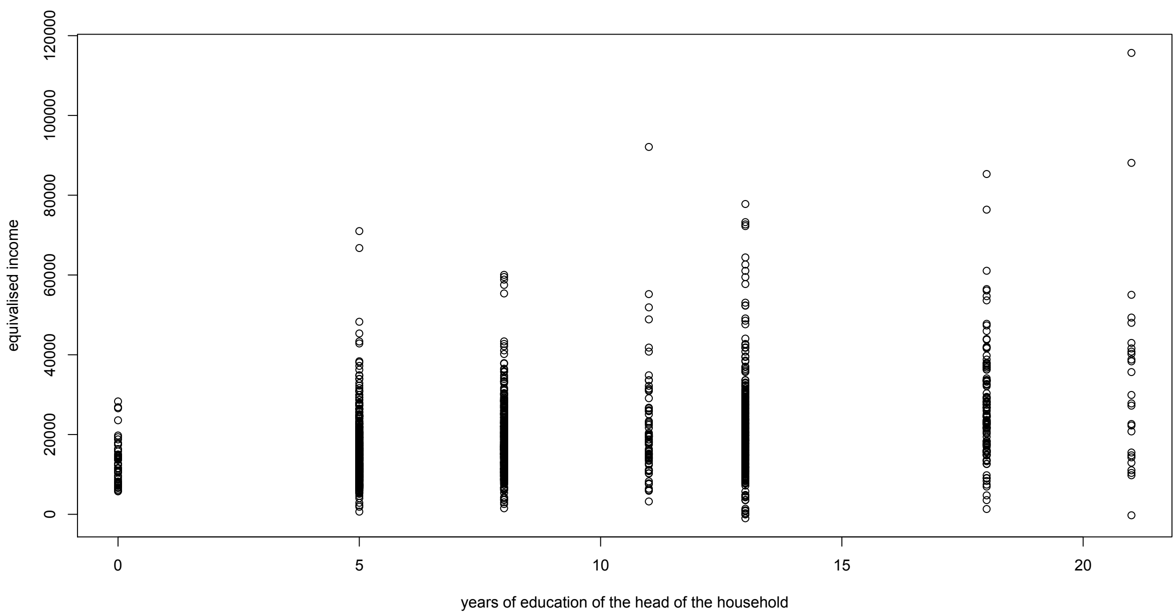


Figure 11.5: Relationship between the equivalised household income and the years of education of the head of the household.



linear M-quantile small area model and at the same time it allows for more complex functional forms that can be estimated from the data.

Small area estimation under the nonparametric M-quantile model has been described in Chapter 10 of Deliverables 12 and 16 and a short description can be find in Chapter 10. The estimates of the average income for provinces are obtained using estimator (10.22) and the estimates of the Mean Squared Error (MSE) are computed by applying estimator (10.29).

The specification of the model is the same to the one we used with the linear M-quantile model, but we further consider the years of education of the head of the household as the smoothing covariate. Figure 11.5 shows a slightly non linear relationship between the outcome variable, equivalised household income, and the years of education of the head of the household. Nevertheless, it appears that the use of the nonparametric M-quantile model is not justified. This is demonstrated by (i) the fact that the λ parameter that controls the level of smoothness of the resulting fit is large for each quantile and (ii) the fact that the small area estimates of average income produced by the nonparametric model are identical to the small area estimates from the linear M-quantile model. The same is true for the corresponding MSE estimates (see Tables 11.4, 11.5, 11.6 and Maps 11.6, 11.7, 11.8).

Table 11.4: Estimated Mean of Household Equivalised Income (MEAN) and estimated Root Mean Squared Error of the Mean estimator (RMSE) for the Lombardia Provinces. Nonparametric M-quantile model.

	PROVINCE	MEAN	RMSE
1	VARESE	21091.52	1305.98
2	COMO	18578.39	1137.01
3	SONDRIO	16307.19	1668.94
4	MILANO	20798.62	497.68
5	BERGAMO	18323.10	820.61
6	BRESCIA	16326.31	581.47
7	PAVIA	21081.27	4080.17
8	CREMONA	16774.19	883.69
9	MANTOVA	17774.93	677.24
10	LECCO	19497.76	1131.63
11	LODI	17052.68	965.49

Table 11.5: Estimated Mean of Household Equivalised Income (MEAN) and estimated Root Mean Squared Error of the Mean estimator (RMSE) for the Toscana Provinces. Nonparametric M-quantile model.

	PROVINCE	MEAN	RMSE
1	MASSA CARRARA	14128.28	664.84
2	LUCCA	15867.73	766.81
3	PISTOIA	18980.78	1119.33
4	FIRENZE	19184.94	498.35
5	LIVORNO	17875.07	919.41
6	PISA	18550.18	876.38
7	AREZZO	18665.97	1014.42
8	SIENA	20229.07	1113.91
9	GROSSETO	16152.52	1151.83
10	PRATO	17702.90	632.74

Figure 11.6: Estimated Mean (Root Mean Squared Error) of Household Equivalised Income for Lombardia Provinces. Nonparametric M-quantile model.

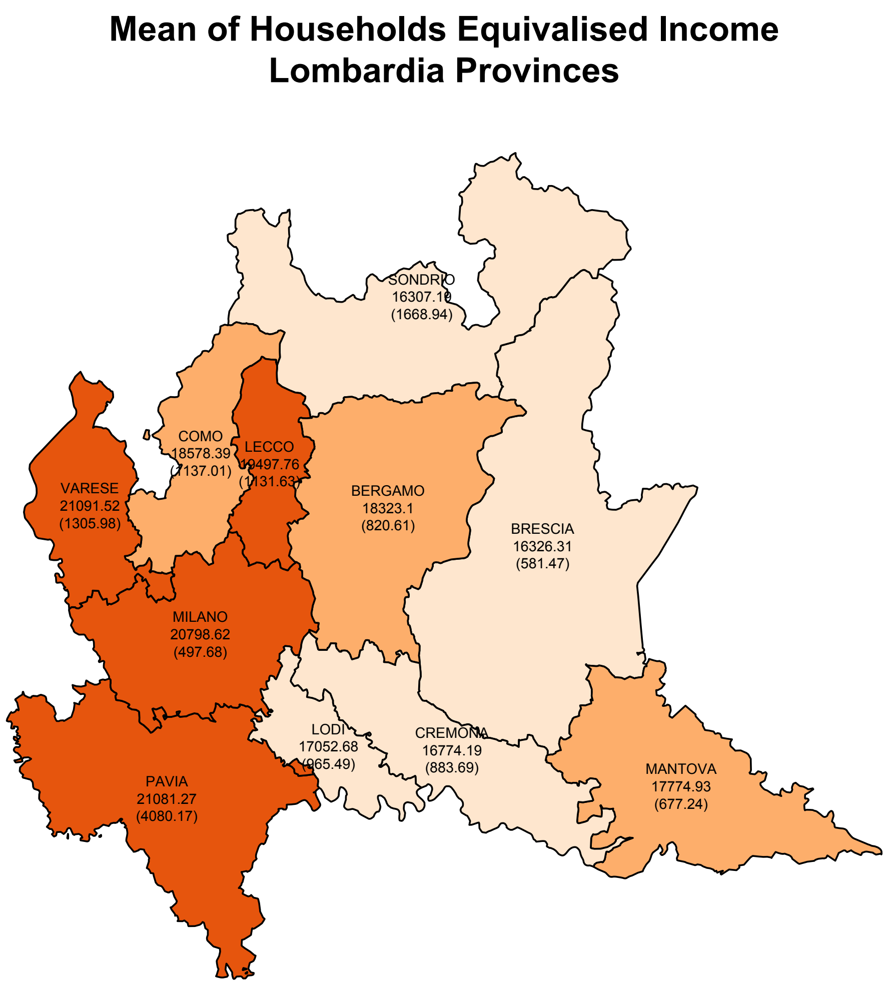


Figure 11.7: Estimated Mean (Root Mean Squared Error) of Household Equivalised Income for Toscana Provinces. Nonparametric M-quantile model.

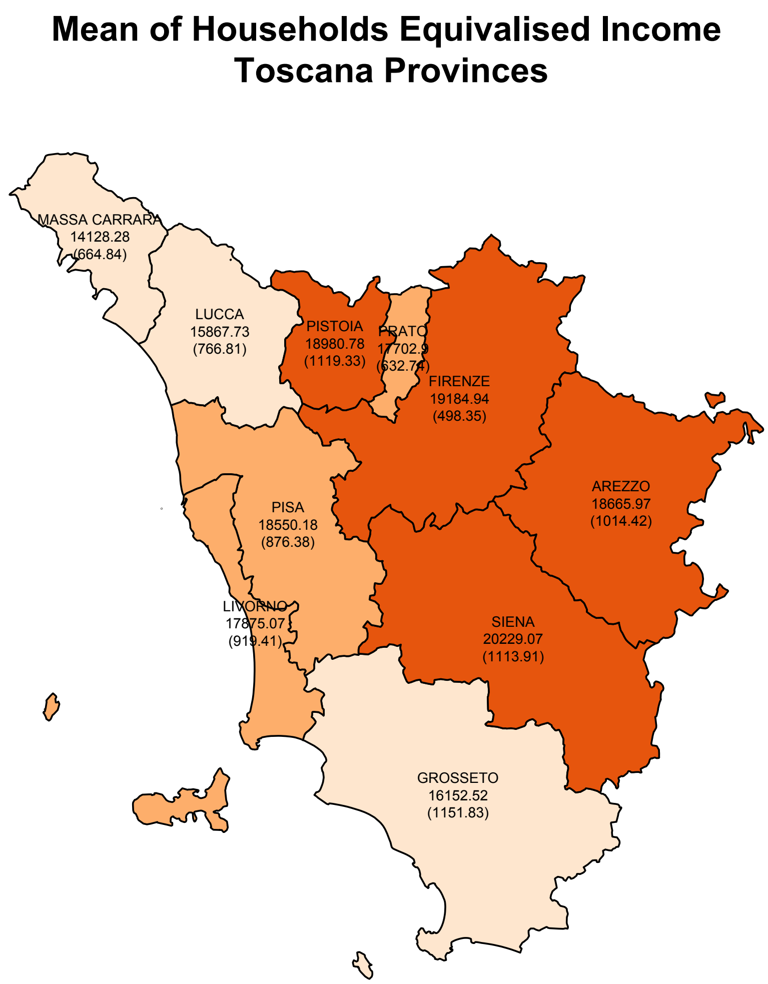


Figure 11.8: Estimated Mean (Root Mean Squared Error) of Household Equivalised Income for Campania Provinces. Nonparametric M-quantile model.

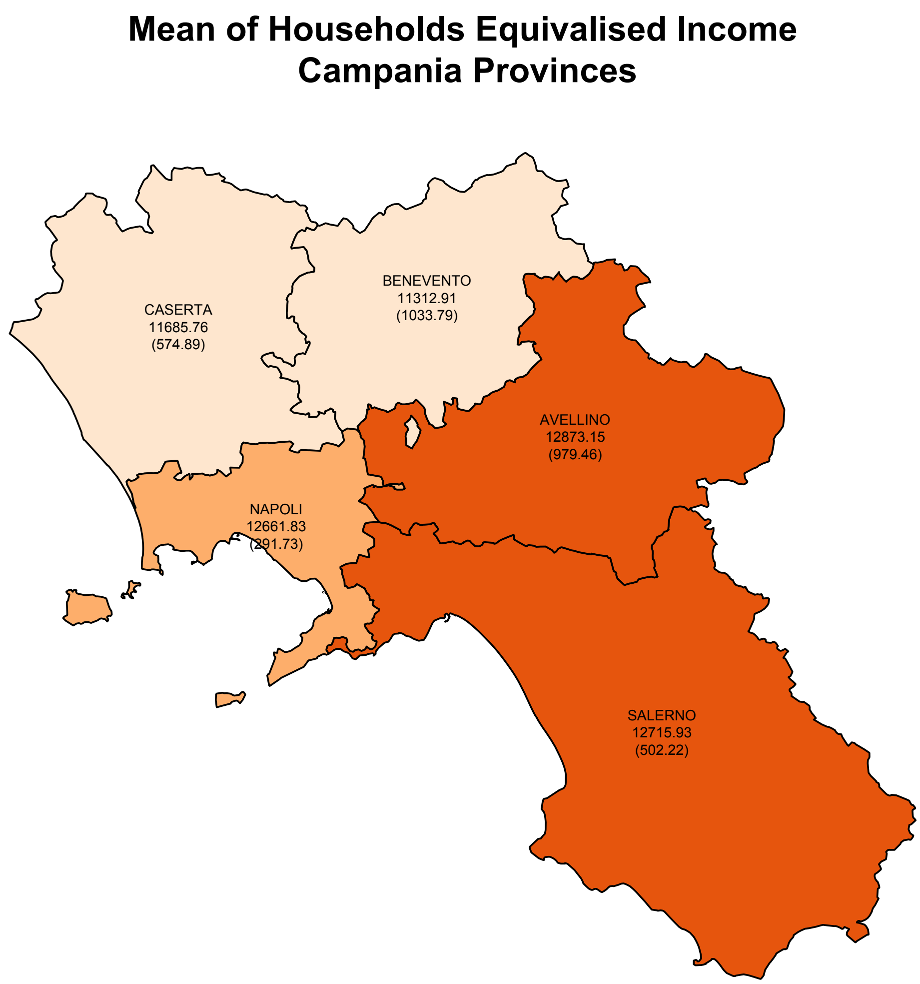


Table 11.6: Estimated Mean of Household Equivalised Income (MEAN) and estimated Root Mean Squared Error of the Mean estimator (RMSE) for the Campania Provinces. Nonparametric M-quantile model.

	PROVINCE	MEAN	RMSE
1	CASERTA	11685.76	574.89
2	BENEVENTO	11312.91	1033.79
3	NAPOLI	12661.83	291.73
4	AVELLINO	12873.15	979.46
5	SALERNO	12715.93	502.22

11.3 M-quantile GWR Model

In this Section we explore the use of M-quantile GWR for the estimation of the equivalised household income in the 287 municipalities of Toscana, of which 228 did not have any observations or at most one observation. A full description of the M-quantile GWR small area methodology has been described in Chapter 10 of Deliverables 12 and 16 and summarized in Chapter 10. The estimates of the average income for municipalities are obtained using estimator (10.38) and the estimates of the Mean Squared Error (MSE) are computed by applying estimator (10.41). The properties of the MQGWR predictors have been studied through model-based and design based simulation studies. The results from these studies reported in Deliverable 12 and 16 suggest that the M-quantile GWR model represents a promising alternative for flexibly incorporating spatial information into SAE. In addition, the performance of the proposed MSE estimator for the M-quantile GWR small area predictors is promising.

The explanatory variables that we used for fitting the M-quantile GWR model are the same to the ones used for fitting the linear and Non-parametric M-quantile models. Moreover, we have used the spatial coordinates of the centroid of each municipality for taking into account that observations that are spatially close may be more related than observations that are further apart.

The M-quantile GWR model captures the spatial variation by using local rather than global parameters in the regression model. That is, a GWR model assumes spatial non-stationarity of the conditional mean of the variable of interest. For using the M-quantile GWR model we need to know the geographical coordinates for the sampled and non-sampled units. In the EU-SILC data we do not have this information. For this reason we have adapted the model to the situation where only spatial information about the centroids of the small areas (municipalities) is available. By using an R function (`get.Pcent()` by library `maptools`) we have obtained the spatial coordinates of the centroids of each municipality. In this case one can use the centroids of each small area to estimate the regression coefficients of the M-quantile GWR model. This means that there is the same vector of coefficients for all sampled and non-sampled units belonging to area ith. But the spatial coefficients vary between municipalities. Before applying the M-quantile GWR model we have used a spatial diagnostic to test the spatial non-stationarity. There is evidence of a non-stationary process. In particular, using an ANOVA test proposed by Brundson et al. (1996) we reject the null hypothesis of stationarity of the model parameters (F = 2.0056 and p-value ¡ 2.2e-16). The spatial weight matrix used in fitting these M-quantile GWR models was constructed using expression (10.34) in Chapter 10 of Deliverables 12 and 16, with bandwidth selected with cross-validation (page 225 in Chapter 10 of Deliverables 12 and 16).

In Figure 11.9 we present a map of the estimated values of average household income for each municipality in Toscana using the M-quantile GWR model. The small area model indicates that there are lower levels of average household income in the North-West and in the South-West of the study region. The M-quantile GWR models also provides high estimates of average household income in some municipalities of the province of Florence, Arezzo and Siena. These results are consistent with the spatial distribution of the average values of household income produced by other nonparametric models (see the results of the nonparametric Fay-Herriot). This indicates that the small area models that allow for more flexible incorporation of the spatial information produce overall consistent results.

Detailed results are presented in appendix A Table A.1.

11.4 Semiparametric Fay and Herriot model

In this section we estimate the mean household equivalilsed income using a semiparametric Fay-Herriot model in the 1546 municipalities of Lombardia of which 1456 did not have any observations or at most one observation, in the 287 municipalities of Toscana of which 228 did not have any observations or at most one observation and in the 551 municipalities of Campania of which 508 did not have any observations or at most one observation. The semiparametric Fay-Herriot model was presented in Chapter 6 of Deliverables 12 and 16; a review of this model can be find in Chapter 10.

Traditional Fay-Herriot small area estimation models is based on the hypothesis of a linear relationship between the variable of interest and the covariates, an hypothesis that can represent a serious restriction in many real data applications.

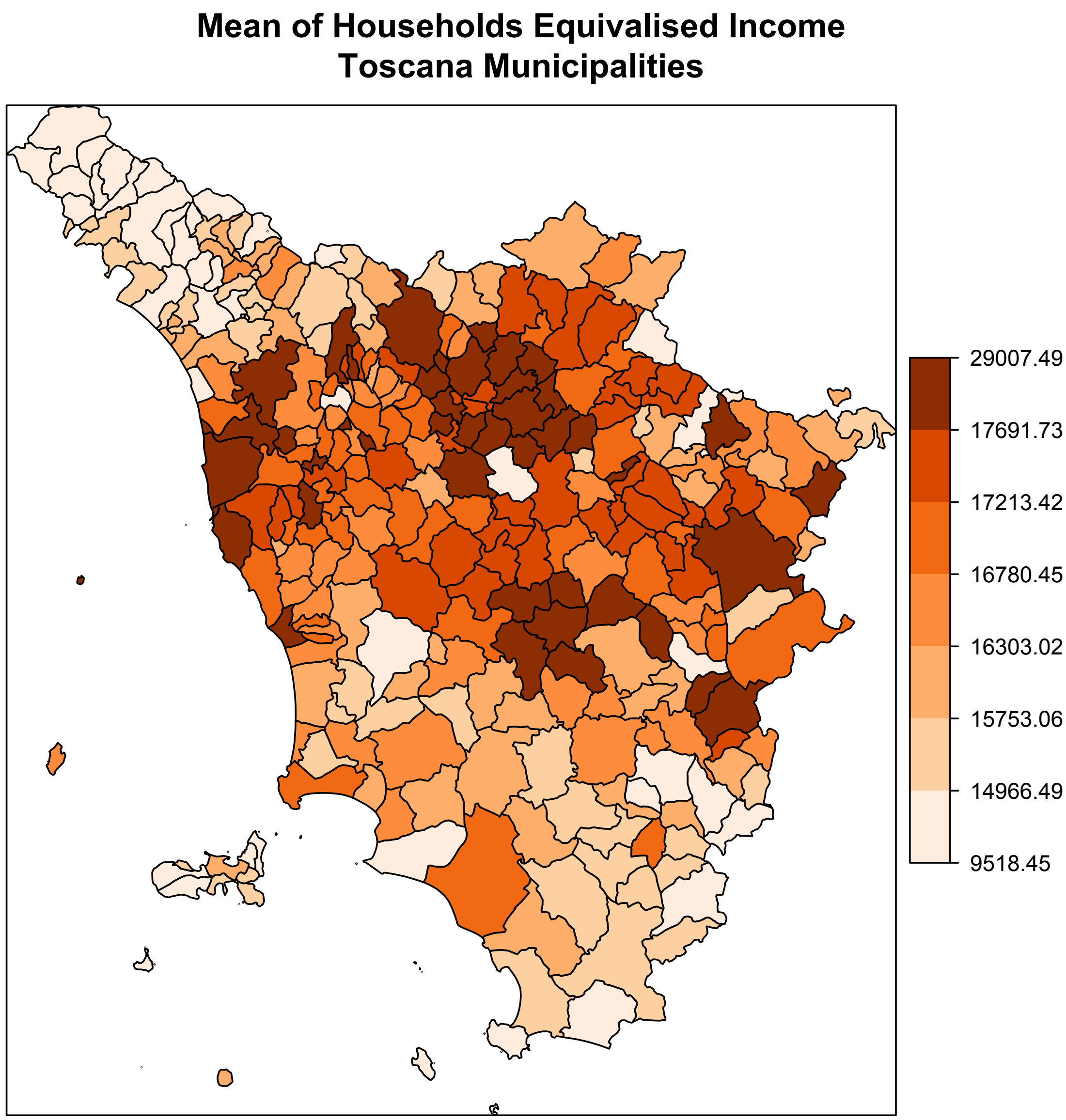
In Deliverables 12 and 16 we presented a semiparametric version of the basic Fay-Herriot model that is based on P-splines and can handle situations where the functional form of the relationship between the variable of interest and the covariates cannot be specified a priori (Giusti et al., in preparation). This is often the case when spatial proximity effects are present. In these cases P-spline bivariate smoothing can easily introduce spatial effects in the area level model.

In the problem of estimating the mean household equivalised income in the regions of Italy we can use a bivariate version of model (10.43) where the ownership of the house, the years, the employment status, the gender and the years of education of the head of the household and the number of household members enter in the chosen mode linearly, while the coordinates (latitude and longitude) of the centroid of the municipalities enter through an unknown smooth bivariate function.

As we can see from the map in Figure 11.10, the municipalities with higher estimated mean household income are concentrated in the north-western and north-eastern parts of the region Lombardia. At the opposite end, the municipalities characterized by lower estimates are concentrated in the central and south-eastern parts of the region. These findings are coherent with the estimates at provincial level from the M-quantile and nonparametric M-quantile models with some exceptions. Some municipalities in the province of Sondrio belong to the richest ones according to the semiparametric Fay-Herriot model while for the M-quantile and Nonparametric M-quantile the province of Sondrio shows a low level income.

Beyond these general indications, note that the semiparametric Fay-Herriot model allows us to discriminate the mean estimates among the municipalities belonging to the same province. In particular, we can see that a certain number of municipalities with estimated household income between approximately 20000 and 28000 Euros (the highest class in Figure 11.10) can be found also in the south-eastern and

Figure 11.9: Estimated Mean of Household Equivalised Income for Toscana Municipalities. M-quantile GWR model.



central-eastern parts of the region. Note also the the municipality of Milano - the biggest municipality in the central-western part of the region in Figure 11.10 - is, as expected, in the higher class of estimated income.

From Figure 11.11 we can see that the north-western part of the Toscana region, corresponding to the municipalities of the province of Massa-Carrara and to the northern part of the province of Lucca, is characterized by the lowest estimates for the mean household equivalised income at municipality level. This result is coherent not only with the estimates deriving from the M-quantile and nonparametric M-quantile models applied to the same data at provincial level, but also with the findings at municipality level derived from the M-quantile GWR model. Thus, this area can be considered the more critical in the region, at least considering the mean estimates of the income. On the other hand, the municipalities in the highest class of estimated income are concentrated in surroundings of the municipalities of Firenze, Siena and Arezzo, in the central-eastern part of the region. These results, that are coherent with those derived from the M-quantile GWR model, confirm that the mean household income is usually higher in the municipalities around the biggest cities. Note that the only relevant differences with respect to the estimates obtained with the M-quantile GWR model characterize the municipalities in the south-eastern part of the region: this may depend on the location of the knots of the nonparametric model, and we will further investigate this aspect in future analyses.

Figure 11.12 highlights that the mean income estimates for the Campania region are more geographically differentiated. Note for example how the municipalities characterized by the highest class of estimated household equivalised income are spread in the region. The only concentration of municipalities with higher income estimates can be found in the coastal areas (central-western part of the region), and in particular near the city of Napoli.

Finally, the estimates of the average income at the municipality level allow us to further investigate the gap between the three regions of Italy. In particular, we can notice that only few municipalities in Lombardia have an estimated average income above the highest level of estimated income in Toscana, while some municipalities in Lombardia have an estimated average below the lowest level in Toscana: thus, we can say that the gap between these two regions is not so pronounced. Considering the estimates for the Campania region we can notice instead a big gap with the northern regions: the highest estimates in Campania are comparable to the lowest ones in Lombardia and Toscana. These results confirm the existence of the so-called 'north-south divide' in Italy as concerns the wealthy conditions of the population.

The Tables with the mean for all the municipalities of Lombardia, Toscana and Campania are shown in the appendix B, Tables B.1, B.2, B.3.

Figure 11.10: Estimated Mean of Household Equivalised Income (x1000) for Lombardia Municipalities. Semiparametric Fay and Herriot model.

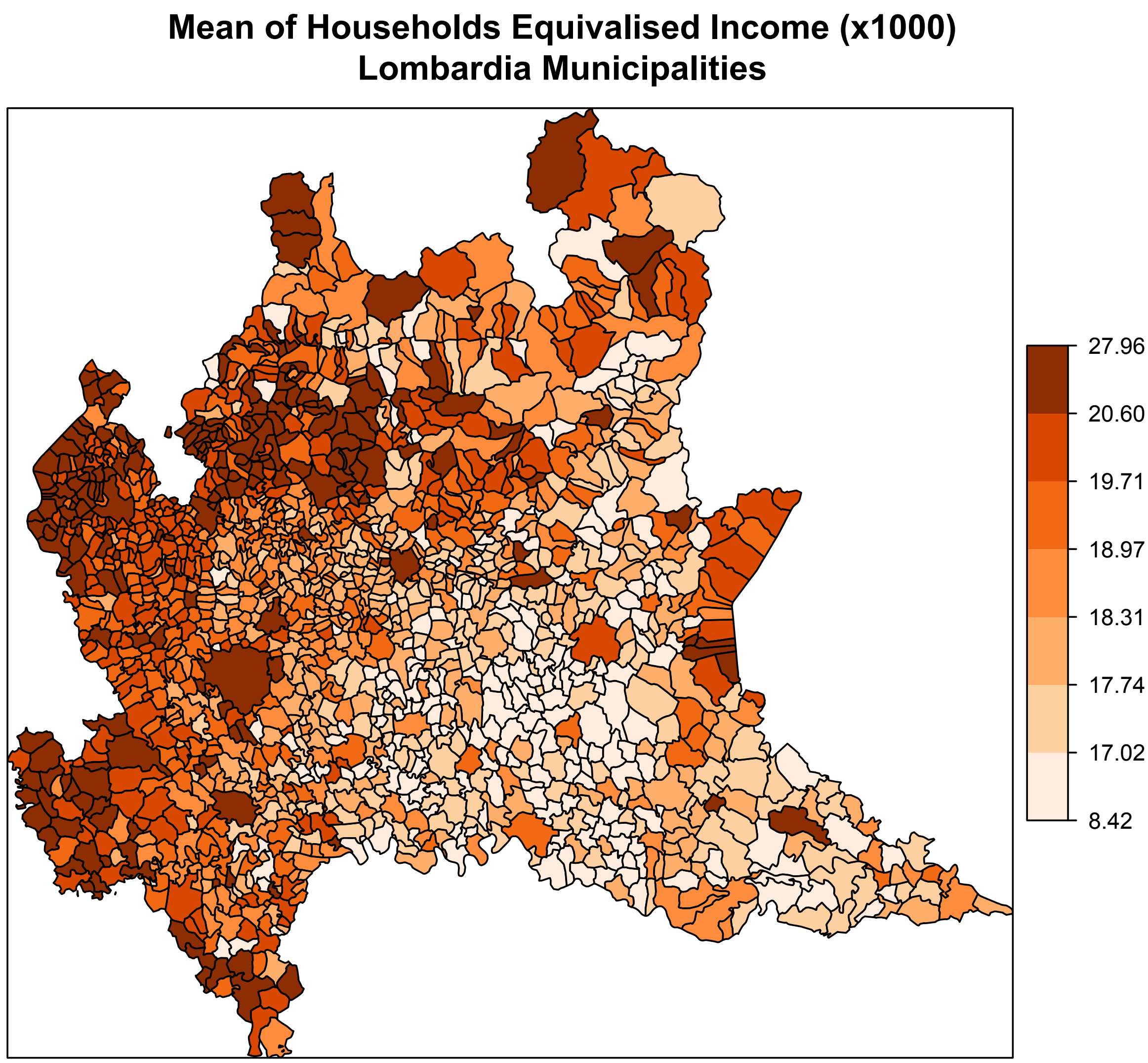


Figure 11.11: Estimated Mean of Household Equivalised Income (x1000) for Toscana Municipalities. Semiparametric Fay and Herriot model.

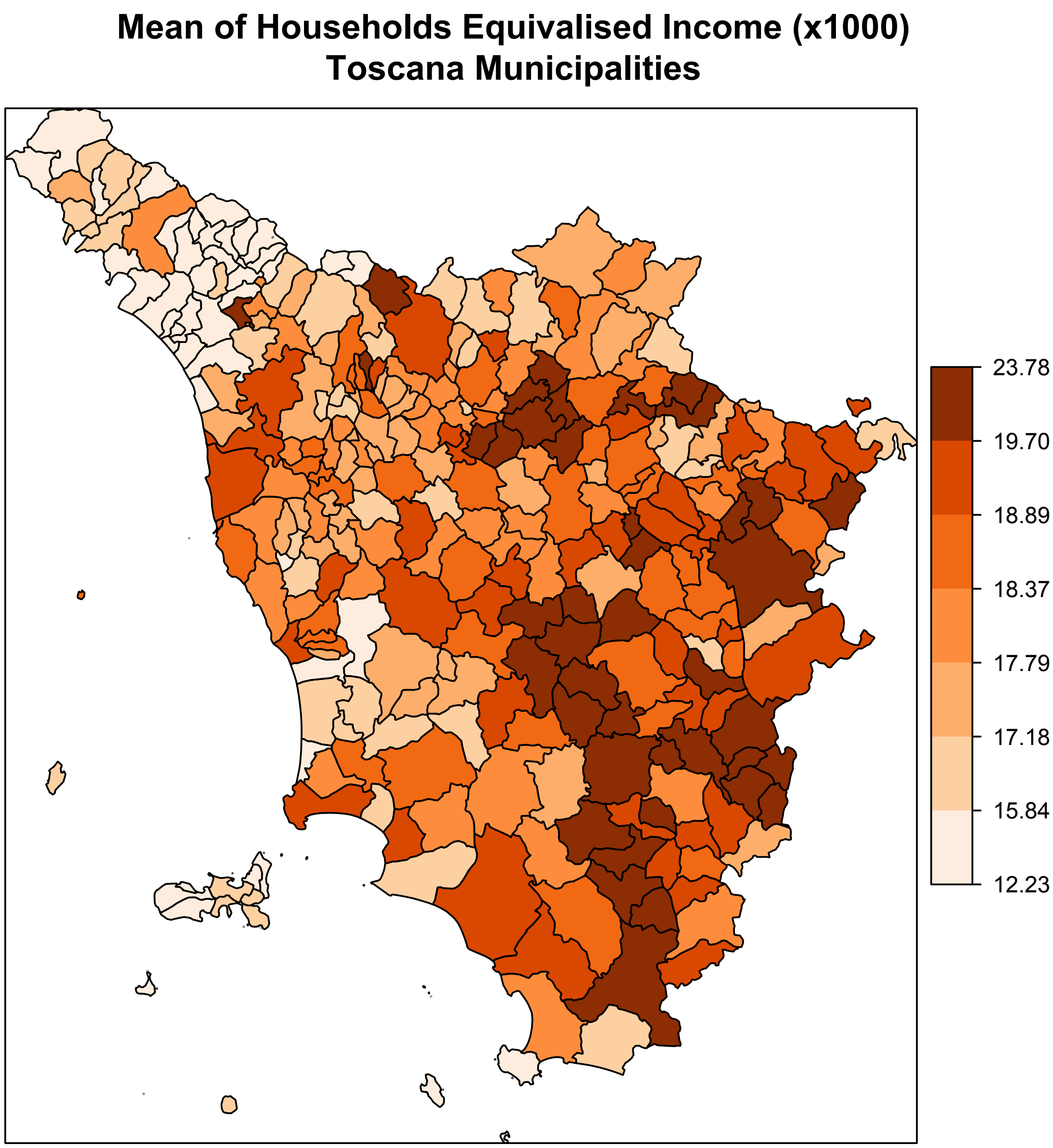
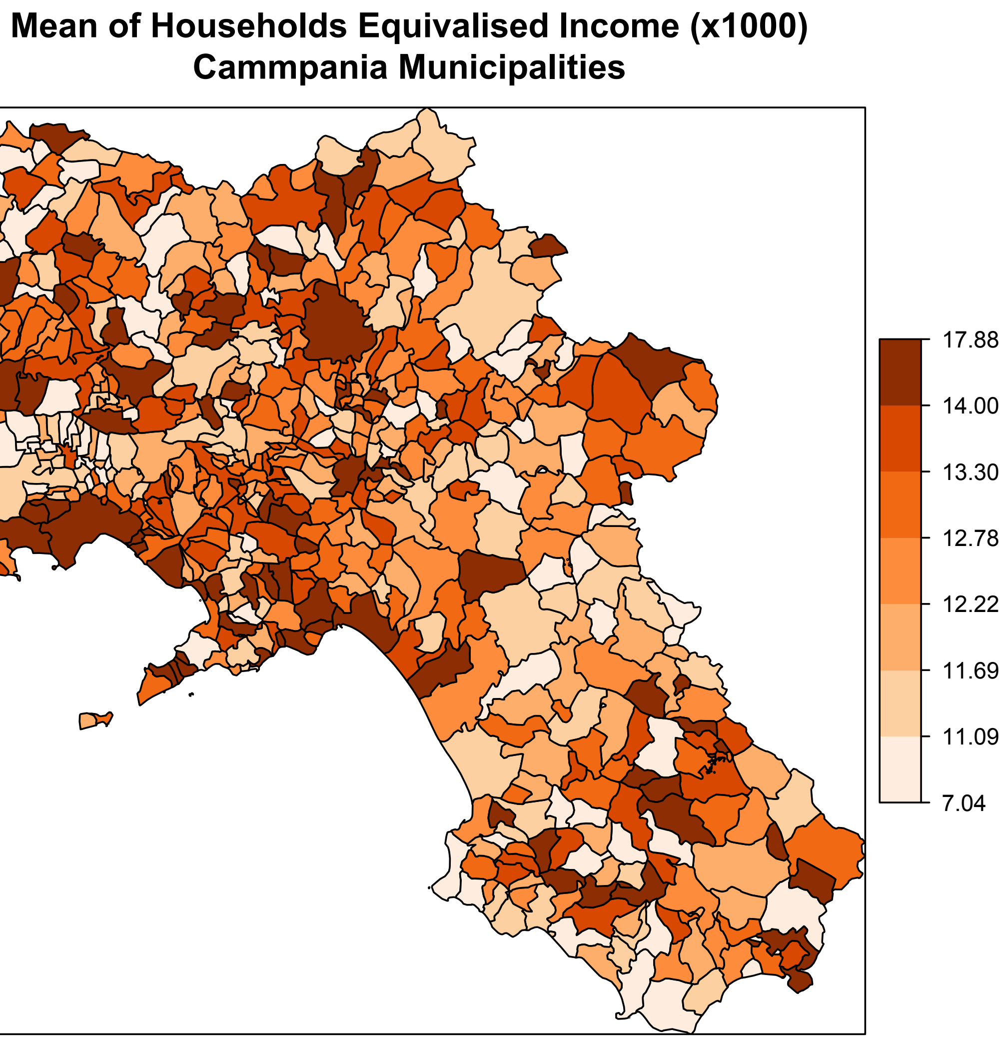


Figure 11.12: Estimated Mean of Household Equivalised Income (x1000) for Campania Municipalities. Semiparametric Fay and Herriot model.



Chapter 12

Estimating Small Area Income Distribution Functions

In the previous chapter we presented small area methodologies for estimating average income. In this chapter we describe the application of small area methodologies for estimating small area distribution functions of household equivalised income for provinces in the regions of Lombardia, Toscana and Campania. Estimating the income distribution function, in addition to small area averages, offers a more comprehensive picture of the wealth and the potential inequalities in the different provinces. The methodologies for estimating small area income distribution functions and the target geographies are listed below.

- M-quantile model for estimating the 25th,50th and 75th percentiles of the income distribution at the Province level
- Nonparametric M-quantile model for estimating the 25th,50th and 75th percentiles of the income distribution at the Province level

12.1 M-quantile Models

The working M-quantile small area model is the same to the model we used for estimating small area averages. Using the M-quantile model, estimates of the percentiles of the household equivalised income distributions for provinces in the three Italian regions (Lombardia, Toscana and Campania) are obtained using estimator (10.10). Corresponding estimates of the Mean Squared Error are obtained by applying the non-parametric bootstrap estimator (10.13). The expressions of the point and MSE estimators can be found in Chapter 10 and also in Chapter 10 of Deliverable 12 and 16.

The results (point and MSE estimation) from the application of the M-quantile model are mapped in Figures 12.1, 12.2, 12.3, 12.4, 12.5, 12.6 and 12.7, 12.8, 12.9 and also presented in Tables 12.1, 12.2, 12.3. The first noticeable result is that the estimated median income in provinces is lower than the corresponding average income. This illustrates the asymmetry of the income distribution and motivates the estimation of small area income distribution functions. Estimating income distribution functions further paves the way of making comparisons between Italian regions. For example, it should be noted that the 25th income percentile in provinces of Lombardia is comparable to the median income of provinces in Campania. The comparison of Lombardia and Toscana is now also easier. Certain provinces in

Toscana appear to have higher gap between the 25th and 75th percentiles of income, see for example, the provinces of Massa-Carrara, Lucca and Grosseto. These provinces are similar to the province of Sondrio in Lombardia. With regards to the wealthier provinces, the province of Siena is comparable to the province of Milano. Finally, in the region of Lombardia, it is interesting to note that provinces that appear to be very wealthy in terms of average income, also appear to have a wide gap between the 25th and 75th percentiles, see for example the provinces of Pavia and Lecco. The results from the estimation of the small area distribution function indicate that relying solely on estimates of average income does not always provide an accurate picture of the wealth of a small area.

Table 12.1: Estimated Quartiles of Household Equivalised Income (Qu.1st, Qu. 2nd and Qu. 3rd) and estimated Root Mean Squared Error of the Mean estimator (RMSE.Qu.1st, RMSE.Qu. 2nd and RMSE.Qu. 3rd) for the Lombardia Provinces. M-quantile model.

	PROVINCE	Qu.1st	RMSE.Qu.1st	Qu.2nd	RMSE.Qu.2nd	Qu.3rd	RMSE.Qu.3rd
1	VARESE	11913.80	510.97	17170.63	498.64	23638.95	723.29
2	COMO	10542.35	591.88	16096.70	709.98	22553.48	920.84
3	SONDRIO	9945.13	1227.77	14664.88	1227.45	19985.80	1839.20
4	MILANO	12617.37	278.01	18569.53	284.97	25412.03	454.00
5	BERGAMO	10378.91	500.24	15997.96	531.00	22975.81	811.78
6	BRESCIA	10087.11	512.11	15205.31	561.61	21188.61	770.20
7	PAVIA	10574.28	920.16	16138.38	1033.24	22796.23	1560.62
8	CREMONA	10867.93	875.52	15625.72	938.05	21829.22	1434.26
9	MANTOVA	11105.03	476.97	16323.66	521.35	22432.72	797.68
10	LECCO	11851.26	809.58	17582.24	792.87	24482.33	1212.89
11	LODI	11417.02	925.51	16707.56	991.17	22515.91	1495.27

Figure 12.1: Estimated First Quartiles (Root Mean Squared Error) of Household Equivalised Income for Lombardia Provinces. M-quantile model.

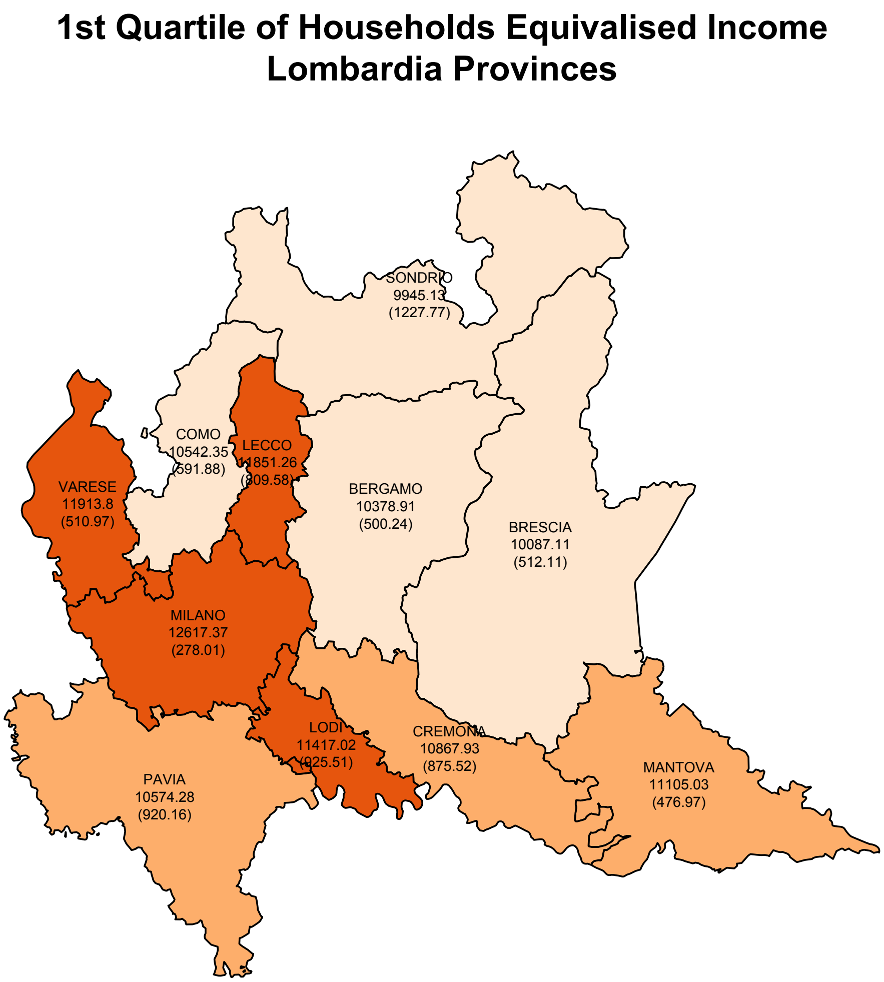


Figure 12.2: Estimated Second Quartiles (Root Mean Squared Error) of Household Equivalised Income for Lombardia Provinces. M-quantile model.

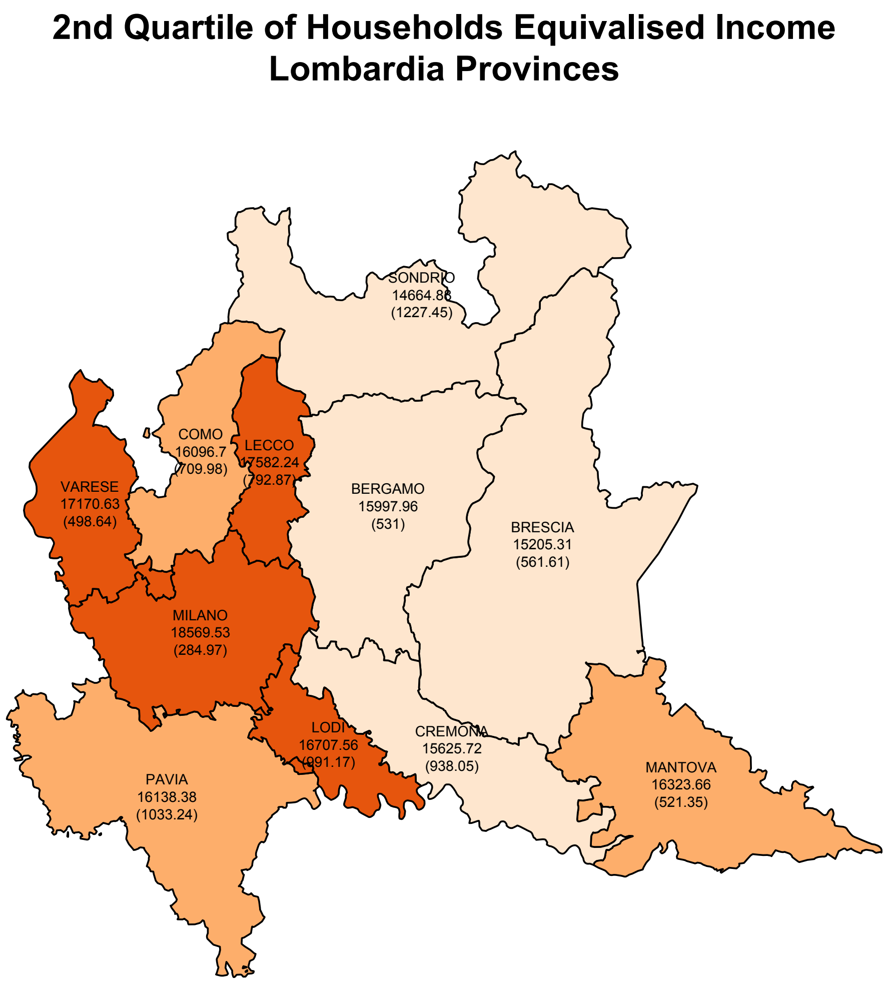


Figure 12.3: Estimated Third Quartiles (Root Mean Squared Error) of Household Equivalised Income for Lombardia Provinces. M-quantile model.

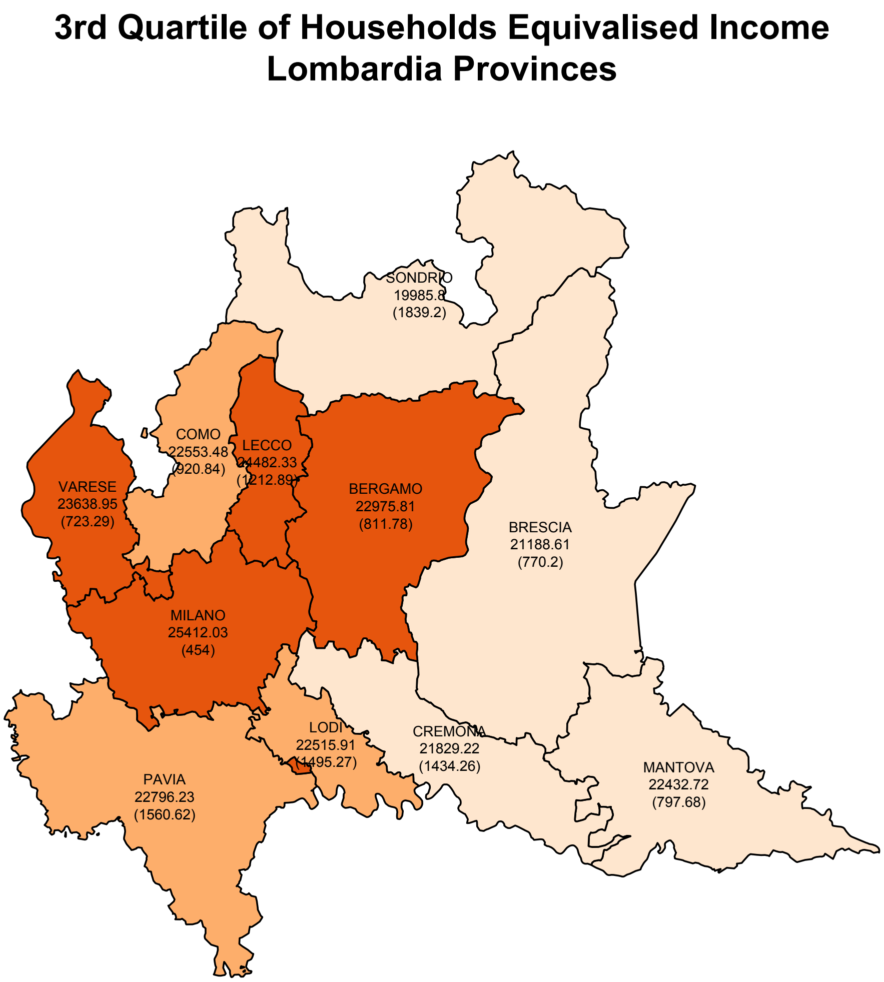


Table 12.2: Estimated Quartiles of Household Equivalised Income (Qu.1st, Qu. 2nd and Qu. 3rd) and estimated Root Mean Squared Error of the Mean estimator (RMSE.Qu.1st, RMSE.Qu. 2nd and RMSE.Qu. 3rd) for the Toscana Provinces. M-quantile model.

	PROVINCE	Qu.1st	RMSE.Qu.1st	Qu.2nd	RMSE.Qu.2nd	Qu.3rd	RMSE.Qu.3rd
1	MASSA CARRARA	8837.42	712.33	13498.69	831.75	18528.25	1164.96
2	LUCCA	9715.74	640.71	14733.29	690.70	20650.60	1087.89
3	PISTOIA	11412.26	669.47	16124.78	685.77	22243.96	1017.31
4	FIRENZE	12628.02	335.54	17364.81	377.37	23328.47	547.68
5	LIVORNO	11338.34	610.02	16662.83	701.85	22991.71	983.31
6	PISA	11571.98	618.04	17161.42	681.67	23867.96	989.89
7	AREZZO	12205.01	578.72	16724.22	638.74	22100.72	949.08
8	SIENA	12639.00	662.31	18373.53	703.94	25471.22	1087.76
9	GROSSETO	9924.80	924.38	15456.41	1016.58	22069.22	1483.27
10	PRATO	12779.53	669.54	16968.74	708.72	21796.88	1101.34

Figure 12.4: Estimated First Quartiles (Root Mean Squared Error) of Household Equivalised Income for Toscana Provinces. M-quantile model.

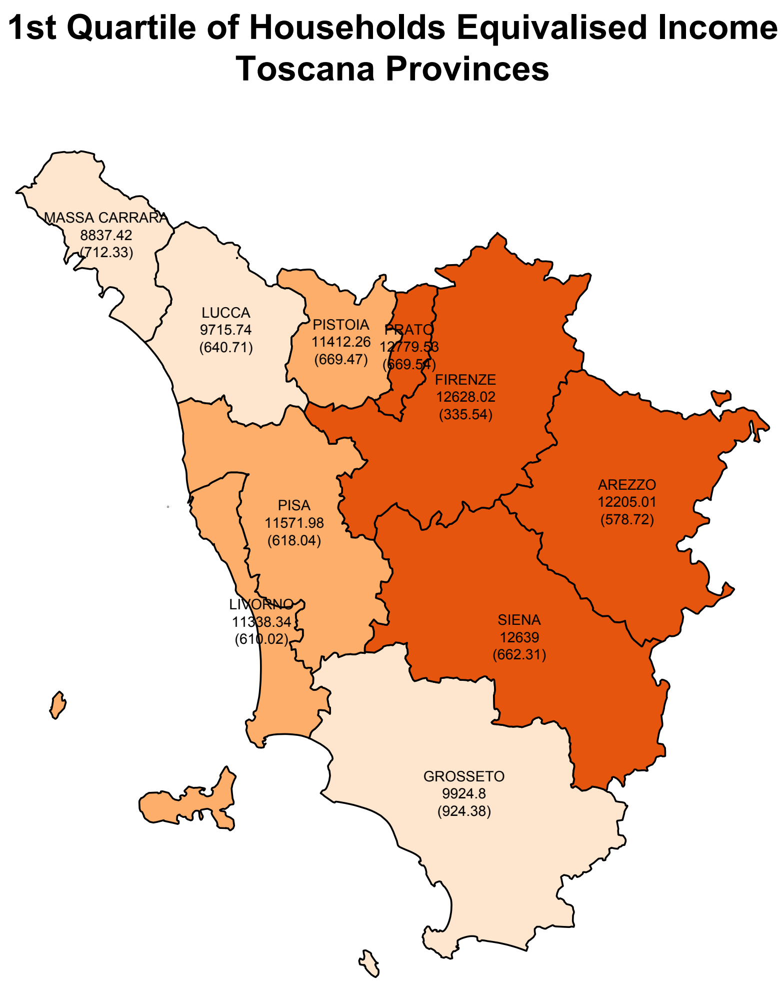


Figure 12.5: Estimated Second Quartiles (Root Mean Squared Error) of Household Equivalised Income for Toscana Provinces. M-quantile model.

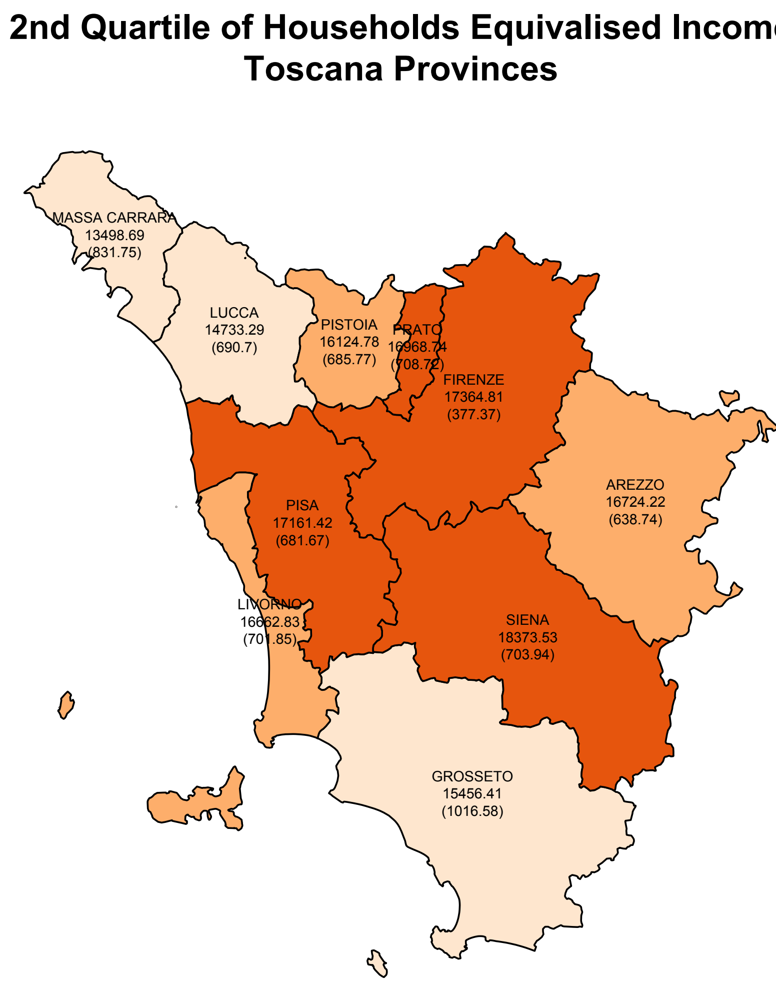


Figure 12.6: Estimated Third Quartiles (Root Mean Squared Error) of Household Equivalised Income for Toscana Provinces. M-quantile model.

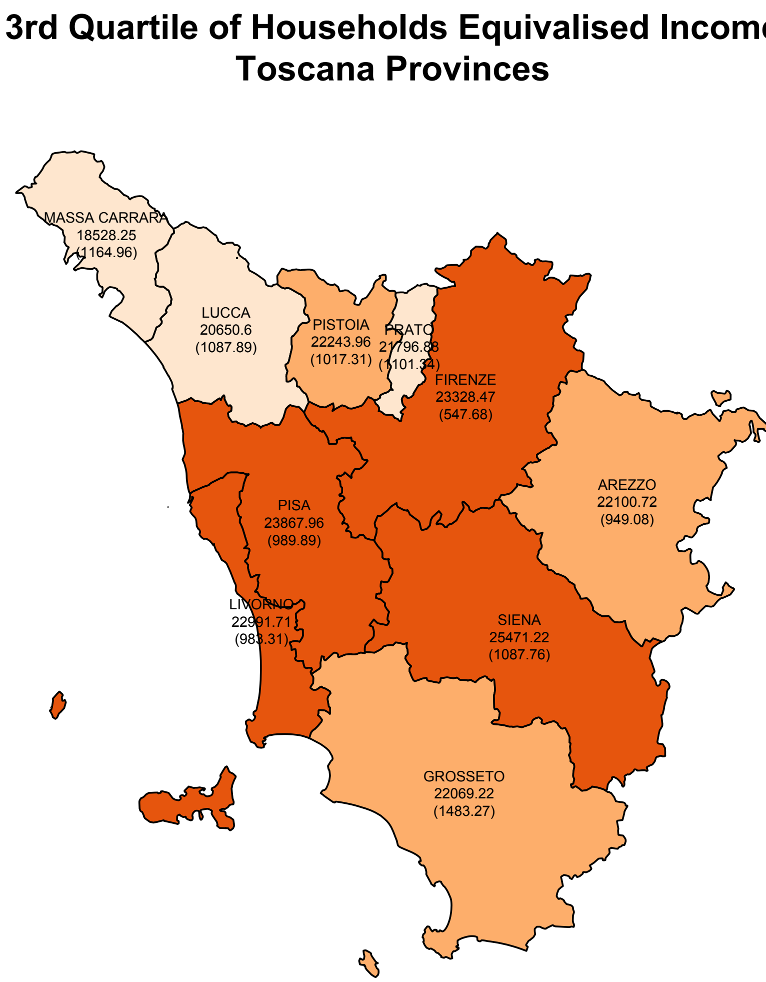


Table 12.3: Estimated Quartiles of Household Equivalised Income (Qu.1st, Qu. 2nd and Qu. 3rd) and estimated Root Mean Squared Error of the Mean estimator (RMSE.Qu.1st, RMSE.Qu. 2nd and RMSE.Qu. 3rd) for the Campania Provinces. M-quantile model.

	PROVINCE	Qu.1st	RMSE.Qu.1st	Qu.2nd	RMSE.Qu.2nd	Qu.3rd	RMSE.Qu.3rd
1	CASERTA	6053.91	470.72	11097.78	547.79	16668.83	696.32
2	BENEVENTO	5856.16	723.83	9751.76	800.80	14609.27	1048.03
3	NAPOLI	6687.48	219.28	11546.04	224.02	17223.05	308.15
4	AVELLINO	6568.94	663.43	11148.13	713.89	17468.50	995.03
5	SALERNO	7410.94	439.10	12010.06	474.30	17249.39	624.75

Figure 12.7: Estimated First Quartiles (Root Mean Squared Error) of Household Equivalised Income for Campania Provinces. M-quantile model.

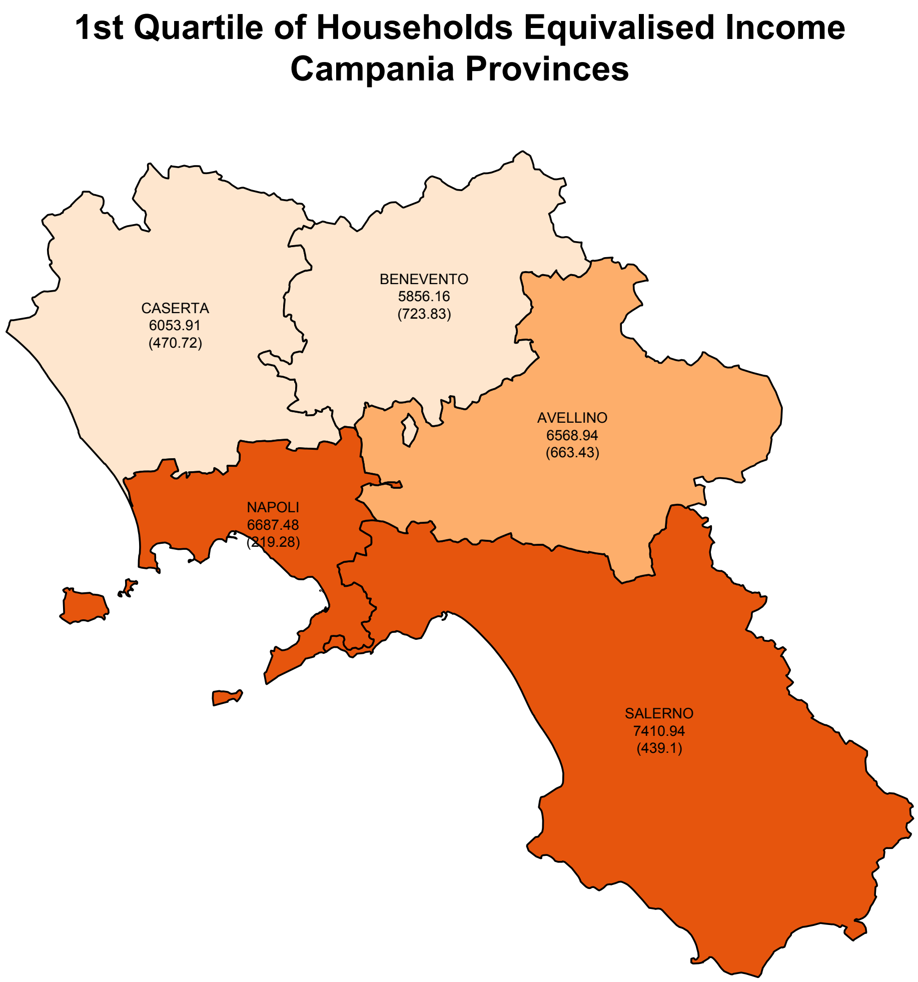


Figure 12.8: Estimated Second Quartiles (Root Mean Squared Error) of Household Equivalised Income for Campania Provinces. M-quantile model.

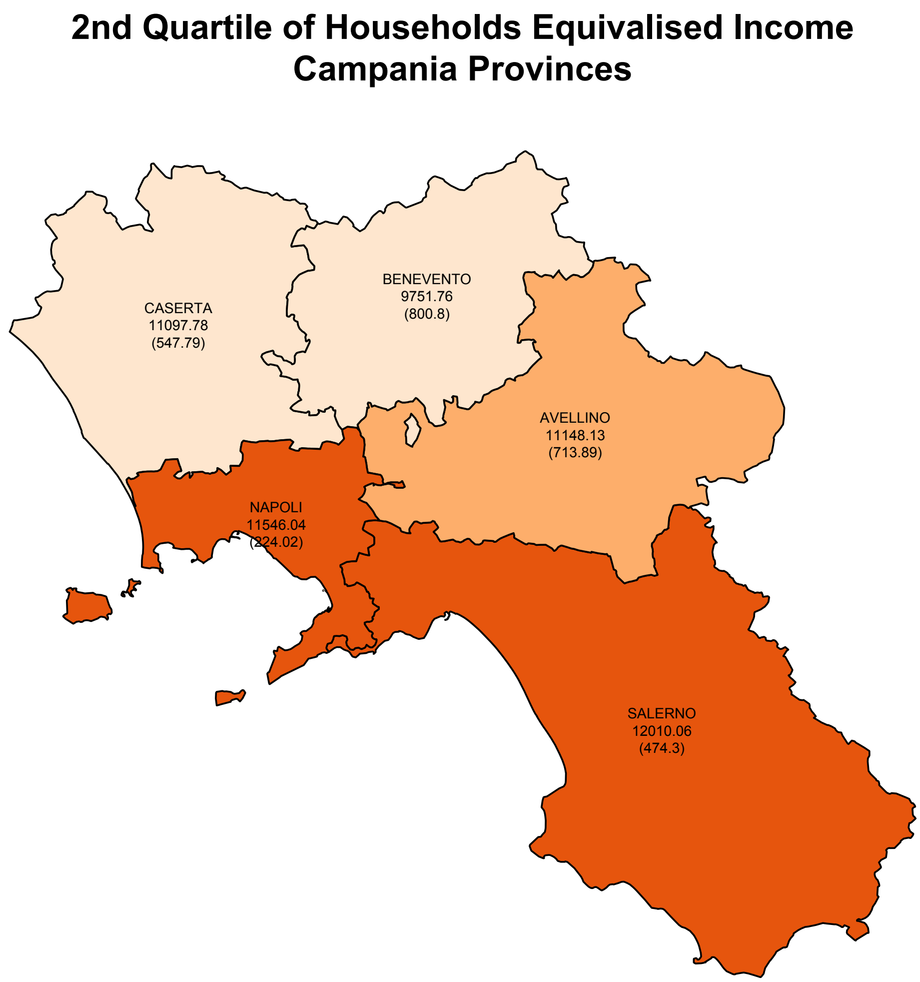
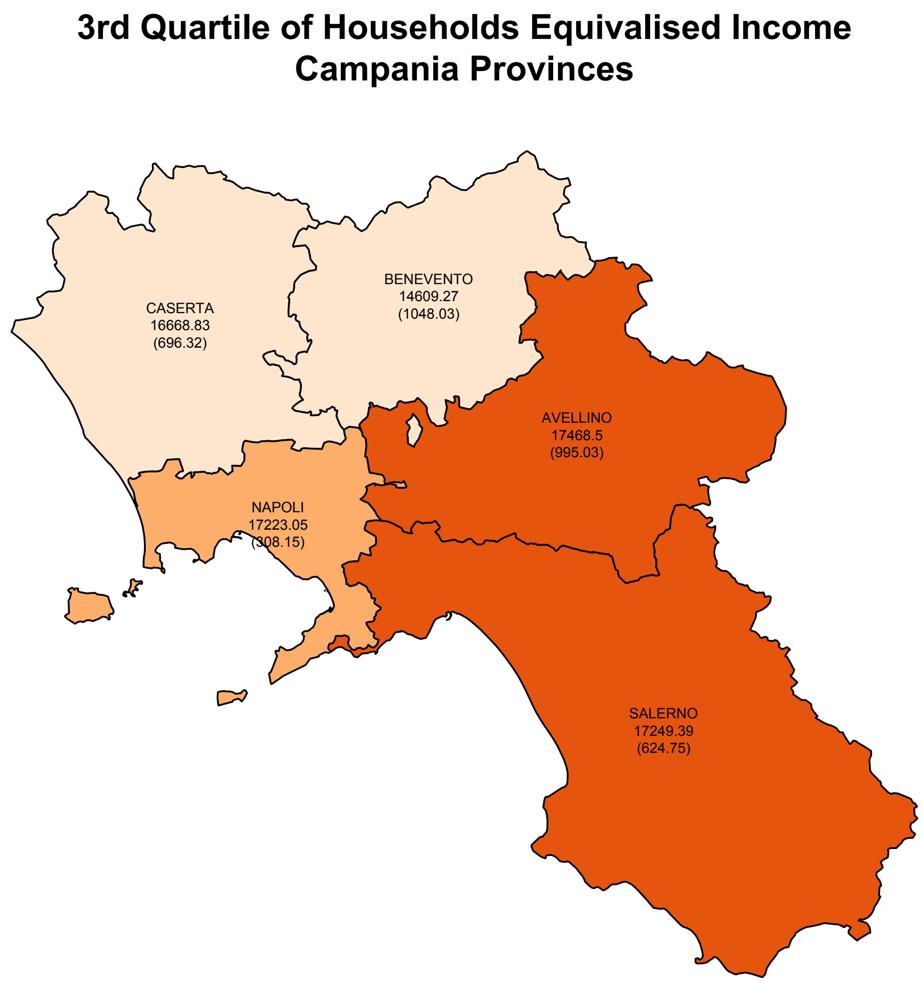


Figure 12.9: Estimated Third Quartiles (Root Mean Squared Error) of Household Equivalised Income for Campania Provinces. M-quantile model.



12.2 Nonparametric M-quantile Model

In the section we extend the analysis with the nonparametric M-quantile model already described in section 12.1 to the estimation of household income quartiles, using estimator (10.25) presented in Chapter 10. As with the case of the household average income, the results for the income quartiles obtained using the parametric and nonparametric specifications of M-quantile models are very similar in the three target regions. Also the gap present in the income distribution described in section 12.1 between the regions remains the same. This is explained by the almost linear relationship between the household equivalised income and the covariates.

As an additional comment, we can notice that in Lombardia the provinces of Varese, Milano and Lecco are characterized by the highest class of estimated income for each of the three quartiles, as we can see in Figures 12.10, 12.11, 12.12 and Table 12.4. The same happens for the provinces of Firenze and Siena in Toscana (see Figures 12.13, 12.14, 12.15 and Table 12.5) and for the province of Napoli in Campania, as we can see in Figures 12.16, 12.17, 12.18 and Table 12.6. On the contrary, the estimates referring to the other provinces have a different relative ranking passing from a quartile to another: for example, the province of Arezzo in Toscana is in the highest class as concerns the estimate of the 1st quartile but it is in a lower class for the 2nd and 3rd quartiles. These results show the importance of estimating the income quartiles, together with the income averages, to get a more complete picture of the wealth in the target areas at different points of the income distribution.

Table 12.4: Estimated Quartiles of Household Equivalised Income (Qu.1st, Qu. 2nd and Qu. 3rd) for the Lombardia Provinces. Nonparametric M-quantile model.

results	PROVINCE	Qu.1st	Qu.2nd	Qu.3rd
1	VARESE	11988.87	17086.34	23595.30
2	COMO	10716.31	16215.10	22430.04
3	SONDRIO	10283.96	14871.41	20262.95
4	MILANO	12738.57	18756.50	25422.99
5	BERGAMO	10467.52	16078.10	22999.28
6	BRESCIA	10407.63	15287.39	21180.40
7	PAVIA	10535.96	16229.13	23183.57
8	CREMONA	10519.25	15574.03	21598.16
9	MANTOVA	11162.35	16316.57	22459.54
10	LECCO	11928.49	17538.00	24473.92
11	LODI	11666.65	16926.19	22268.77

Figure 12.10: Estimated First Quartiles of Household Equivalised Income for Lombardia Provinces. Nonparametric M-quantile model.

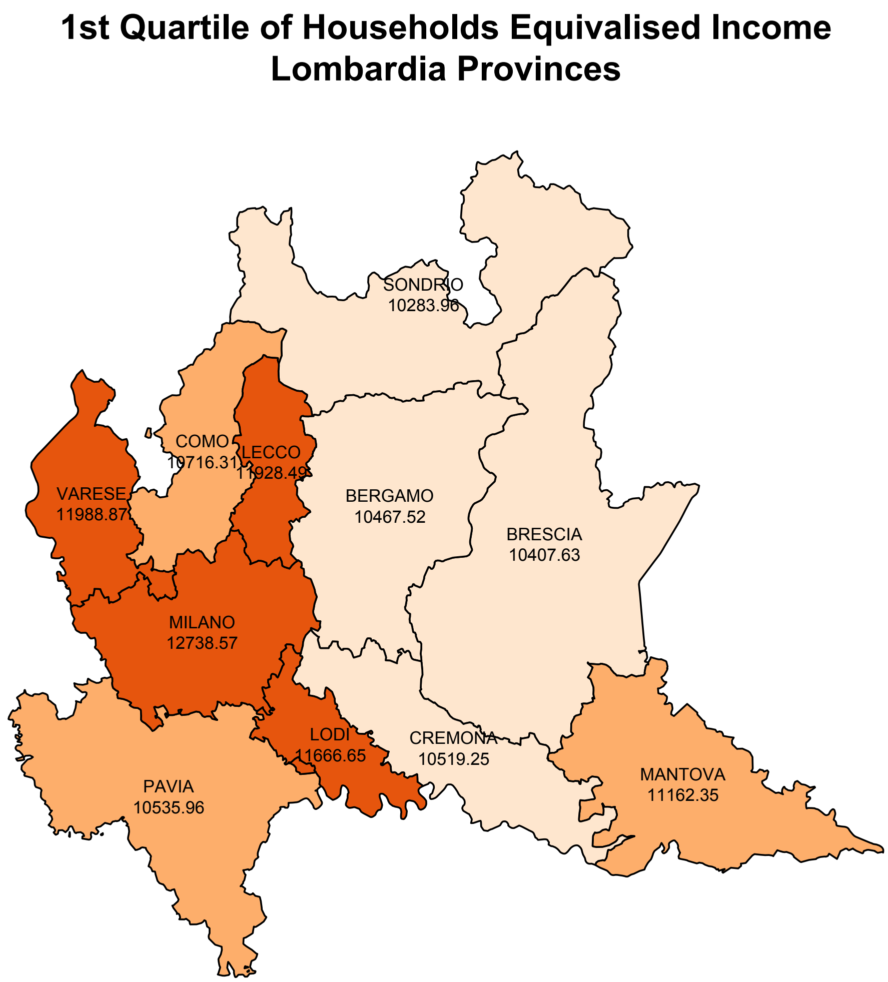


Figure 12.11: Estimated Second Quartiles of Household Equivalised Income for Lombardia Provinces. Nonparametric M-quantile model.

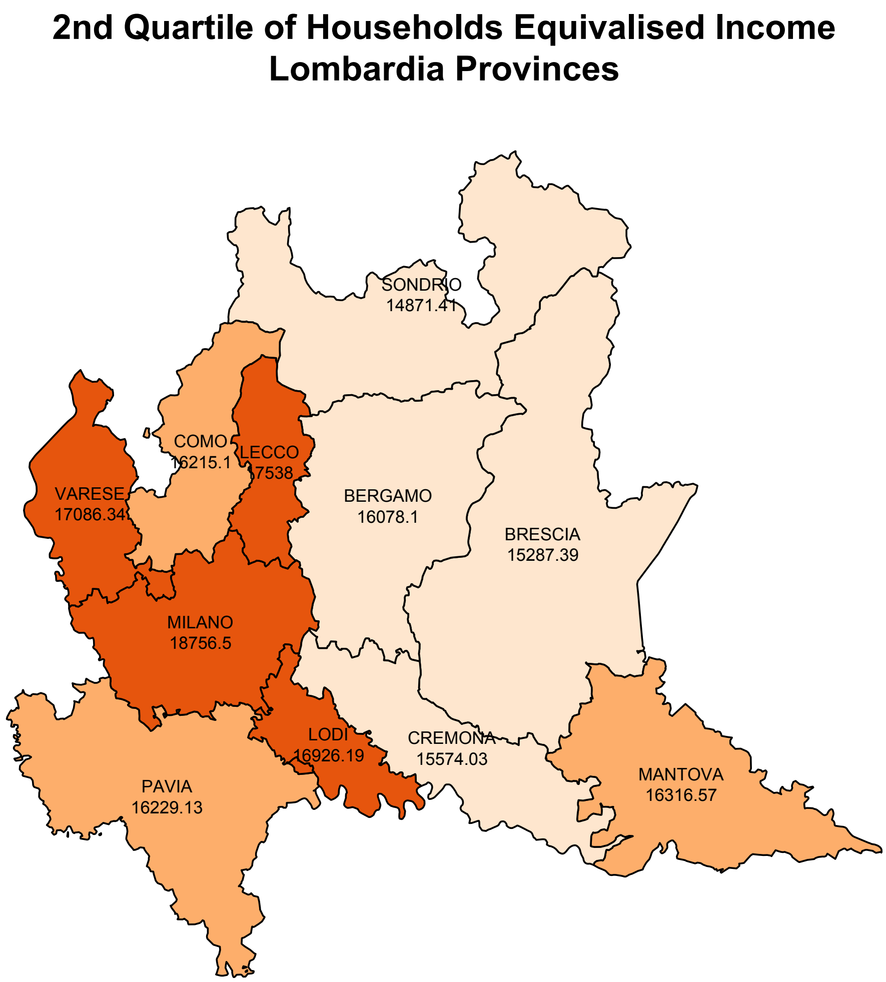


Figure 12.12: Estimated Third Quartiles of Household Equivalised Income for Lombardia Provinces. Nonparametric M-quantile model.

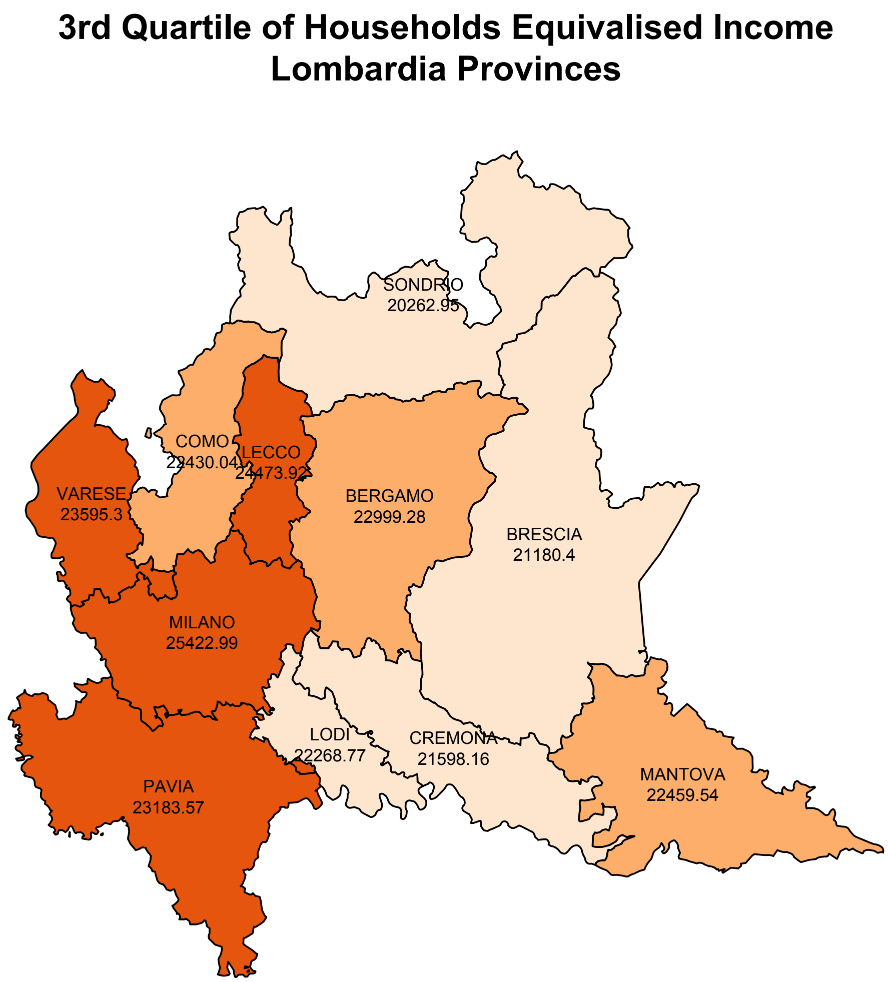


Table 12.5: Estimated Quartiles of Household Equivalised Income (Qu.1st, Qu. 2nd and Qu. 3rd) for the Toscana Provinces. Nonparametric M-quantile model.

results	PROVINCE	Qu.1st	Qu.2nd	Qu.3rd
1	MASSA CARRARA	8784.95	13276.74	18303.39
2	LUCCA	9764.79	14809.48	20652.12
3	PISTOIA	11416.86	16417.14	22536.24
4	FIRENZE	12724.62	17378.57	23194.72
5	LIVORNO	11113.53	16644.73	22558.18
6	PISA	11577.20	16929.11	23539.89
7	AREZZO	11881.20	16633.32	22216.63
8	SIENA	12412.42	18237.81	25227.04
9	GROSSETO	9892.01	15519.11	21714.83
10	PRATO	13034.21	17163.42	21888.17

Figure 12.13: Estimated First Quartiles of Household Equivalised Income for Toscana Provinces. Nonparametric M-quantile model.

Figure 12.14: Estimated Second Quartiles of Household Equivalised Income for Toscana Provinces. Nonparametric M-quantile model.

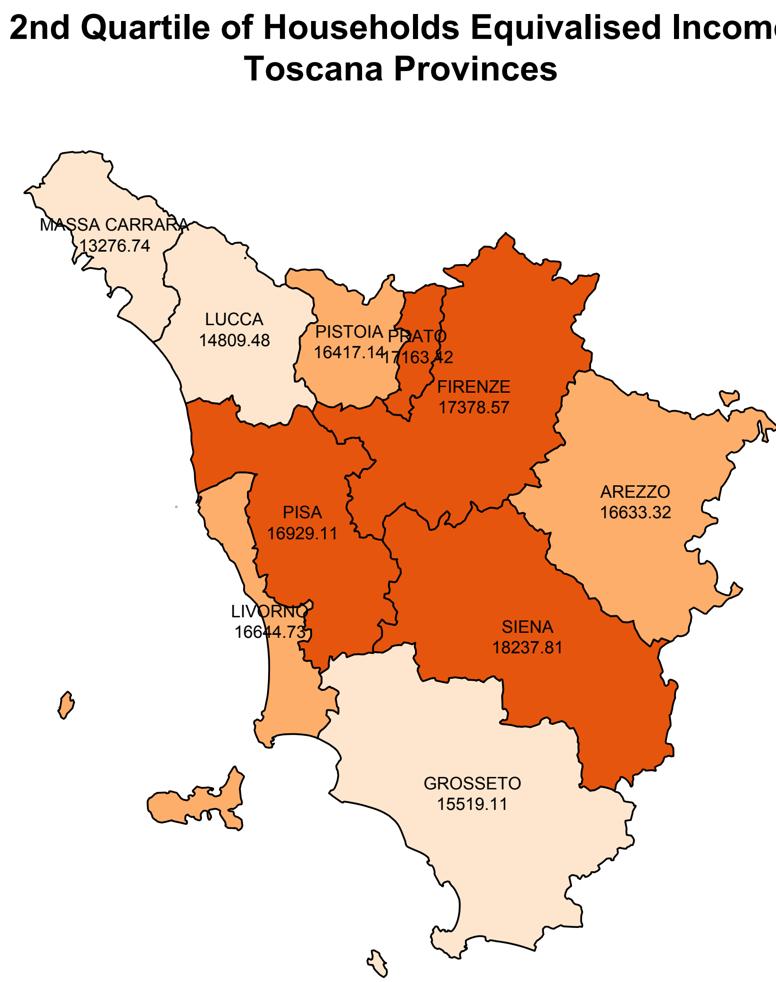


Figure 12.15: Estimated Third Quartiles of Household Equivalised Income for Toscana Provinces. Nonparametric M-quantile model.

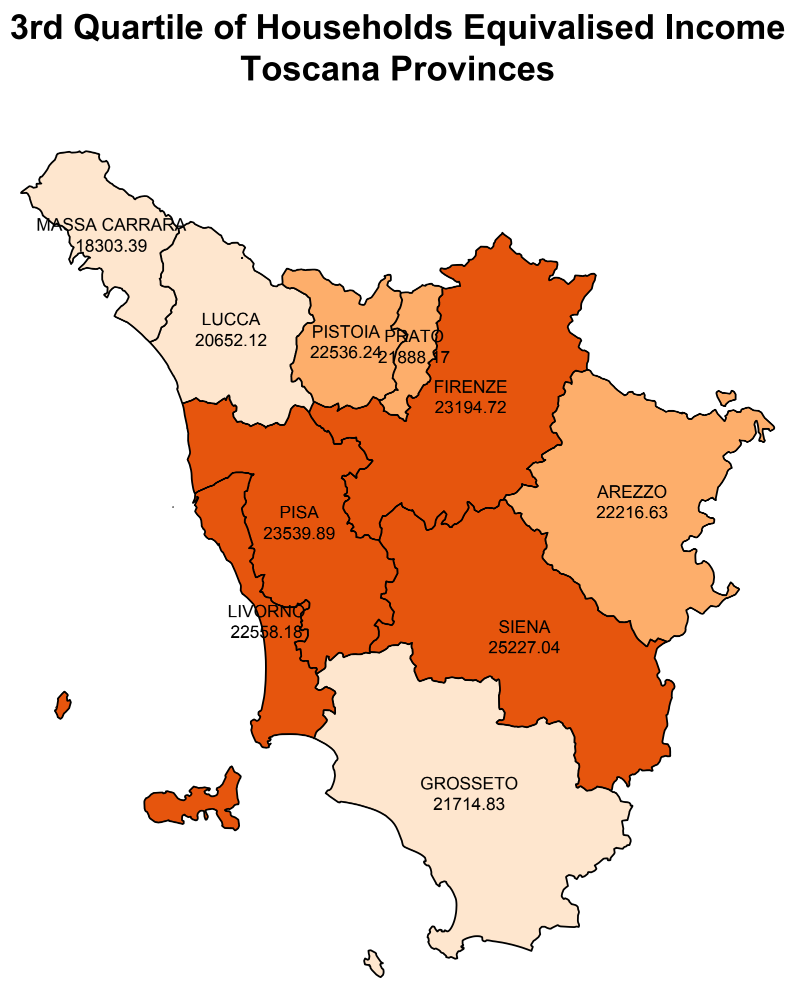


Table 12.6: Estimated Quartiles of Household Equivalised Income (Qu.1st, Qu. 2nd and Qu. 3rd) for the Campania Provinces. Nonparametric M-quantile model.

results	PROVINCE	Qu.1st	Qu.2nd	Qu.3rd
1	CASERTA	6153.17	11094.48	16619.98
2	BENEVENTO	5946.44	9883.93	14734.74
3	NAPOLI	6788.77	11671.06	17302.27
4	AVELLINO	6719.47	11431.13	17738.89
5	SALERNO	7390.11	11996.88	17285.07

Figure 12.16: Estimated First Quartiles of Household Equivalised Income for Campania Provinces. Nonparametric M-quantile model.

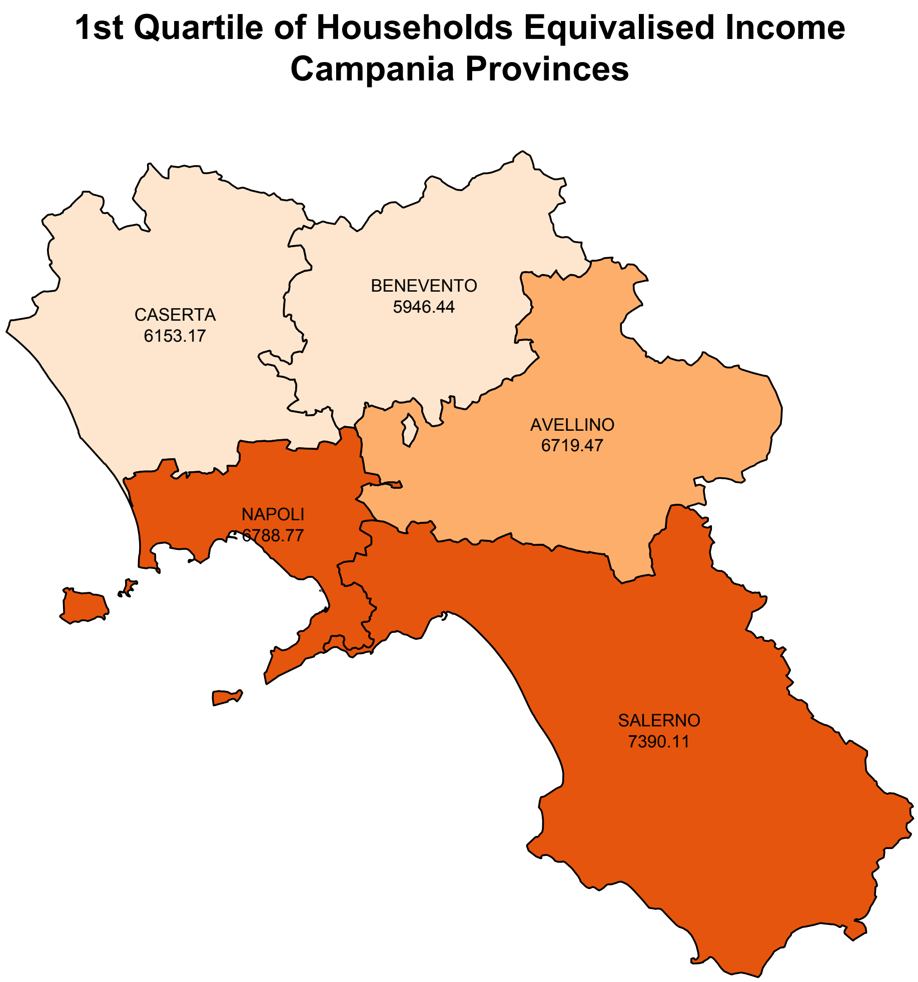


Figure 12.17: Estimated Second Quartiles of Household Equivalised Income for Campania Provinces. Nonparametric M-quantile model.

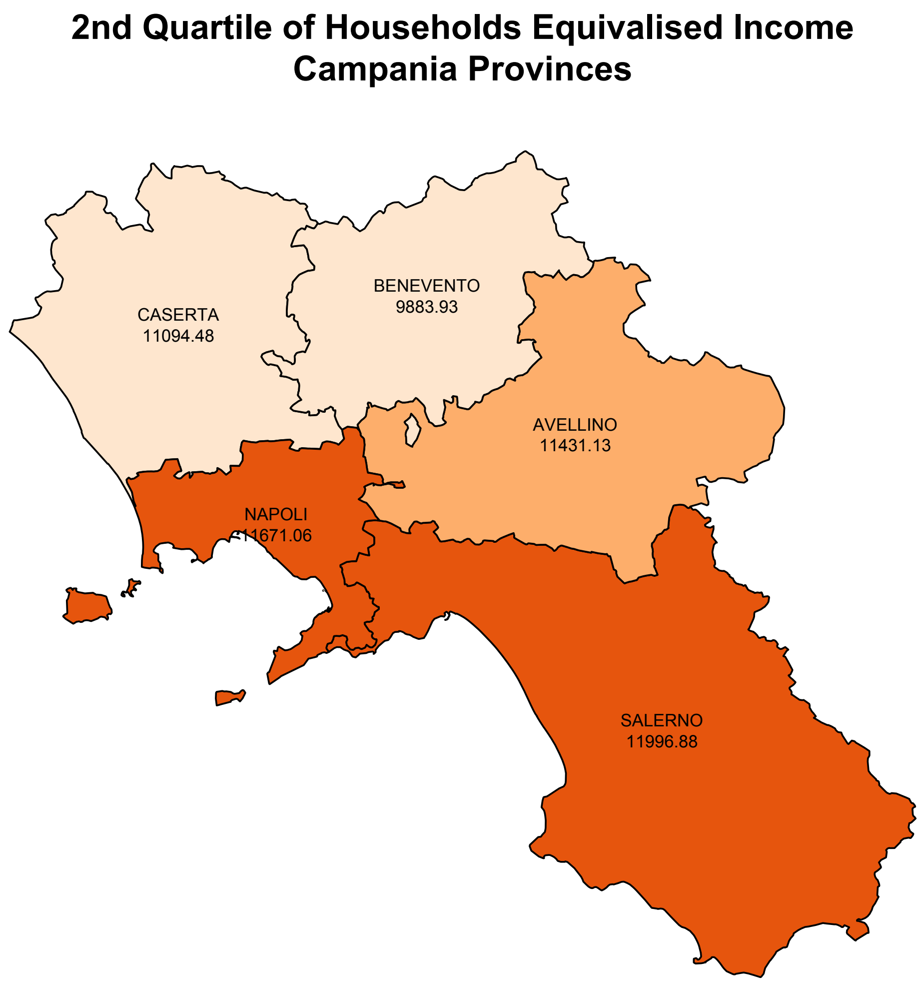
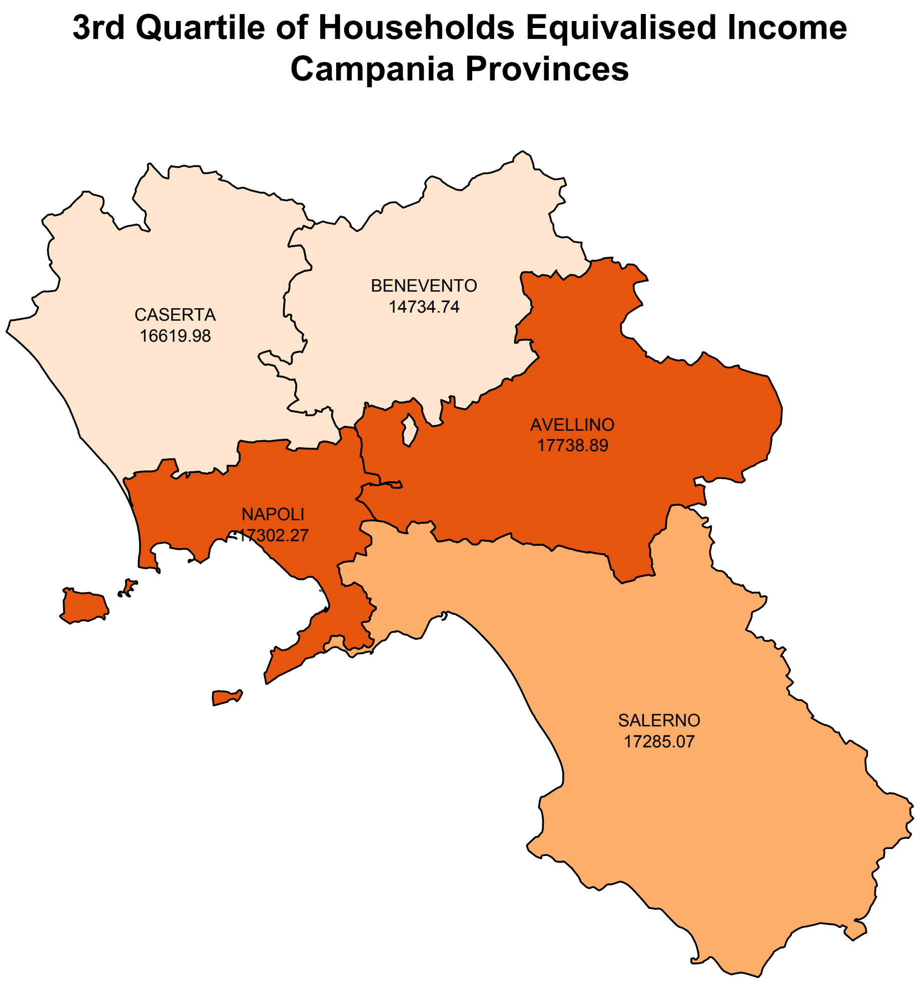


Figure 12.18: Estimated Third Quartiles of Household Equivalised Income for Campania Provinces. Nonparametric M-quantile model.



Chapter 13

Estimating Poverty Indicators

In Chapter 10 of Deliverables 12 and 16 we focused on the estimation of poverty indicators using M-quantile small area models. Indeed, poverty estimation often corresponds to a special case of quantile estimation since in this case we are interested in estimating the number of individuals/households below a given income threshold (poverty line). Since the M-quantile method does not impose strong distributional assumptions and is outlier robust, the use of the M-quantile model for poverty estimation may protect us against departures from assumptions of the unit-level nested error regression model. In particular, in Deliverables 12 and 16 we considered the estimation for two popular poverty indicators, namely the Head Count Ratio (HCR) and the Poverty Gap (PG), and using a model-based and a design-based simulation study we compared the performance of the M-quantile estimators to that of the recently proposed estimators by Molina and Rao (2009) under the Empirical Best Prediction (EBP) approach. Furthermore, in Chapter 10 of deliverable 12 and 16 we considered also the so called fuzzy approach to the poverty measure. In particular we focused on the so called Fuzzy Monetary Index (FM) and we compared the M-quantile estimators to the Empirical Best estimator (proposed by Molina and Rao, 2009) using a model-based simulation. In this chapter we describe the application of small area methodologies for estimating three poverty indicators, the Head Count Ratio (HCR), the Poverty Gap (PG) and the Fuzzy Monetary Index (FM) for provinces in the regions of Lombardia, Toscana and Campania.

The HCR indicator is a widely used measure of poverty. The popularity of this indicator is due to its ease of construction and interpretation. At the same time this indicator also assumes that all poor household/individuals are in the same situation. For example, the easiest way of reducing the headcount index is by targeting benefits to people just below the poverty line because they are the ones who are cheapest to move across the line. Hence, policies based on the headcount index might be sub-optimal. For this reason we also obtain estimates of the PG indicator. The PG can be interpreted as the average shortfall of poor people. It shows how much would have to be transferred to the poor to bring their expenditure up to the poverty line.

As an alternative to the HCR and PG indicators, the FM considers poverty as a matter of degree rather than an attribute that is simply present or absent for individuals in the population (Betti et al. 2009). Fuzzy indicators have been made such that they vary from 0 to 1, where 0 indicates the richest person in the population while 1 indicates the poorest.

A short description of the methodology for poverty indicators estimation is also given in Chapter 10 of this deliverable.

The methodologies for estimating poverty indicators and the target geographies are listed below.

- M-quantile model for estimating the HCR and PG indicators at the Province level
- M-quantile model for estimating the FM indicator at the Province level

For producing poverty estimates we use region-specific poverty lines defined as 60% of the regional median income. The poverty lines for Lombardia, Tuscany and Campania are resepctively equal to 10322, 10248 and 6923 Euros. Although we use different poverty lines, comparisons between Lombardia and Tuscany are possible due to the similarity in the poverty lines of these two regions.

13.1 M-quantile Models

The working M-quantile small area model is the same to the one we used for estimating small area averages. Using the M-quantile model, we obtain estimates of the HCR, PG and FM indicators for provinces in the three Italian regions (Lombardia, Toscana and Campania) using the procedure outlined in Section 10.2.2 and 10.2.3 of Chapter 10. Corresponding estimates of the Mean Squared Error are obtained by applying the non-parametric bootstrap estimator (10.13). The results about HCR and PG (point and MSE estimation) from the application of the M-quantile model are mapped in Figures 13.1, 13.2 and 13.3 and also presented in Tables 13.1, 13.2 and 13.3, while results about FM (point estimation) are mapped in Figures 13.4, 13.5, 13.6 and Tables 13.4, 13.5, 13.6.

The HCR in Lombardia ranges from 0.172 (Milano) to 0.27 (Sondrio) while the PG for provinces in the same region ranges from 0.073 (Milano) to 0.124 (Sondrio). For provinces in Toscana the HCR ranges from 0.28 (Massa-Carrara) to 0.161 (Siena) and the PG in the same region from 0.117 (Massa-Carrara) to 0.06 (Siena). Finally, for provinces in the region of Campania the HCR ranges from 0.238 (Salerno) to 0.280 (Benevento) and the PG from 0.127 (Salerno) to 0.161 (Caserta). The picture that emerges is as expected i.e. Campania is a region that has consistently higher poverty than Toscana and Lombardia. Taking into account the lower poverty line used for producing the estimates in Campania, it becomes apparent that the differences in poverty between the North of Italy and the South of Italy are even more pronounced.

The use of the PG indicator significantly enhances the picture of wealth in the different regions. Noticeable are also some aspects of the comparison between the regions of Toscana and Lombardia. The analysis in the previous chapters indicated that Lombardia is somewhat wealthier than Toscana. However, looking at the estimates of HCR and PG a different picture emerges. Overall, provinces in Toscana have lower HCR and PG than provinces in Lombardia. For example, Pavia, one of the wealthiest provinces, in terms of average income and income distribution, in Lombardia appears to have higher poverty than a number of provinces in Toscana such as Siena and Florence. Of course, in these comparisons one must take into account the precision of the estimates and the fact that each region has its own poverty line (computed as 0.6 times the median of the equivalised household income of the region). Nevertheless, these results indicate that inequalities in Lombardia may be more pronounced than inequalities in Toscana.

The FM in Lombaria ranges from 0.157 (Milano) to 0.161 (Sondrio) and in all the other provinces is between 0.158 and 0.159. For the provinces in Toscana the FM ranges from 0.156 (Firenze, Pisa and Siena) to 0.159 (Massa-Carrara) and is 0.157 for all the other provinces with the exception of Grosseto

Table 13.1: Estimated Head Count Ratio (HCR) and Poverty Gap (PG), and estimated Root Mean Squared Error of the Head Count Ratio and Poverty Gap estimators (RMSE.HCR and RMSE.PG) for the Lombardia Provinces. M-quantile model.

	PROVINCE	HCR	RMSE.HCR	PG	RMSE.PG
1	VARESE	0.203	0.014	0.087	0.010
2	COMO	0.222	0.018	0.096	0.013
3	SONDRIO	0.270	0.037	0.124	0.028
4	MILANO	0.172	0.010	0.073	0.007
5	BERGAMO	0.224	0.015	0.097	0.011
6	BRESCIA	0.250	0.018	0.113	0.013
7	PAVIA	0.223	0.027	0.098	0.019
8	CREMONA	0.236	0.025	0.105	0.018
9	MANTOVA	0.226	0.015	0.100	0.011
10	LECCO	0.196	0.021	0.083	0.015
11	LODI	0.215	0.025	0.095	0.018

Table 13.2: Estimated Head Count Ratio (HCR) and Poverty Gap (PG), and estimated Root Mean Squared Error of the Head Count Ratio and Poverty Gap estimators (RMSE.HCR and RMSE.PG) for the Toscana Provinces. M-quantile model.

	PROVINCE	HCR	RMSE.HCR	PG	RMSE.PG
1	MASSA CARRARA	0.280	0.039	0.117	0.022
2	LUCCA	0.239	0.026	0.094	0.015
3	PISTOIA	0.195	0.019	0.073	0.011
4	FIRENZE	0.166	0.012	0.061	0.007
5	LIVORNO	0.193	0.020	0.075	0.012
6	PISA	0.175	0.018	0.065	0.010
7	AREZZO	0.182	0.018	0.068	0.010
8	SIENA	0.161	0.023	0.060	0.012
9	GROSSETO	0.231	0.029	0.093	0.019
10	PRATO	0.172	0.021	0.062	0.011

for which the FM is 0.158. Finally, for provinces in the region of Campania the FM ranges from 0.197 (Avellino and Salerno) to 0.199 (Benevento, Caserta and Napoli). We remind that lower values of the FM correspond to higher income level. As expected in Campania people are poorer than in Lombardia and Toscana.

These results illustrate that for constructing a good picture of the wealth locally, we must produce a wide range of small area statistics. Finally, the present work demonstrates how small area estimation and inference for such statistics can be implemented in practice.

Figure 13.1: Estimated Head Count Ratio (Root Mean Squared Error) and Poverty Gap (Root Mean Squared Error) for Lombardia Provinces. M-quantile model.

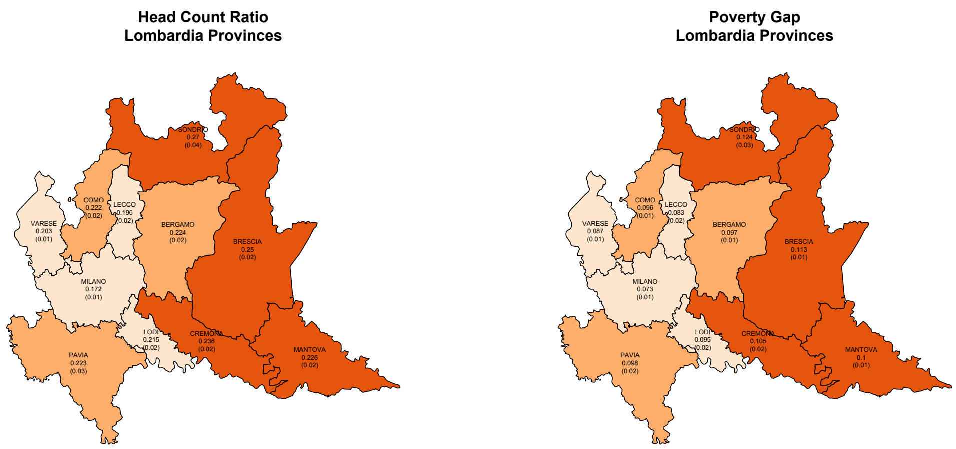


Figure 13.2: Estimated Head Count Ratio (Root Mean Squared Error) and Poverty Gap (Root Mean Squared Error) for Toscana Provinces. M-quantile model.

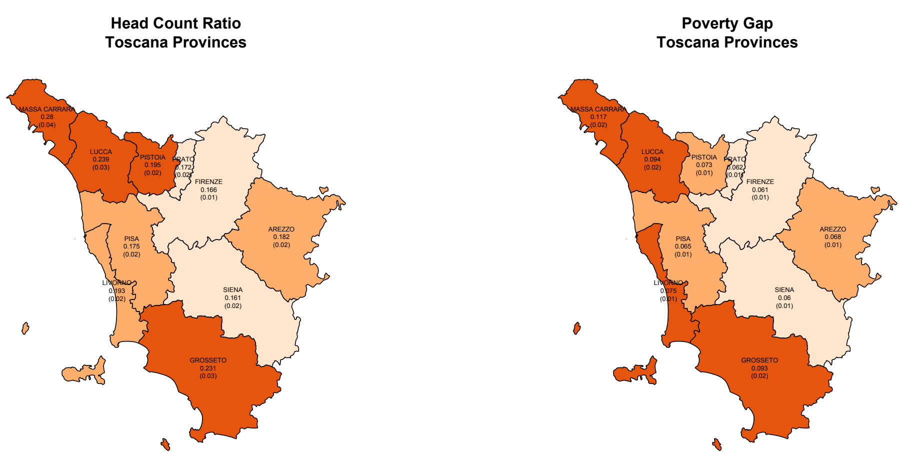


Table 13.3: Estimated Head Count Ratio (HCR) and Poverty Gap (PG), and estimated Root Mean Squared Error of the Head Count Ratio and Poverty Gap estimators (RMSE.HCR and RMSE.PG) for the Campania Provinces. M-quantile model.

	PROVINCE	HCR	RMSE.HCR	PG	RMSE.PG
1	CASERTA	0.277	0.018	0.161	0.017
2	BENEVENTO	0.280	0.033	0.153	0.026
3	NAPOLI	0.266	0.010	0.158	0.011
4	AVELLINO	0.245	0.024	0.131	0.020
5	SALERNO	0.238	0.016	0.127	0.014

Figure 13.3: Estimated Head Count Ratio (Root Mean Squared Error) and Poverty Gap (Root Mean Squared Error) for Campania Provinces. M-quantile model.

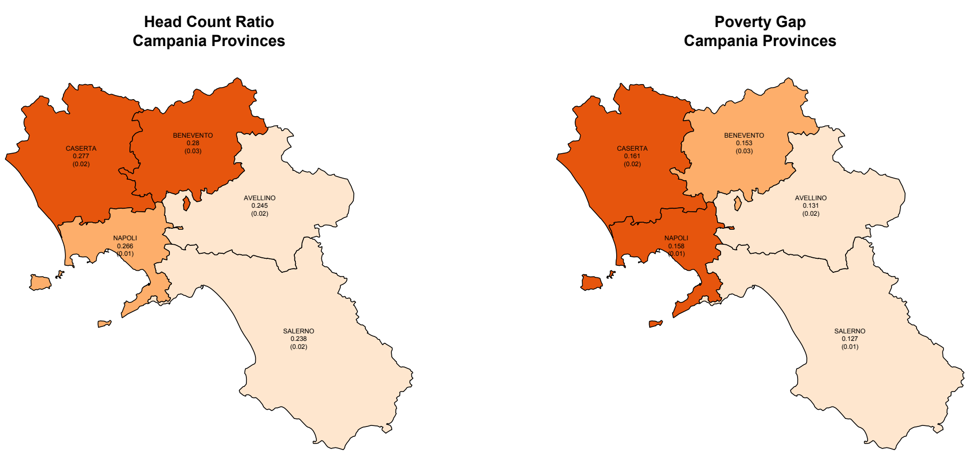


Table 13.4: Estimated Fuzzy Monetary Index (FM) for the Lombardia Provinces. M-quantile model.

results	PROVINCE	FM
1	VARESE	0.158
2	COMO	0.158
3	SONDRIO	0.161
4	MILANO	0.157
5	BERGAMO	0.158
6	BRESCIA	0.159
7	PAVIA	0.158
8	CREMONA	0.159
9	MANTOVA	0.159
10	LECCO	0.158
11	LODI	0.159

Table 13.5: Estimated Fuzzy Monetary Index (FM) for the Toscana Provinces. M-quantile model.

results	PROVINCE	FM
1	MASSA CARRARA	0.159
2	LUCCA	0.157
3	PISTOIA	0.157
4	FIRENZE	0.156
5	LIVORNO	0.157
6	PISA	0.156
7	AREZZO	0.157
8	SIENA	0.156
9	GROSSETO	0.158
10	PRATO	0.157

Table 13.6: Estimated Fuzzy Monetary Index (FM) for the Campania Provinces. M-quantile model.

results	PROVINCE	FM
1	CASERTA	0.199
2	BENEVENTO	0.199
3	NAPOLI	0.199
4	AVELLINO	0.197
5	SALERNO	0.197

Figure 13.4: Estimated Fuzzy Monetary Index (FM) for Lombardia Provinces. M-quantile model.

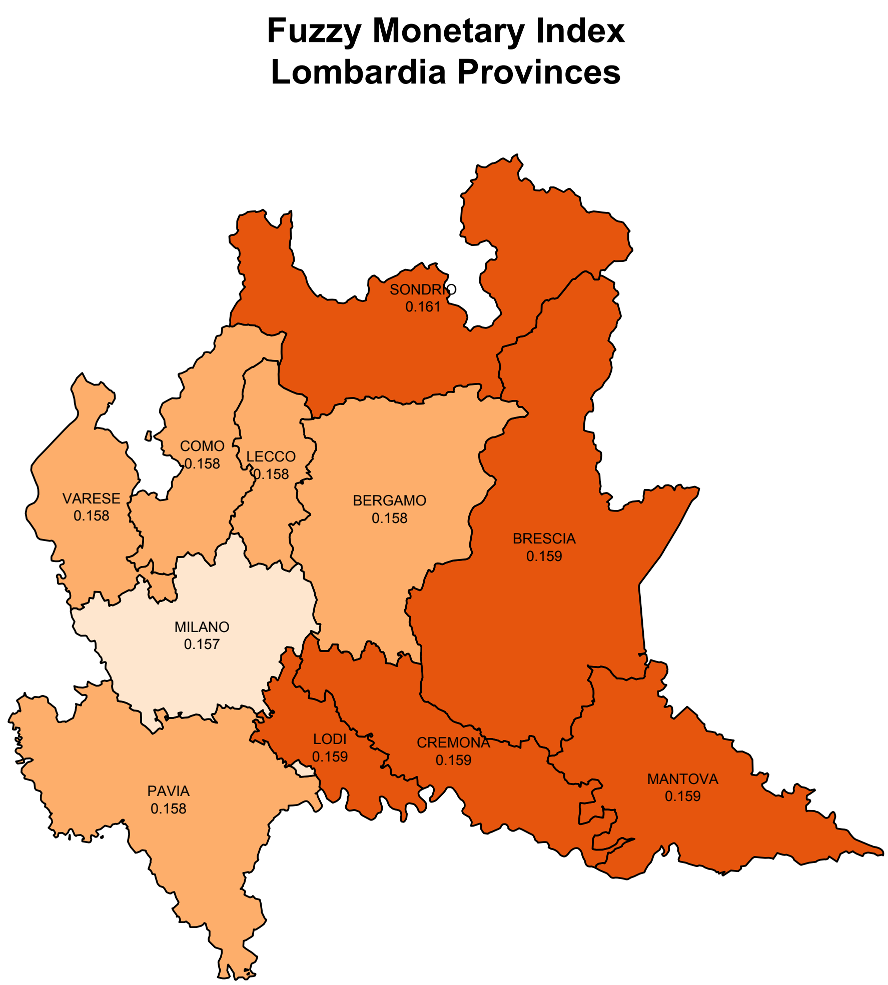


Figure 13.5: Estimated Fuzzy Monetary Index (FM) for Toscana Provinces. M-quantile model.

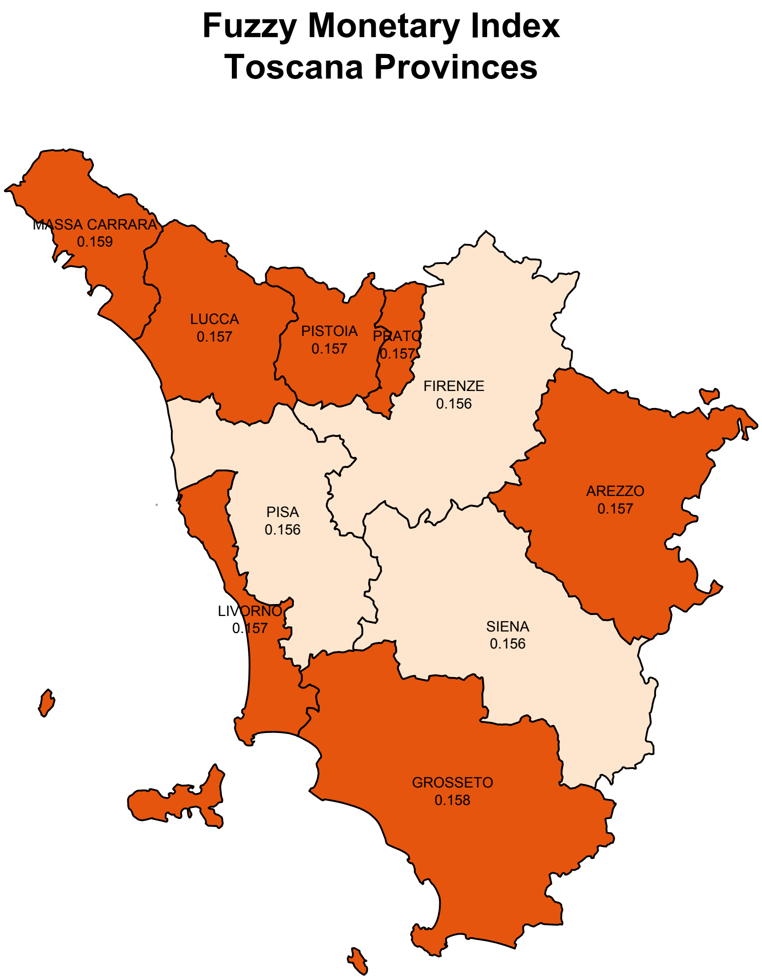
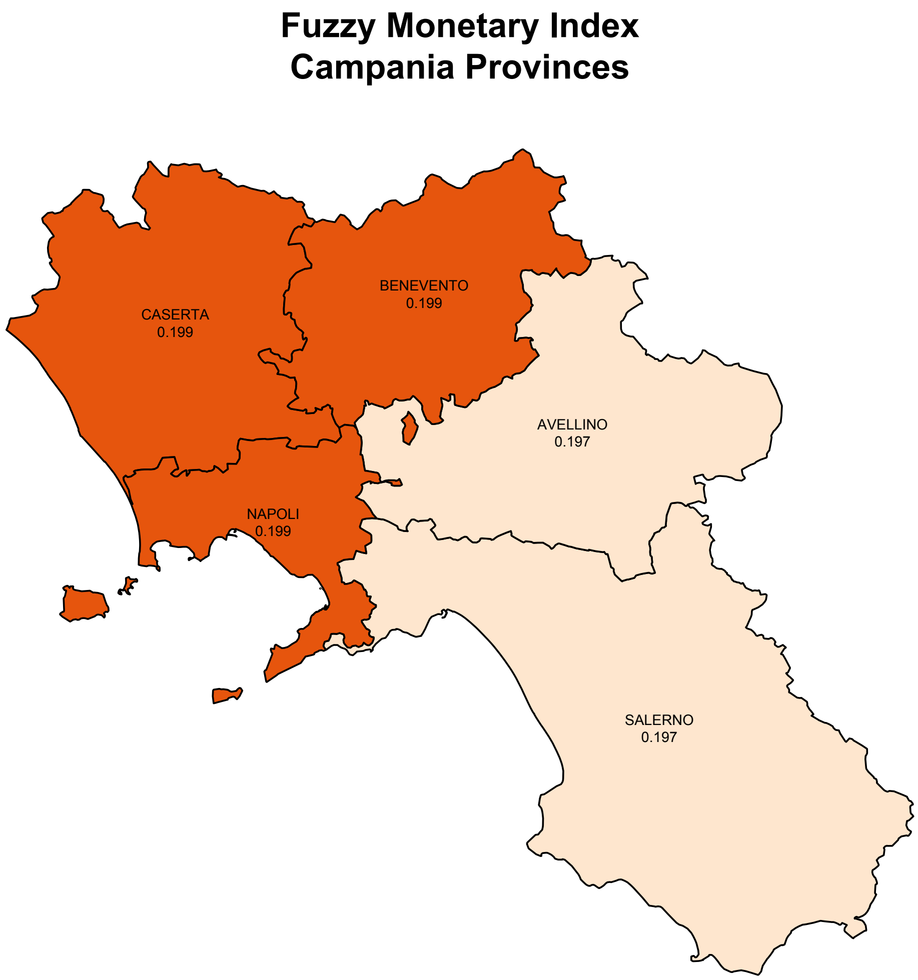


Figure 13.6: Estimated Fuzzy Monetary Index (FM) for Campania Provinces. M-quantile model.



Chapter 14

References

Betti G., Cheli B., Lemmi A., Verma V. (2006). Multidimensional and Longitudinal Poverty: an Integrated Fuzzy Approach, in Lemmi A., Betti G. (eds.) *Fuzzy Set Approach to Multidimensional Poverty Measurement*, 111-137, New York: Springer.

Betti, G. and Verma, V. (1999). Measuring the degree of poverty in a dynamic and comparative context: a multi-dimensional approach using fuzzy set theory. *Proceedings, ICCS-VI, Lahore, Pakistan, August 27-31* **11**, 289-301.

Anselin, L. (1988). *Spatial Econometrics. Methods and Models*. Boston: Kluwer Academic Publishers.

Aragon, Y., Casanova, S., Chambers, R.L., Leconte, E. (2005). Conditional ordering using nonparametric expectiles. *Journal of Official Statistics*, **21**, 617-633.

Ballini, F., Betti, G., Carrette, S. and Neri, L. (2006). Poverty and inequality mapping in the Commonwealth of Dominica. *Estudios Economicos*, **2**, 123-162.

Banerjee, S., Carlin, B. and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. New York: Chapman and Hall.

Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data. *Journal of the American Statistical Association*, **83**, 28-36.

Bell, W. (1997). Models for county and state poverty estimates. Preprint, Statistical Research Division, U. S. Census Bureau.

Betti G., Cheli B., Lemmi A., Verma V. (2006). Multidimensional and Longitudinal Poverty: an Integrated Fuzzy Approach, in Lemmi A., Betti G. (eds.) *Fuzzy Set Approach to Multidimensional Poverty Measurement*, 111-137, New York: Springer.

Betti, G. and Verma, V. (1999). Measuring the degree of poverty in a dynamic and comparative context: a multi-dimensional approach using fuzzy set theory. *Proceedings, ICCS-VI, Lahore, Pakistan, August 27-31* **11**, 289-301.

Betti, G., Ferretti, C., Gagliardi, F., Lemmi, A. and Verma, V. (2009). Proposal for new multidimensional and fuzzy measures of poverty and inequality at national and regional level. *C.R.I.DI.RE. research centre working paper, University of Siena 83*

Bowman, A.W.,Hall, P., Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika*, **85**, 799-808.

Breckling, J. and Chambers, R.L. (1988). M-quantiles. *Biometrika*, **75**, 761-71.

Brunsdon, C., Fotheringham, A.S., Charlton, M. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, **28**, 281-298.

Cardot, H., Ferraty, F. and Sarda, P. (1999). Linear Functional Model. *Statistics and Probability Letters*, **45**, 11–22.

Castro, P.E., Lawton, W.H. and Sylvestre, E.A. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics*, **28**, 329–337.

Chambers, R.L. (1986). Outlier robust finite population estimation. *Journal of the American Statistical Association*, **81**, 1063-1069.

Chambers, R.L. and Dunstan, R. (1986). Estimating distribution functions from survey data. *Biometrika*, **73**, 597-604.

Chambers, R.L., Dorfman, A.H., Hall, P. (1992). Properties of estimators of the finite population distribution function. *Biometrika*, **79**, 577-82.

Chambers, R.L. and Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika*, **93**, 255-268.

Chambers, R.L., Chandra, H., Tzavidis, N. (2008). On robust mean squared error estimation for linear predictors for domains. [Paper submitted for publication. A copy is available upon request].

Chambers, R., Chandra, H. and Tzavidis, N. (2009). On Bias-Robust Mean Squared Error Estimation for Linear Predictors for Domains. *Working Papers, 09-08* Centre for Statistical and Survey Methodology, The University of Wollongong, Australia. In http://cssm.uow.edu.au/publications.

Chatterjee, S., Lahiri, P. and Huilin, L. (2008). Parametric Bootstrap Approximation to the Distribution of EBLUP and Related Prediction Intervals in Linear Mixed Models. *Annals of Statistics*, **36**, 1221-1245.

Cheli, B. and Lemmi, A. (1995). A Totally Fuzzy and Relative Approach to the Multidimensional Analysis of Poverty. *Economic Notes*, 24, 115-134.

Cheli, B. and Betti, G. (1999). Totally Fuzzy and Relative Measures of Poverty in Dynamics Context. *Metron*, **57(1-2)**, 83-104

Ciampalini G., Betti G., Verma V. (2009). Comparability in self-employment income. Working Paper 82/09, Department of Quantitative Methods, University of Siena.

Cole, R.J. (1988). Fitted smoothed centile curves to reference data. *Journal of the Royal Statistical Society, Series A*, **151**, 385-418.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik*, **31**, 377-403.

Cressie, N. (1991). Small-area prediction of undercount using the general linear model. *Proceedings of Statistic Symposium 90: Measurement and Improvement of Data Quality, Ottawa: Statistics Canada*, 93-105.

Cressie, N. (1993). *Statistics for Spatial Data*. New York: John Wiley & Sons.

Cuadras, C.M. and Fortiana, J. (1995). A continuous metric sacling solution for a random variable. *Journal of Multivariate Amalysis*, **52**, 1-14.

Cuadras, C.M. and Fortiana, J. (1998). Visualizing categorical data with related metric scaling. In J. Blasius and M. Greenacre (Eds.), *Visualization of Categorical Data*, pp. 365-376. Academic Press.

Das, K., Jiang, J. and Rao, J. N. K. (2004). Mean squared error of empirical predictor. *The Annals of Statistics*, **32**, 818-840.

Datta, G.S., Lahiri, P., Maiti, T. and Lu, K.L. (1999). Hierarchical Bayes estimation of unemployment rates for the U.S. states. *Journal of the American Statistical Association*, **94**, 1074-1082.

Datta, G.S., Lahiri, P. and Maiti, T. (2002). Empirical Bayes estimation of median income of four-person families by state using time series and cross-sectional data. *Journal of Statistical Planning and Inference*, **102**, 83–97.

Datta, G.S., Rao, J.N.K. and Smith, D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika* **92**, 183-196.

Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a random vector function: some application to statistical inference. *Journal of Multivariate Amalysis*, **12**, 136–154.

Duan, N. (1983). Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association*, **78**, 605-610.

Eilers, P.H. and Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89-121.

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

Eurostat (2006). Treatment of negative income: empirical assessment of the impact of methods used. Report N. ISR I.04, Project EU-SILC (Community statistics on income and living conditions) 2005/S 116-114302 – Lot 1 (Methodological studies to estimate the impact on comparability of the national methods used).

Fay, R. E. and Herriot, R. A. (1979). Estimation of income from small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269–277.

Farber, S. and Páez, A. (2007). A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations. *Journal of Geographical Systems*, **9**, 371-396.

Fellner, W.H. (1986). Robust Estimation of Variance Components. *Technometrics*, **28**, 51-60.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. New York: Springer-Verlag.

Foster, J., Greer, J. and Thorbecke, E. (1984). A class of decomposable poverty measures, *Econometrica*, **52**, 761–766.

Fotheringham, A.S., Brundson, C., and Charlton M. (2002). *Geographically Weighted Regression - The analysis of spatially varying relationship*. West Sussex, England: John Wiley & Sons Ltd.

Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, **9**, 55-93.

Ghosh, M., Nangia, N. and Kim, D. (1996). Estimation of median income of four-person families: a Bayesian time series approach. *Journal of the American Statistical Association*, **91**, 1423–1431.

Giusti, C., Pratesi, M., and Salvati, S. (2009a). Estimation of poverty indicators: a comparison of small area methods at LAU1-2 level in Tuscany. *Abstract Book, NTTS - Conferences on New Techniques and Technologies for Statistics*, Brussels, 18-20 Febbraio 2009.

Giusti, C., Pratesi, M., and Salvati, S. (2009b). Small area methods in the estimation of poverty indicators: the case of Tuscany. *Politica Economica*, **3**, 369-380.

Giusti, C., Marchetti, S., Pratesi, M., and Salvati, S. (in preparation). Semiparametric Fay-Herriot model using Penalized Splines.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D. and Santamaría, L. (2007). Estimation of the mean squared error of predictors of small area linear parameters under logistic mixed model. *Computational Statistics and Data Analyis*, **51**, 2720-2733.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D. and Santamaría, L. (2008a). Bootstrap mean squared error of a small-area EBLUP. *Journal of Statistical Computation and Simulation*, **78**, 443–462.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D. and Santamaría, L. (2008b). Analytic and bootstrap approximations of prediction errors under a multivariate Fay–Herriot model. *Computational Statistics and Data Analysis*, **52**, 5242-5252.

Griffith, D. and Csillag, F. (1993). Exploring relationships between semi-variogram and spatial autoregressive. *Papers in Regional Science*, **72**, 283-296.

Hall, P. and Maiti, T. (2006a). On Parametric Bootstrap Methods for Small Area Prediction. *Journal Royal Statistical Society, Series B*, **68**, 221-238.

Hall, P. and Maiti, T. (2006b). Nonparametric estimation of mean-squared prediction error in nested-error regression models. *The Annals of Statistics*, **34**, 1733-1750.

Hayfield, T. and Racine, J. (2008). Nonparametric Econometrics: The *np* Package. *Journal of statistical Software*, **27**, number 5.

Harville, D. and Jeske, D. (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association*, **87**, 724-731.

Haslett, S. and Jones, G. (2005). Small area estimation using surveys and some practical and statistical issues. *Statistics in Transition*, **7**, 541–555.

He, X. (1997). Quantile curves without crossing. *The American Statistician*, **51**, 186-192.

Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423-447.

Huber, P.J. (1981). *Robust Statistics*. John Wiley & Sons.

Huggins, R.M. (2009). A Robust Approach to the Analysis of Repeated Measures. *Biometrics*, **49**, 255–268.

Jiang, J. (1996). REML estimation: asymptotic behavior and related topics. *Annals of Statistics*, **24**, 255-286.

Jiang, J. and Lahiri, P. (2002). A unified jacknife theory for empirical best prediction with *D*-estimation. *The Annals of Statistics*, **30**, 2720-2733.

Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, **15**, 1-96.

Kackar, R. and Harville, D. (1984). Approximations for standard errors of fixed and random effects in mixed models. *Journal of the American Statistical Association*, **79**, 853-862.

Kakamu, K. (2005). Bayesian estimation of a distance functional weight matrix model. *Economics Bulletin*, **3**, 1-6.

Kaufman, L. and Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33-50.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, **91**, 74-89.

Kokic, P., Chambers, R.L., Breckling, J., Beare, S. (1997). A measure of production performance. *Journal of Business and Economic Statistics*, **15**, 445-451.

LeSage, J.P. and Pace, R.K. (2007). A matrix exponential spatial specification. *Journal of Econometrics*, **140**, 190-214.

Li, Q. and Racine, J.S. (2007). Nonparametric Econometrics: Theory and Practice. *Princeton University Press*.

Lombardia M.J., Gonzalez-Manteiga W. and Prada-Sanchez J.M. (2003). Bootstrapping the Dorfman-Hall-Chambers-Dunstan estimator of a finite population distribution function. *Journal of Nonparametric. Statistics*, **16**, 63-90.

Molina, I., Salvati, N. and Pratesi, M. (2008). Bootstrap for estimating the MSE of the Spatial EBLUP. *Computational Statistics*, **24**, 441-458.

Molina, I. and Rao, J.N.K. (2010). Small Area Estimation of Poverty Indicators. *Canadian Journal of Statistics*, **38**, 369-385.

Molina, I. and Rao, J.N.K. (2009). Small area estimation of poverty indicators. Working Paper 09-15, Statistics and Econometric Series 05, Universidad Carlos III de Madrid, http://halweb.uc3m.es/esp/Personal/personas/imolina/MiDocencia/SmallAreaEstimation/ imjrWorkingPaperPoverty.pdf

Neri, L., Ballini, F. and Betti, G. (2005). Poverty and inequality in transition countries. *Statistics in Transition*, **7**, 135–157.

Neri L., Gagliardi F., Ciampalini G., Verma V., Betti G. (2009). Outliers at upper end of income distribution. Working Paper 86/09, Department of Quantitative Methods, University of Siena.

Newey, W.K. and powell, J.L. (1987). Asymmetric least squared estimation and testing. *Econometrica*, **55**, 819-47.

Nychka, D. and Saltzman, N. (1998). Design of air quality monitoring networks. In Kychka, Douglas Piegorsch, Walter and Cox (eds), *Case studies in environmental statistics*.

Opsomer, J. D., Claeskens, G., Ranalli, M. G., Kauermann, G. and Breidt, F. J. (2008). Nonparametric Small Area Estimation Using Penalized Spline Regression. *Journal of the Royal Statistical Society, Series B*, **70**, 265-283.

Petrucci, A. and Salvati, N. (2006). Small Area Estimation considering Spatial Correlation in Watershed Erosion Assessment Survey. *Journal of Agricultural , Biological, and Environmental Statistics*, Vol. 11, 14,169-182.

Pfeffermann, D. and Buck, L. (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methodology*, **16**, 217—237.

Pfeffermann, D. (2002). Small Area Estimation - New Developments and Directions. *International Statistical Review*, **70**, 1, 125-143.

Pfeffermann, D. and Tiller, R. (2005). Bootstrap approximation to prediction MSE for state-space models with estimated parameters. *Journal of Time Series Analysis*, **26**, 893-916.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, **85**, 163-171.

Pratesi, M., Salvati, N. (2008). Small Area Estimation: the EBLUP estimator based on spatially correlated random area effects. *Statistical Methods and Applications*, **17**, 113-141.

Pratesi, M., Salvati, N. (2009). Small Area Estimation in the presence of correlated random area effects. *Journal of Official Statistics*, **25**, 37-53.

Pratesi, M., Ranalli, M.G., Salvati, N. (2006). Nonparametric M-quantile regression via penalized splines. *ASA Proceedings on Survey Research Methods*, Alexandria, VA.

Pratesi, M., Ranalli, M.G., Salvati, N. (2008). Semiparametric M-quantile regression for estimating the proportion of acidic lakes in 8-digit HUCs of the Northeastern US. *Environmetrics*, **19**, 687-701.

Pratesi, M., Ranalli, M.G., Salvati, N. (2009). Nonparametric M-quantile regression using penalized splines. *Journal of Nonparametric Statistics*, **21**, 287-304.

R Development Core Team (2005). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria. URL: http://www.R-project.org.

Rao, J.N.K., Kovar, J.G., Mantel, H.J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika*, **77**, 365-75.

Rao, J.N.K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, **25**, 175-186.

Rao, J.N.K. and Yu, M. (1994). Small area estimation by combining time series and cross sectional data. *Canadian Journal of Statistics*, **22**, 511—528.

Rao, J.N.K. (2003). *Small Area Estimation*. John Wiley.

Richardson, A.M. and Welsh, A.H.(1995). Robust Estimation in the Mixed Linear Model. *Biometrics*, **51**, 1429-1439.

Royall, R. M. (1976). The linear least squared prediction approach to two-stage sampling. *Journal of the American Statistical Association*, **71**, 657-664.

Royall, R.M. and Cumberland, W.G. (1978). Variance estimation in finite population sampling. *Journal of the American Statistical Association*, **73**, 351 - 58.

Ruppert, D., Wand, M.P., Carroll, R. (2003). *Semiparametric regression*. Cambridge University Press, Cambridge, New York.

Saei, A. and Chambers, R.L. (2005). Empirical best linear unbiased prediction for out of sample areas. In *S3Ri Methodology Working Papers*. Southampton Statistical Science Research Institute, pp. 1-15.

Salvati, N., Tzavidis, N., Pratesi, M., Chambers, R.L. (2008). Small area estimation via M-quantile geographically weighted regression. [Paper submitted for publication. A copy is available upon request].

Salvati, N. and Tzavidis, N. (2010). M-quantile GWR function. Available from *URL: http://www.dipstat.ec.unipi.it/persone/docenti/salvati/.*

Salvati, N., Chandra, H., Ranalli, M.G. and Chambers, R. (2010a). Small Area Estimation Using a Nonparametric Model Based Direct Estimator. *Computational Statistics and Data Analysis*, **54**, 2159-2171.

Salvati, N., Ranalli, M. G. and Pratesi, M. (2010b). Small area estimation of the mean using non-parametric M-quantile regression: a comparison when a linear mixed model does not hold. To appear in *Journal of Statistical Computation and Simulation*.

Särndal, C.E., Swensson, B. and Wretman J. (1992) *Model assisted survey sampling*, Springer-Verlag.

Sinha, S.K. and Rao, J.N.K. (2009). Robust Small Area Estimation. *Canadian Journal of Statistics*, **37**, 381-399.

Singh, B., Shukla, G. and Kundu, D. (2005). Spatio-temporal models in small area estimation. *Survey Methodology*, **31**, 183-195.

Street, J.O., Carroll, R.J. and Ruppert, D. (1988). A note on computing robust regression estimates via iteratively reweighted least squares. *American Statistician*, **42**, 152-154.

Tarozzi, A. and Deaton, A. (forthcoming). Using census and survey data to estimate poverty and inequality for small areas. *Review of Economics and Statistics*.

Tzavidis, N., Marchetti, S., and Chambers, R.L. (2009). Robust prediction of small are means and quantiles. To appear in the Australian and New Zealand Journal of Statistics.

Tzavidis, N. and Chambers, R.L. (2006). Bias adjusted distribution estimation for small areas with outlying values. In *S3Ri Methodology Working Papers*. Southampton Statistical Science Research Institute, pp. 1-30.

Tzavidis, N. and Chambers, R. (2007). Robust prediction of small area means and distributions. *CCSR Working Paper 2007-08, University of Manchester.*

Tzavidis, N., Marchetti, S. and Chambers, R. (2010). Robust estimation of small area means and quantiles. *Australian and New Zealand Journal of Statistics* DOI 10.1111/j.1467-842X.2010.00572.x

Ugarte, M. D., Militino, A.F. and Goicoa, T. (2008). Prediction error estimators in Empirical Bayes disease mapping. *Environmetrics*, **19**, 287-300.

Venables, W.N. and Ripley, B.D. (2002). *Modern Applied Statistics with S*. New York: Springer.

Wang, S. and Dorfman, A.H. (1996). A new estimator of the finite population distribution function. *Biometrika*, **83**, 639-52.

You, Y. and Rao, J.N.K. (2000). Hierarchical Bayes estimation of small area means using multi-level models. *Survey Methodology*, **26**, 173-181.

You, Y., Rao, J.N.K. and Gambino, J. (2001). Model-based unemployment raye estimation for the Canadian Labour Force Survey: a hierarchical approach. Technical report, Household Survey Method Division. Statistics Canada.

Welsh, A.H. and Ronchetti, E.(1998). Bias-calibrated estimation from sample surveys containing outliers. *Journal of the Royal Statistical Society, Series B*, **60**, 413-428.

Zimmerman, D. and Cressie, N. (1992). Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Annals of the Institute of Statistical Mathematics*, **44**, 27-43.

Appendix A

Table of the average equivalised income at municipality level in Toscana; M-quantile GWR model

Table A.1: Estimated Mean of Household Equivalised Income (MEAN) for the Lombardia Provinces. M-quantile GWR model.

	MUNICIPALITY	MEAN
1	Aulla	15145.65
2	Bagnone	14031.35
3	Carrara	15125.06
4	Casola in Lunigiana	14679.20
5	Comano	14272.58
6	Filattiera	14162.57
7	Fivizzano	14820.18
8	Fosdinovo	15399.41
9	Licciana Nardi	12975.37
10	Massa	11353.92
11	Montignoso	15538.55
12	Mulazzo	14095.95
13	Podenzana	14985.54
14	Pontremoli	14478.88
15	Tresana	13961.23
16	Villafranca in Lunigiana	14848.44
17	Zeri	12781.93
18	Altopascio	9519.40
19	Bagni di Lucca	15248.68
20	Barga	16752.06
21	Borgo a Mozzano	16193.51

Continued on next page

Table A.1 – continued from previous page

	MUNICIPALITY	MEAN
22	Camaiore	15978.75
23	Camporgiano	15845.63
24	Capannori	16771.80
25	Careggine	14595.15
26	Castelnuovo di Garfagnana	16414.99
27	Castiglione di Garfagnana	13753.09
28	Coreglia Antelminelli	16072.19
29	Fabbriche di Vallico	14969.97
30	Forte dei Marmi	15929.64
31	Fosciandora	15769.24
32	Gallicano	15699.20
33	Giuncugnano	14820.08
34	Lucca	19464.46
35	Massarosa	16521.51
36	Minucciano	14664.21
37	Molazzana	15405.52
38	Montecarlo	17073.64
39	Pescaglia	15746.42
40	Piazza al Serchio	15333.28
41	Pietrasanta	15885.74
42	Pieve Fosciana	16247.36
43	Porcari	16781.53
44	San Romano in Garfagnana	15758.36
45	Seravezza	15700.23
46	Sillano	14231.79
47	Stazzema	14819.67
48	Vagli Sotto	14875.01
49	Vergemoli	14292.60
50	Viareggio	13505.10
51	Villa Basilica	15613.64
52	Villa Collemandina	15328.68
53	Abetone	14707.64
54	Agliana	27899.02
55	Buggiano	21403.63
56	Cutigliano	15340.94
57	Lamporecchio	16798.64
58	Larciano	16729.67
59	Marliana	16170.03
60	Massa e Cozzile	17266.73
61	Monsummano Terme	16779.64
62	Montale	17134.67

Continued on next page

Table A.1 – continued from previous page

	MUNICIPALITY	MEAN
63	Montecatini-Terme	17207.59
64	Pescia	20515.86
65	Pieve a Nievole	16128.43
66	Pistoia	19242.42
67	Piteglio	15384.71
68	Ponte Buggianese	16407.61
69	Quarrata	17954.16
70	Sambuca Pistoiese	15034.19
71	San Marcello Pistoiese	16272.40
72	Serravalle Pistoiese	17418.05
73	Uzzano	17525.30
74	Chiesina Uzzanese	16522.53
75	Bagno a Ripoli	20280.53
76	Barberino di Mugello	17302.28
77	Barberino Val d'Elsa	17574.61
78	Borgo San Lorenzo	17570.39
79	Calenzano	17804.86
80	Campi Bisenzio	17240.00
81	Capraia e Limite	17789.13
82	Castelfiorentino	15755.97
83	Cerreto Guidi	17054.37
84	Certaldo	16826.96
85	Dicomano	16825.65
86	Empoli	16455.70
87	Fiesole	18451.22
88	Figline Valdarno	16568.62
89	Firenze	21962.34
90	Firenzuola	16007.51
91	Fucecchio	17006.12
92	Gambassi Terme	17153.25
93	Greve in Chianti	17684.39
94	Impruneta	17990.57
95	Incisa in Val d'Arno	15113.42
96	Lastra a Signa	23040.28
97	Londa	17387.29
98	Marradi	16164.74
99	Montaione	17058.65
100	Montelupo Fiorentino	17417.53
101	Montespertoli	29006.54
102	Palazzuolo sul Senio	16333.72
103	Pelago	17688.75

Continued on next page

Table A.1 – continued from previous page

	MUNICIPALITY	MEAN
104	Pontassieve	16924.25
105	Reggello	16848.52
106	Rignano sull'Arno	18083.70
107	Rufina	17399.75
108	San Casciano in Val di Pesa	14751.05
109	San Godenzo	14488.51
110	San Piero a Sieve	17052.45
111	Scandicci	22357.68
112	Scarperia	17636.06
113	Sesto Fiorentino	28968.88
114	Signa	17449.50
115	Tavarnelle Val di Pesa	17539.76
116	Vaglia	18610.30
117	Vicchio	17316.58
118	Vinci	17179.81
119	Bibbona	16304.09
120	Campiglia Marittima	15443.13
121	Campo nell'Elba	14941.33
122	Capoliveri	15526.59
123	Capraia Isola	16349.86
124	Castagneto Carducci	16007.37
125	Cecina	19444.17
126	Collesalvetti	17258.28
127	Livorno	18191.81
128	Marciana	14945.63
129	Marciana Marina	15121.17
130	Piombino	16782.43
131	Porto Azzurro	15394.01
132	Portoferraio	15775.21
133	Rio Marina	14775.01
134	Rio nell'Elba	14198.18
135	Rosignano Marittimo	16793.21
136	San Vincenzo	16335.38
137	Sassetta	15444.70
138	Suvereto	16354.54
139	Bientina	16989.69
140	Buti	16774.13
141	Calci	17798.60
142	Calcinaia	19055.17
143	Capannoli	17030.47
144	Casale Marittimo	16964.39

Continued on next page

Table A.1 – continued from previous page

	MUNICIPALITY	MEAN
145	Casciana Terme	17103.13
146	Cascina	17109.20
147	Castelfranco di Sotto	16647.55
148	Castellina Marittima	16532.63
149	Castelnuovo di Val di Cecina	15930.44
150	Chianni	16320.38
151	Crespina	17403.14
152	Fauglia	17311.94
153	Guardistallo	16663.99
154	Lajatico	16326.23
155	Lari	18579.13
156	Lorenzana	16916.92
157	Montecatini Val di Cecina	16155.77
158	Montescudaio	17012.33
159	Monteverdi Marittimo	15694.30
160	Montopoli in Val d'Arno	17132.64
161	Orciano Pisano	16301.59
162	Palaia	16850.01
163	Peccioli	16373.61
164	Pisa	18830.44
165	Pomarance	14400.80
166	Ponsacco	17039.98
167	Pontedera	17506.71
168	Riparbella	16307.70
169	San Giuliano Terme	21903.06
170	San Miniato	17308.32
171	Santa Croce sull'Arno	26489.68
172	Santa Luce	16383.71
173	Santa Maria a Monte	16793.69
174	Terricciola	16858.09
175	Vecchiano	17110.31
176	Vicopisano	17031.02
177	Volterra	17407.30
178	Anghiari	16832.29
179	Arezzo	20980.25
180	Badia Tedalda	16211.21
181	Bibbiena	19065.26
182	Bucine	17136.62
183	Capolona	17353.08
184	Caprese Michelangelo	16065.20
185	Castel Focognano	16676.27

Continued on next page

Table A.1 – continued from previous page

	MUNICIPALITY	MEAN
186	Castelfranco di Sopra	17534.71
187	Castel San Niccolo’	15874.99
188	Castiglion Fibocchi	17585.34
189	Castiglion Fiorentino	15409.50
190	Cavriglia	17327.47
191	Chitignano	16435.45
192	Chiusi della Verna	16523.63
193	Civitella in Val di Chiana	17316.20
194	Cortona	16808.24
195	Foiano della Chiana	16867.92
196	Laterina	17204.16
197	Loro Ciuffenna	17266.88
198	Lucignano	16527.92
199	Marciano della Chiana	17190.52
200	Montemignaio	15032.40
201	Monterchi	16234.93
202	Monte San Savino	16694.54
203	Montevarchi	17606.51
204	Ortignano Raggiolo	16088.81
205	Pergine Valdarno	17508.94
206	Pian di Sco	17811.19
207	Pieve Santo Stefano	16667.73
208	Poppi	12999.41
209	Pratovecchio	17235.63
210	San Giovanni Valdarno	17572.99
211	Sansepolcro	19289.23
212	Sestino	15628.34
213	Stia	17240.87
214	Subbiano	17402.17
215	Talla	15840.95
216	Terranuova Bracciolini	17490.91
217	Abbadia San Salvatore	16122.26
218	Asciano	15755.71
219	Buonconvento	16733.72
220	Casole d’Elsa	17029.39
221	Castellina in Chianti	17475.37
222	Castelnuovo Berardenga	18632.71
223	Castiglione d’Orcia	14823.93
224	Cetona	15539.19
225	Chianciano Terme	17471.31
226	Chiusdino	16073.11

Continued on next page

Table A.1 – continued from previous page

	MUNICIPALITY	MEAN
227	Chiusi	16692.65
228	Colle di Val d'Elsa	17537.78
229	Gaiole in Chianti	16524.14
230	Montalcino	16525.24
231	Montepulciano	21652.20
232	Monteriggioni	18512.45
233	Monteroni d'Arbia	17713.99
234	Monticiano	16090.07
235	Murlo	16634.15
236	Piancastagnaio	15562.69
237	Pienza	16562.53
238	Poggibonsi	17474.05
239	Radda in Chianti	17006.58
240	Radicofani	14517.07
241	Radicondoli	16324.59
242	Rapolano Terme	20100.34
243	San Casciano dei Bagni	14553.52
244	San Gimignano	17228.01
245	San Giovanni d'Asso	15948.62
246	San Quirico d'Orcia	16377.37
247	Sarteano	15976.99
248	Siena	20752.49
249	Sinalunga	11603.94
250	Sovicille	24062.68
251	Torrita di Siena	17946.87
252	Trequanda	15919.05
253	Arcidosso	15392.57
254	Campagnatico	15846.57
255	Capalbio	14388.73
256	Castel del Piano	16036.07
257	Castell'Azzara	15164.94
258	Castiglione della Pescaia	14924.13
259	Cinigiano	15481.92
260	Civitella Paganico	15704.62
261	Follonica	16075.56
262	Gavorrano	15935.53
263	Grosseto	16892.58
264	Isola del Giglio	12481.48
265	Magliano in Toscana	15796.67
266	Manciano	15581.49
267	Massa Marittima	16313.84

Continued on next page

Table A.1 – continued from previous page

	MUNICIPALITY	MEAN
268	Monte Argentario	14258.23
269	Montieri	15227.14
270	Orbetello	15123.06
271	Pitigliano	15351.17
272	Roccalbegna	14983.44
273	Roccastrada	15777.72
274	Santa Fiora	17083.69
275	Scansano	15131.31
276	Scarlino	16305.15
277	Seggiano	14613.24
278	Sorano	14302.97
279	Monterotondo Marittimo	15705.33
280	Semproniano	15413.16
281	Cantagallo	16067.57
282	Carmignano	17970.49
283	Montemurlo	16606.89
284	Poggio a Caiano	17655.20
285	Prato	17709.61
286	Vaiano	19620.34
287	Vernio	16169.19

Appendix B

Table of the average equivalised income at municipality level in Lombardia, Toscana and Campania; semiparametric Fay-Herriot model

Table B.1: Estimated Mean of Household Equivalised Income (x1000) (MEAN) and estimated Root Mean Squared Error of the Mean estimator (RMSE) for the Lombardia Municipalities. Semi-parametric Fay and Herriot model.

	MUNICIPALITY	MEAN	RMSE
1	Como	22.17	1.92
2	Consiglio di Rumo	19.49	1.45
3	Corrido	19.87	1.18
4	Cremia	19.89	1.59
5	Cucciago	19.30	0.96
6	Cusino	18.20	2.42
7	Dizzasco	19.90	2.91
8	Domaso	21.66	1.66
9	Dongo	19.59	1.29
10	Dosso del Liro	19.88	2.30
11	Drezzo	20.42	0.94
12	Erba	18.95	1.87
13	Eupilio	19.41	0.98
14	Faggeto Lario	20.91	1.37
15	Faloppio	20.12	0.99
16	Fenger	20.57	1.32

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
17	Figino Serenza	17.86	0.88
18	Fino Mornasco	20.02	0.75
19	Garzeno	19.00	2.02
20	Gera Lario	21.54	2.00
21	Germasino	21.99	2.27
22	Gironico	20.37	1.32
23	Grandate	20.51	1.07
24	Grandola ed Uniti	20.11	1.20
25	Gravedona	21.57	2.68
26	Griante	21.11	1.79
27	Guanzate	19.73	0.98
28	Inverigo	18.63	0.77
29	Laglio	21.08	1.10
30	Laino	22.00	1.71
31	Lambrugo	20.02	0.85
32	Lanzo d'Intelvi	22.53	2.08
33	Lasnigo	19.91	1.65
34	Lenno	20.38	1.31
35	Vimercate	19.14	0.77
36	Vimodrone	19.28	0.71
37	Vittuone	20.44	1.81
38	Vizzolo Predabissi	17.06	0.98
39	Zelo Surrigone	18.43	1.71
40	Zibido San Giacomo	19.11	1.62
41	Villa Cortese	19.02	0.90
42	Vanzaghello	19.66	0.98
43	Baranzate		
44	Garlasco	20.15	1.23
45	Genzone	19.03	1.83
46	Gerenzago	17.15	1.01
47	Giussago	18.17	1.02
48	Godiasco	20.75	2.42
49	Golferenzo	20.02	3.29
50	Gravellona Lomellina	19.87	1.63
51	Gropello Cairoli	20.52	1.52
52	Inverno e Monteleone	17.29	1.51
53	Landriano	19.26	0.88
54	Langosco	23.06	2.60
55	Lardirago	19.70	1.58
56	Linarolo	18.54	1.04
57	Lirio	19.22	3.43

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
58	Lomello	20.02	1.63
59	Lungavilla	17.96	1.73
60	Magherno	18.89	0.85
61	Marcignago	18.54	1.84
62	Marzano	18.20	0.92
63	Mede	20.99	1.77
64	Menconico	19.73	3.90
65	Mezzana Bigli	21.30	2.19
66	Mezzana Rabattone	19.46	1.63
67	Mezzanino	18.43	1.44
68	Miradolo Terme	19.97	1.29
69	Montalto Pavese	18.46	2.58
70	Montebello della Battaglia	18.54	1.75
71	Montecalvo Versiggia	18.16	2.47
72	Montescano	17.86	2.11
73	Montesegale	15.99	4.22
74	Monticelli Pavese	16.94	1.63
75	Beccaria	20.11	2.04
76	Mornico Losana	19.98	2.27
77	Mortara	22.51	1.64
78	Nicorvo	19.78	1.88
79	Olevano di Lomellina	20.69	1.78
80	Oliva Gessi	20.01	2.45
81	Ottobiano	18.55	1.84
82	Palestro	20.06	2.07
83	Pancarana	18.80	2.52
84	Parona	21.02	1.26
85	Pavia	21.06	2.49
86	Lezzeno	20.74	2.09
87	Limido Comasco	21.25	1.42
88	Lipomo	20.18	0.98
89	Livo	16.04	4.14
90	Locate Varesino	20.43	0.89
91	Lomazzo	20.50	0.98
92	Longone al Segrino	18.90	1.17
93	Luisago	19.64	0.78
94	Lurago d’Erba	19.63	0.73
95	Lurago Marinone	20.14	1.03
96	Lurate Caccivio	20.10	0.81
97	Magreglio	22.05	2.52
98	Mariano Comense	18.69	0.75

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
99	Maslianico	19.34	0.95
100	San Colombano al Lambro	17.56	1.11
101	San Donato Milanese	19.28	0.88
102	San Giorgio su Legnano	19.01	1.14
103	San Giuliano Milanese	18.37	0.80
104	Santo Stefano Ticino	18.97	0.83
105	San Vittore Olona	18.76	0.83
106	San Zenone al Lambro	17.07	1.51
107	Sedriano	19.51	0.88
108	Segrate	20.91	2.59
109	Senago	17.81	0.88
110	Seregno	18.57	0.71
111	Sesto San Giovanni	20.21	1.73
112	Settala	18.18	1.12
113	Settimo Milanese	19.30	1.01
114	Seveso	18.37	0.79
115	Solaro	18.90	0.90
116	Sovico	18.86	1.15
117	Sulbiate	18.47	0.80
118	Trezzano Rosa	18.72	1.61
119	Trezzano sul Naviglio	18.25	0.80
120	Trezzo sull'Adda	19.39	0.82
121	Tribiano	18.28	2.19
122	Triuggio	18.63	0.76
123	Truccazzano	17.26	0.97
124	Turbigo	20.33	0.97
125	Usmate Velate	17.89	0.98
126	Vanzago	19.25	0.94
127	Vaprio d'Adda	19.21	0.84
128	Varedo	17.91	0.77
129	Vedano al Lambro	19.48	0.59
130	Veduggio con Colzano	18.57	0.98
131	Verano Brianza	18.06	0.76
132	Vermezzo	18.84	1.32
133	Vernate	19.26	0.96
134	Vignate	18.60	0.98
135	Villasanta	18.54	0.72
136	Pietra de' Giorgi	19.06	2.10
137	Pieve Albignola	19.48	1.58
138	Pieve del Cairo	20.34	1.79
139	Pieve Porto Morone	18.42	1.59

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
140	Agra	20.39	2.94
141	Albizzate	20.13	1.01
142	Angera	21.44	1.39
143	Arcisate	19.02	1.05
144	Arsago Seprio	20.58	1.29
145	Azzate	21.68	1.09
146	Azzio	20.17	1.59
147	Barasso	21.38	1.11
148	Bardello	21.45	1.36
149	Bedero Valcuvia	19.13	1.81
150	Besano	19.11	1.12
151	Besnate	19.66	1.11
152	Besozzo	21.40	1.29
153	Biandronno	20.71	1.14
154	Bisuschio	19.16	1.03
155	Bodio Lomnago	19.73	1.25
156	Brebbia	21.43	1.37
157	Bregano	20.74	1.60
158	Brenta	19.34	1.47
159	Brezzo di Bedero	22.90	3.56
160	Brinzio	21.67	1.48
161	Brissago-Valtravaglia	20.37	2.03
162	Brunello	20.40	1.68
163	Brusimpiano	22.46	2.32
164	Buguggiate	19.79	1.08
165	Busto Arsizio	19.93	2.90
166	Cadegliano-Viconago	20.78	1.82
167	Cadrezzate	20.31	1.33
168	Cairate	19.22	0.95
169	Cantello	19.15	0.93
170	Caravate	20.09	1.35
171	Cardano al Campo	19.80	1.00
172	Carnago	19.71	0.84
173	Caronno Pertusella	19.44	0.67
174	Caronno Varesino	19.50	0.97
175	Casale Litta	19.56	1.24
176	Casalzuigno	21.69	1.47
177	Casciago	19.11	1.33
178	Casorate Sempione	20.33	1.05
179	Cassano Magnago	21.04	2.66
180	Cassano Valcuvia	20.37	1.43

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
181	Castellanza	19.94	0.91
182	Castello Cabiaglio	23.18	2.19
183	Castelseprio	19.75	1.14
184	Castelveccana	22.67	1.72
185	Castiglione Olona	19.61	0.86
186	Castronno	18.00	1.16
187	Cavaria con Premezzo	19.15	1.06
188	Cazzago Brabbia	22.58	1.97
189	Cislago	19.35	0.66
190	Cittiglio	20.68	1.43
191	Clivio	19.40	0.92
192	Cocquio-Trevisago	21.10	1.24
193	Comabbio	20.19	1.32
194	Comerio	21.32	1.13
195	Cremenaga	18.58	1.61
196	Crosio della Valle	21.46	1.58
197	Cuasso al Monte	21.77	1.64
198	Cugliate-Fabiasco	18.54	1.60
199	Cunardo	19.90	1.60
200	Curiglia con Monteviasco	19.25	3.29
201	Cuveglio	19.71	1.56
202	Cuvio	22.39	1.65
203	Daverio	19.91	1.10
204	Dumenza	22.82	1.94
205	Duno	23.64	3.62
206	Fagnano Olona	19.00	1.17
207	Ferno	19.63	1.02
208	Ferrera di Varese	21.47	1.43
209	Gallarate	21.69	1.10
210	Galliate Lombardo	20.14	1.30
211	Gavirate	21.73	2.39
212	Gazzada Schianno	20.29	1.04
213	Gemonio	21.66	1.30
214	Gerenzano	19.44	0.77
215	Germignaga	21.23	1.55
216	Golasecca	22.02	1.26
217	Gorla Maggiore	19.96	1.03
218	Gorla Minore	19.87	0.80
219	Gornate-Olona	19.80	0.92
220	Grantola	20.79	1.66
221	Inarzo	19.12	1.45

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
222	Induno Olona	20.26	0.97
223	Ispra	22.78	1.64
224	Jerago con Orago	20.01	0.94
225	Lavena Ponte Tresa	20.68	1.59
226	Laveno-Mombello	22.49	1.64
227	Leggiuno	21.81	1.70
228	Lonate Ceppino	20.05	1.04
229	Lonate Pozzolo	19.63	1.10
230	Lozza	20.40	1.35
231	Luino	18.84	2.08
232	Luvinate	20.71	1.08
233	Maccagno	22.30	1.85
234	Malgesso	19.68	1.44
235	Malnate	19.87	0.89
236	Marchirolo	19.69	1.84
237	Marnate	18.88	0.87
238	Marzio	21.14	2.51
239	Masciago Primo	20.50	2.26
240	Mercallo	20.01	2.41
241	Mesenzana	20.19	1.40
242	Montegrino Valtravaglia	20.93	1.83
243	Monvalle	23.48	2.05
244	Morazzone	19.88	0.90
245	Mornago	19.86	1.18
246	Oggiona con Santo Stefano	19.66	1.07
247	Olgiate Olona	19.55	0.82
248	Origgio	20.00	0.92
249	Orino	22.34	2.08
250	Osmate	22.72	1.45
251	Pino sulla Sponda del Lago Maggiore	20.59	3.29
252	Porto Ceresio	20.18	1.50
253	Porto Valtravaglia	22.69	2.02
254	Rancio Valcuvia	20.25	1.62
255	Ranco	23.58	1.89
256	Saltrio	19.14	1.00
257	Samarate	17.79	1.20
258	Saronno	21.81	1.43
259	Sesto Calende	21.13	1.31
260	Solbiate Arno	20.83	1.10
261	Solbiate Olona	19.65	0.97
262	Somma Lombardo	21.35	1.10

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
263	Sumirago	21.08	2.70
264	Taino	22.05	1.33
265	Ternate	20.64	1.23
266	Tradate	20.01	0.85
267	Travedona-Monate	21.06	1.20
268	Tronzano Lago Maggiore	23.89	2.39
269	Uboldo	19.31	0.72
270	Valganna	21.76	1.45
271	Varano Borghi	22.04	1.52
272	Varese	21.72	2.20
273	Vedano Olona	19.77	0.85
274	Veddasca	27.96	3.49
275	Venegono Inferiore	20.09	2.39
276	Venegono Superiore	19.27	0.81
277	Vergiate	21.16	1.06
278	Viggi	20.16	0.91
279	Vizzola Ticino	20.04	1.17
280	Sangiano	20.10	1.58
281	Albavilla	20.26	0.82
282	Albese con Cassano	18.17	1.26
283	Albiolo	19.11	0.84
284	Alserio	19.70	1.53
285	Alzate Brianza	18.61	0.89
286	Anzano del Parco	19.58	0.84
287	Appiano Gentile	20.14	0.93
288	Argegno	19.77	1.75
289	Arosio	18.83	0.80
290	Asso	20.25	0.89
291	Barni	20.69	3.16
292	Bellagio	21.67	1.50
293	Bene Lario	17.67	1.76
294	Beregazzo con Figliaro	20.59	1.16
295	Binago	20.57	1.05
296	Bizzarone	18.78	1.03
297	Blessagno	21.67	1.58
298	Blevio	20.99	1.49
299	Bregnano	19.22	0.77
300	Brenna	19.05	1.18
301	Brienno	22.53	1.99
302	Brunate	21.16	1.41
303	Bulgarograsso	19.40	0.84

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
304	Cabiate	18.45	0.83
305	Cadorago	19.97	0.94
306	Caglio	21.07	2.80
307	Cagno	19.71	0.97
308	Campione d'Italia	21.66	2.13
309	Cant	18.30	1.54
310	Canzo	19.96	1.39
311	Capiago Intimiano	19.15	0.75
312	Carate Urio	20.47	1.86
313	Carbonate	18.11	0.97
314	Carimate	18.00	0.83
315	Carlazzo	20.32	1.30
316	Carugo	18.18	0.78
317	Casasco d'Intelvi	22.61	2.60
318	Caslino d'Erba	20.19	1.17
319	Casnate con Bernate	18.48	0.84
320	Cassina Rizzardi	19.82	1.21
321	Castelmarte	17.70	1.35
322	Castelnuovo Bozzente	21.77	1.73
323	Castiglione d'Intelvi	20.69	2.43
324	Cavallasca	19.37	0.95
325	Cavargna	15.64	3.79
326	Cerano d'Intelvi	23.18	2.48
327	Cermenate	19.72	0.83
328	Cernobbio	19.92	1.07
329	Cirimido	19.96	0.83
330	Civenna	20.27	2.53
331	Claino con Osteno	17.88	2.81
332	Colonno	22.85	1.96
333	Menaggio	22.90	1.85
334	Merone	18.96	1.10
335	Mezzegra	20.61	1.16
336	Moltrasio	20.09	1.36
337	Monguzzo	17.99	0.96
338	Montano Lucino	19.31	0.75
339	Montemezzo	18.82	2.45
340	Montorfano	18.56	0.82
341	Mozzate	19.79	0.69
342	Musso	19.65	1.42
343	Nesso	19.19	1.62
344	Novedrate	18.20	1.03

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
345	Olgiate Comasco	20.31	0.88
346	Oltrona di San Mamette	20.82	1.26
347	Orsenigo	17.97	0.98
348	Ossuccio	19.49	1.57
349	Par	19.80	0.91
350	Peglio	21.55	2.27
351	Pellio Intelvi	23.80	2.41
352	Pianello del Lario	20.57	2.06
353	Pigra	21.13	2.13
354	Plesio	22.09	2.39
355	Pognana Lario	20.97	1.68
356	Ponna	22.98	2.67
357	Ponte Lambro	19.85	0.73
358	Porlezza	18.99	2.15
359	Proserpio	18.58	1.08
360	Pusiano	19.52	0.67
361	Ramponio Verna	21.46	2.64
362	Rezzago	20.15	2.83
363	Rodero	21.97	1.69
364	Ronago	21.64	1.91
365	Rovellasca	20.05	0.72
366	Rovello Porro	18.79	0.79
367	Sala Comacina	19.06	2.11
368	San Bartolomeo Val Cavargna	19.78	1.47
369	San Fedele Intelvi	22.83	2.42
370	San Fermo della Battaglia	18.95	0.88
371	San Nazzaro Val Cavargna	21.51	1.96
372	Schignano	20.59	2.09
373	Senna Comasco	18.46	1.45
374	Solbiate	19.87	0.98
375	Sorico	20.12	1.65
376	Sormano	22.78	2.60
377	Stazzona	19.57	1.51
378	Tavernerio	19.18	1.76
379	Torno	21.39	1.46
380	Tremezzo	21.65	1.51
381	Trezzone	21.05	2.66
382	Turate	20.05	0.88
383	Uggiate-Trevano	20.03	0.94
384	Valbrona	21.25	1.56
385	Valmorea	20.17	2.54

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
386	Val Rezzo	13.53	5.78
387	Valsolda	20.39	1.92
388	Veleso	22.63	2.69
389	Veniano	19.25	1.02
390	Vercana	20.10	2.28
391	Vertemate con Minoprio	19.47	0.85
392	Villa Guardia	20.15	0.83
393	Zelbio	22.69	3.04
394	San Siro		
395	Albaredo per San Marco	15.40	2.45
396	Albosaggia	18.38	1.44
397	Andalo Valtellino	19.97	1.54
398	Aprica	20.40	3.04
399	Ardenno	17.90	1.32
400	Bema	22.19	2.99
401	Berbenno di Valtellina	17.81	1.52
402	Bianzone	19.42	1.56
403	Bormio	20.34	3.42
404	Buglio in Monte	16.34	1.91
405	Caiolo	17.20	1.93
406	Campodolcino	22.08	3.54
407	Caspoggio	18.86	3.13
408	Castello dell'Acqua	16.43	2.24
409	Castione Andevenno	18.63	1.17
410	Cedrasco	19.09	1.51
411	Cercino	19.45	1.37
412	Chiavenna	19.28	1.84
413	Chiesa in Valmalenco	19.73	2.49
414	Chiuro	18.19	1.48
415	Cino	16.84	3.07
416	Civo	17.64	2.66
417	Colorina	15.09	2.18
418	Cosio Valtellino	18.24	1.14
419	Dazio	20.41	2.25
420	Delebio	19.59	1.09
421	Dubino	17.66	1.79
422	Faedo Valtellino	20.50	1.73
423	Forcola	18.40	1.57
424	Fusine	20.81	1.53
425	Gerola Alta	22.86	3.54
426	Gordona	18.59	1.76

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
427	Grosio	16.90	2.38
428	Grosotto	19.56	1.99
429	Madesimo	23.04	4.01
430	Lanzada	18.37	2.76
431	Livigno	20.67	5.33
432	Lovero	19.72	1.77
433	Mantello	17.07	1.61
434	Mazzo di Valtellina	19.52	2.38
435	Mello	17.41	2.36
436	Menarola	17.35	4.23
437	Mese	17.66	2.00
438	Montagna in Valtellina	18.67	1.34
439	Morbegno	16.62	1.41
440	Novate Mezzola	18.39	1.89
441	Pedesina	20.97	3.82
442	Piantedo	18.86	1.71
443	Piateda	17.12	1.65
444	Piuro	18.45	2.37
445	Poggiridenti	18.26	1.24
446	Ponte in Valtellina	19.72	1.47
447	Postalesio	18.39	2.08
448	Prata Camportaccio	18.97	1.96
449	Rasura	18.19	2.32
450	Rogolo	20.28	1.52
451	Samolaco	18.63	2.32
452	San Giacomo Filippo	21.00	3.02
453	Sernio	19.79	2.08
454	Sondalo	20.99	2.51
455	Sondrio	20.11	2.07
456	Spriana	26.14	3.16
457	Talamona	17.51	1.23
458	Tartano	18.68	2.90
459	Teglio	18.54	1.72
460	Tirano	19.03	1.50
461	Torre di Santa Maria	17.88	2.10
462	Tovo di Sant'Agata	18.45	2.13
463	Traona	17.15	1.45
464	Tresivio	19.64	1.39
465	Valdidentro	20.05	3.71
466	Valdisotto	18.53	3.17
467	Valfurva	17.30	3.53

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
468	Val Masino	20.72	2.25
469	Verceia	17.31	2.01
470	Vervio	18.91	2.17
471	Villa di Chiavenna	19.04	2.12
472	Villa di Tirano	19.12	1.58
473	Abbiategrasso	20.54	2.32
474	Agrate Brianza	18.67	0.69
475	Aicurzio	18.13	0.81
476	Albairate	18.97	1.01
477	Albiate	21.02	1.25
478	Arconate	20.41	1.16
479	Arcore	18.37	0.77
480	Arese	18.08	0.79
481	Arluno	20.00	0.82
482	Assago	23.08	2.17
483	Bareggio	19.15	0.79
484	Barlassina	18.85	0.82
485	Basiano	18.45	1.07
486	Basiglio	23.59	4.01
487	Bellinzago Lombardo	18.41	0.94
488	Bellusco	18.37	0.76
489	Bernareggio	18.30	0.97
490	Bernate Ticino	19.23	1.03
491	Besana in Brianza	18.57	0.64
492	Besate	19.54	1.12
493	Biassono	17.77	0.83
494	Binasco	19.51	0.85
495	Boffalora sopra Ticino	19.32	0.91
496	Bollate	19.33	0.79
497	Bovisio-Masciago	18.54	0.65
498	Bresso	19.72	0.73
499	Briosco	17.72	0.79
500	Brugherio	20.21	2.06
501	Bubbiano	18.08	2.06
502	Buccinasco	18.12	1.26
503	Burago di Molgora	17.57	0.73
504	Buscate	19.76	0.85
505	Busnago	17.87	1.06
506	Bussero	18.36	1.03
507	Busto Garolfo	21.12	1.75
508	Calvignasco	19.16	1.73

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
509	Cambiago	18.77	1.11
510	Camparada	18.02	0.96
511	Canegrate	18.81	0.85
512	Caponago	19.27	1.42
513	Carate Brianza	19.17	0.78
514	Carnate	17.56	1.01
515	Carpiano	18.06	1.09
516	Carugate	17.98	0.91
517	Casarile	19.55	2.15
518	Casorezzo	20.15	1.14
519	Cassano d'Adda	18.71	0.67
520	Cassina de' Pecchi	17.96	0.67
521	Cassinetta di Lugagnano	20.09	1.06
522	Castano Primo	21.26	2.29
523	Cavenago di Brianza	18.15	0.96
524	Ceriano Laghetto	18.43	0.82
525	Cernusco sul Naviglio	18.20	0.69
526	Cerro al Lambro	16.27	1.26
527	Cerro Maggiore	19.79	0.80
528	Cesano Boscone	19.89	1.28
529	Cesano Maderno	17.51	0.91
530	Cesate	19.58	1.06
531	Cinisello Balsamo	17.89	1.30
532	Cisliano	17.93	0.95
533	Cogliate	18.08	0.79
534	Cologno Monzese	17.93	1.31
535	Colturano	18.18	2.11
536	Concorezzo	19.23	1.46
537	Corbetta	19.47	0.75
538	Cormano	18.95	1.33
539	Cornaredo	18.80	0.97
540	Cornate d'Adda	18.05	0.63
541	Correzzana	17.68	0.85
542	Corsico	19.26	0.82
543	Cuggiono	19.84	1.06
544	Cusago	19.46	1.11
545	Cusano Milanino	19.61	0.76
546	Dairago	19.43	0.95
547	Desio	18.46	0.56
548	Dresano	17.32	0.94
549	Gaggiano	18.62	0.76

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
550	Garbagnate Milanese	19.15	1.03
551	Gessate	19.95	2.12
552	Giussano	18.31	0.71
553	Gorgonzola	18.62	0.86
554	Grezzago	17.16	1.09
555	Gudo Visconti	18.32	1.29
556	Inveruno	19.34	1.08
557	Inzago	18.12	1.68
558	Lacchiarella	18.74	0.78
559	Lainate	19.98	2.51
560	Lazzate	18.75	0.73
561	Legnano	19.98	0.71
562	Lentate sul Seveso	16.67	1.09
563	Lesmo	17.87	0.80
564	Limbiate	17.90	1.62
565	Liscate	18.49	1.51
566	Lissone	18.20	0.65
567	Locate di Triulzi	14.32	1.59
568	Macherio	17.95	0.88
569	Magenta	20.02	0.90
570	Magnago	19.30	0.96
571	Marcallo con Casone	19.91	1.36
572	Masate	8.42	1.87
573	Meda	17.67	0.91
574	Mediglia	17.78	1.40
575	Melegnano	20.40	0.95
576	Melzo	19.19	0.76
577	Mesero	19.14	1.12
578	Mezzago	18.08	0.77
579	Milano	23.27	1.08
580	Misinto	20.27	1.50
581	Monza	21.06	2.61
582	Morimondo	20.39	2.46
583	Motta Visconti	19.00	0.91
584	Muggi	18.05	0.66
585	Nerviano	19.14	0.68
586	Nosate	19.82	1.96
587	Nova Milanese	18.68	2.40
588	Novate Milanese	19.80	0.74
589	Noviglio	17.31	1.70
590	Opera	19.48	0.94

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
591	Ornago	18.70	0.97
592	Ossona	19.20	0.87
593	Ozzero	19.28	1.17
594	Voltido	14.38	2.24
595	Acquanegra sul Chiese	17.82	1.11
596	Asola	17.96	0.84
597	Bagnolo San Vito	17.17	1.33
598	Bigarello	17.26	1.66
599	Borgoforte	16.87	1.24
600	Borgofranco sul Po	19.60	2.66
601	Bozzolo	17.82	1.08
602	Canneto sull'Oglio	16.05	1.37
603	Carbonara di Po	18.31	2.10
604	Casalmoro	17.31	1.27
605	Casaloldo	17.88	1.07
606	Casalromano	18.37	0.98
607	Castelbelforte	17.37	1.14
608	Castel d'Ario	17.79	1.47
609	Castel Goffredo	19.29	1.94
610	Castellucchio	17.60	1.08
611	Castiglione delle Stiviere	19.57	1.68
612	Cavriana	18.06	1.40
613	Ceresara	17.18	1.14
614	Commessaggio	16.83	1.34
615	Curtatone	16.47	1.30
616	Dosolo	17.08	1.42
617	Felonica	17.81	2.76
618	Gazoldo degli Ippoliti	20.79	1.45
619	Gazzuolo	16.83	1.49
620	Goito	17.69	1.18
621	Gonzaga	17.61	1.82
622	Guidizzolo	18.15	0.90
623	Magnacavallo	17.40	2.13
624	Mantova	22.15	2.06
625	Semiana	20.55	2.93
626	Silvano Pietra	19.33	1.73
627	Siziano	18.80	1.53
628	Sommo	19.15	1.51
629	Spessa	17.51	1.34
630	Stradella	19.03	1.34
631	Suardi	20.73	2.54

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
632	Torrazza Coste	18.51	2.04
633	Torre Beretti e Castellaro	23.15	2.43
634	Torre d'Arese	19.79	1.28
635	Torre de' Negri	17.09	1.89
636	Capralba	15.72	1.42
637	Casalbuttano ed Uniti	18.37	1.19
638	Casale Cremasco-Vidolasco	18.40	1.62
639	Casaletto Ceredano	16.97	1.31
640	Casaletto di Sopra	14.95	1.77
641	Casaletto Vaprio	18.04	0.90
642	Casalmaggiore	18.36	1.14
643	Casalmorano	15.65	1.75
644	Casteldidone	18.19	1.67
645	Castel Gabbiano	15.55	1.60
646	Castelleone	17.69	0.88
647	Castelverde	17.84	1.08
648	Castelvisconti	16.93	1.56
649	Cella Dati	17.52	1.44
650	Chieve	17.02	1.05
651	Cicognolo	16.46	1.46
652	Cingia de' Botti	16.41	1.81
653	Corte de' Cortesi con Cignone	17.34	1.17
654	Torre d'Isola	18.48	1.29
655	Torrevecchia Pia	17.68	1.54
656	Torricella Verzate	17.58	1.83
657	Siccomario	17.54	1.20
658	Trivolzio	17.35	1.64
659	Tromello	20.43	1.33
660	Trovo	19.21	1.26
661	Val di Nizza	20.87	3.29
662	Valeggio	20.67	2.16
663	Valle Lomellina	21.05	1.78
664	Valle Salimbene	18.25	1.20
665	Valverde	19.17	4.32
666	Varzi	20.83	2.51
667	Velezzo Lomellina	19.76	3.06
668	Vellezzo Bellini	18.77	1.69
669	Verretto	20.96	1.98
670	Verrua Po	16.68	2.27
671	Vidigulfo	19.28	1.60
672	Vigevano	20.86	1.23

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
673	Villa Biscossi	22.51	2.17
674	Villanova d'Ardenghi	17.50	1.84
675	Villanterio	19.68	0.92
676	Vistarino	18.28	1.33
677	Voghera	20.24	1.74
678	Volpara	17.92	3.18
679	Paderno Dugnano	18.37	0.63
680	Pantigliate	17.90	0.96
681	Parabiago	20.84	2.08
682	Paullo	17.97	0.64
683	Pero	18.97	0.86
684	Peschiera Borromeo	18.86	1.15
685	Pessano con Bornago	18.46	1.28
686	Pieve Emanuele	21.34	3.56
687	Pioltello	17.95	1.49
688	Pogliano Milanese	19.01	0.84
689	Pozzo d'Adda	18.11	1.00
690	Pozzuolo Martesana	18.54	1.22
691	Pregnana Milanese	18.56	0.73
692	Renate	18.70	0.77
693	Rescaldina	19.28	0.88
694	Rho	19.60	0.61
695	Robecchetto con Induno	19.28	1.02
696	Robecco sul Naviglio	19.42	1.12
697	Rodano	17.92	0.97
698	Roncello	18.16	0.95
699	Ronco Briantino	18.17	1.37
700	Rosate	19.71	1.09
701	Rozzano	19.99	1.71
702	Adrara San Martino	17.98	1.19
703	Adrara San Rocco	18.40	1.60
704	Albano Sant'Alessandro	17.38	1.16
705	Albino	19.48	0.98
706	Alm	17.46	0.73
707	Almenno San Bartolomeo	19.19	1.04
708	Almenno San Salvatore	18.31	0.76
709	Alzano Lombardo	14.66	1.21
710	Ambivere	17.46	0.83
711	Antegnate	17.84	0.78
712	Arcene	17.70	0.87
713	Ardesio	18.33	1.25

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
714	Arzago d'Adda	16.13	1.14
715	Averara	15.57	4.64
716	Aviatico	20.38	3.08
717	Azzano San Paolo	17.76	0.72
718	Azzone	15.15	2.65
719	Bagnatica	18.03	1.32
720	Barbata	15.85	1.14
721	Bariano	16.04	0.99
722	Barzana	18.19	1.42
723	Bedulita	18.63	1.92
724	Berbenno	18.29	1.85
725	Bergamo	22.51	1.97
726	Berzo San Fermo	17.40	0.89
727	Bianzano	19.53	1.66
728	Blello	17.41	2.92
729	Bolgare	17.00	1.31
730	Boltiere	17.92	1.18
731	Bonate Sopra	17.47	0.87
732	Bonate Sotto	17.94	0.73
733	Borgo di Terzo	17.75	1.18
734	Bossico	19.15	2.78
735	Bottanuco	18.90	1.41
736	Bracca	19.19	1.12
737	Branzi	19.85	2.47
738	Brembate	18.91	1.05
739	Brembate di Sopra	17.27	0.83
740	Brembilla	17.95	0.91
741	Brignano Gera d'Adda	17.96	0.78
742	Brumano	23.26	3.16
743	Brusaporto	18.29	1.79
744	Calcinate	18.63	1.96
745	Calcio	17.12	0.71
746	Calusco d'Adda	17.37	0.76
747	Calvenzano	17.41	1.00
748	Camerata Cornello	17.38	2.17
749	Canonica d'Adda	19.79	2.51
750	Capizzone	18.04	0.89
751	Capriate San Gervasio	18.36	1.60
752	Caprino Bergamasco	17.81	0.71
753	Caravaggio	17.76	0.61
754	Carobbio degli Angeli	17.46	1.02

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
755	Carona	20.74	2.59
756	Carvico	17.64	0.68
757	Casazza	16.57	1.72
758	Casirate d'Adda	18.41	1.09
759	Casnigo	19.39	1.25
760	Cassiglio	19.71	2.34
761	Castelli Calepio	17.49	0.85
762	Castel Rozzone	17.89	0.80
763	Castione della Presolana	19.99	3.36
764	Castro	16.94	1.73
765	Cavernago	17.71	2.01
766	Cazzano Sant'Andrea	17.95	1.25
767	Cenate Sopra	17.38	1.01
768	Cenate Sotto	16.76	1.30
769	Cene	19.26	1.00
770	Cerete	19.26	1.44
771	Chignolo d'Isola	18.37	1.18
772	Chiuduno	17.68	0.78
773	Cisano Bergamasco	18.91	0.90
774	Ciserano	19.13	1.58
775	Cividate al Piano	17.57	0.72
776	Clusone	20.45	1.55
777	Colere	16.31	2.05
778	Cologno al Serio	16.23	1.16
779	Colzate	18.32	0.87
780	Comun Nuovo	17.64	1.07
781	Corna Imagna	21.35	3.02
782	Cortenuova	16.79	0.94
783	Costa di Mezzate	18.41	1.70
784	Costa Valle Imagna	19.53	2.12
785	Costa Volpino	18.83	2.31
786	Covo	17.27	0.77
787	Credaro	18.41	1.05
788	Curno	18.26	0.63
789	Cusio	17.50	3.06
790	Dalmine	17.98	2.27
791	Dossena	17.54	1.92
792	Endine Gaiano	16.71	1.18
793	Entratico	16.88	0.81
794	Fara Gera d'Adda	18.54	0.85
795	Fara Olivana con Sola	18.22	1.55

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
796	Filago	17.99	1.11
797	Fino del Monte	17.80	2.11
798	Fiorano al Serio	19.21	1.35
799	Fontanella	18.19	1.01
800	Fonteno	16.82	1.99
801	Foppolo	22.42	3.35
802	Foresto Sparso	17.63	1.11
803	Fornovo San Giovanni	16.75	1.05
804	Fuipiano Valle Imagna	20.18	3.01
805	Gandellino	18.33	2.20
806	Gandino	19.20	0.95
807	Gandosso	15.77	1.50
808	Gaverina Terme	17.73	2.02
809	Gazzaniga	20.26	1.33
810	Gerosa	20.32	2.02
811	Ghisalba	16.63	1.21
812	Gorlago	17.54	1.47
813	Gorle	18.72	0.90
814	Gorno	19.71	1.21
815	Grassobbio	17.13	1.28
816	Gromo	20.96	2.23
817	Grone	17.66	2.61
818	Grumello del Monte	17.63	0.92
819	Isola di Fondra	20.21	2.94
820	Isso	17.27	1.65
821	Lallio	18.71	1.79
822	Leffe	18.76	1.03
823	Lenna	22.26	1.77
824	Levate	16.72	0.86
825	Locatello	20.13	2.31
826	Lovere	18.31	1.25
827	Lurano	18.20	1.15
828	Luzzana	17.87	0.95
829	Madone	18.63	1.34
830	Mapello	18.05	0.83
831	Martinengo	17.88	1.27
832	Mezzoldo	21.29	2.46
833	Misano di Gera d’Adda	16.71	0.94
834	Moio de’ Calvi	15.69	3.61
835	Monasterolo del Castello	18.30	1.40
836	Montello	17.60	1.30

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
837	Morengo	17.62	1.02
838	Mornico al Serio	17.69	1.54
839	Mozzanica	16.58	1.09
840	Mozzo	20.80	1.73
841	Nembro	19.13	0.98
842	Olmo al Brembo	19.95	1.61
843	Oltre il Colle	20.75	2.97
844	Oltressenda Alta	21.62	2.89
845	Oneta	19.01	2.07
846	Onore	19.86	3.06
847	Orio al Serio	19.11	1.26
848	Ornica	22.50	3.00
849	Osio Sopra	17.35	0.90
850	Osio Sotto	17.98	0.78
851	Pagazzano	16.42	1.04
852	Paladina	18.98	0.77
853	Palazzago	17.32	0.75
854	Palosco	17.12	1.12
855	Parre	20.59	1.89
856	Parzanica	14.78	3.27
857	Pedrengo	17.20	1.23
858	Peia	17.53	1.16
859	Pianico	16.47	1.17
860	Piario	20.83	1.88
861	Piazza Brembana	20.25	1.90
862	Piazzatorre	20.41	3.04
863	Piazzolo	18.24	3.13
864	Pognano	17.47	0.81
865	Ponte Nossa	19.80	1.50
866	Ponteranica	18.46	0.66
867	Ponte San Pietro	19.23	0.72
868	Pontida	17.68	0.73
869	Pontirolo Nuovo	19.94	1.35
870	Pradalunga	18.19	0.84
871	Predore	19.50	0.96
872	Premolo	19.41	1.76
873	Presezzo	18.10	0.81
874	Pumenengo	16.35	0.97
875	Ranica	18.59	0.83
876	Ranzanico	19.42	2.16
877	Riva di Solto	18.03	2.15

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
878	Rogno	17.65	1.14
879	Romano di Lombardia	18.03	0.75
880	Roncobello	18.11	3.23
881	Roncola	22.24	3.43
882	Rota d'Imagna	21.25	2.40
883	Rovetta	19.84	1.86
884	San Giovanni Bianco	17.26	1.05
885	San Paolo d'Argon	18.86	1.91
886	San Pellegrino Terme	18.05	1.05
887	Santa Brigida	18.64	2.41
888	Sant'Omobono Terme	20.57	1.98
889	Sarnico	19.30	0.91
890	Scanzorosciate	17.68	0.77
891	Schilpario	17.99	2.24
892	Sedrina	18.74	0.68
893	Selvino	20.78	3.45
894	Torricella del Pizzo	17.83	1.84
895	Trescore Cremasco	17.53	0.89
896	Trigolo	17.63	1.25
897	Vaiano Cremasco	16.30	1.22
898	Vailate	16.01	1.56
899	Vescovato	17.50	0.89
900	Volongo	17.70	1.62
901	Marcaria	17.09	1.20
902	Mariana Mantovana	17.41	1.78
903	Marmirolo	17.12	1.04
904	Medole	19.43	1.66
905	Moglia	17.20	1.53
906	Valbrembo	17.39	0.98
907	Valgoglio	20.29	2.34
908	Valleve	20.40	3.01
909	Valnegra	16.27	3.54
910	Valsecca	18.66	2.58
911	Valtorta	21.57	2.48
912	Vedeseta	21.53	3.20
913	Verdellino	17.86	1.17
914	Verdello	18.11	0.73
915	Vertova	19.30	1.46
916	Viadanica	17.23	1.20
917	Vigano San Martino	16.41	1.22
918	Vigolo	21.87	2.29

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
919	Villa d'Adda	18.57	1.00
920	Villa Alm	17.76	0.83
921	Villa di Serio	17.85	1.12
922	Villa d'Ogna	18.60	1.04
923	Villongo	18.10	1.16
924	Vilminore di Scalve	18.65	1.64
925	Zandobbio	17.97	0.83
926	Zanica	17.32	0.93
927	Zogno	19.34	0.98
928	Costa Serina	19.60	2.91
929	Algua	18.43	2.36
930	Cornalba	19.72	3.03
931	Medolago	17.64	1.36
932	Solza	18.05	1.16
933	Acquafredda	17.06	0.80
934	Adro	16.85	0.80
935	Agnosine	16.65	0.95
936	Alfianello	15.96	1.21
937	Anfo	21.28	2.60
938	Angolo Terme	19.30	1.76
939	Artogne	19.41	3.05
940	Azzano Mella	16.65	1.03
941	Bagnolo Mella	17.23	0.70
942	Bagolino	16.48	1.72
943	Barbariga	16.92	1.14
944	Barghe	16.73	0.94
945	Bassano Bresciano	18.02	1.04
946	Bedizzole	16.98	1.00
947	Berlingo	15.16	1.40
948	Berzo Demo	14.41	2.56
949	Berzo Inferiore	17.12	1.37
950	Bienno	17.95	1.39
951	Bione	17.75	1.88
952	Borgo San Giacomo	15.87	1.37
953	Borgosatollo	16.16	1.13
954	Borno	19.22	2.32
955	Botticino	17.82	0.82
956	Bovegno	16.02	2.34
957	Bovezzo	17.49	0.94
958	Brandico	15.37	1.37
959	Braone	19.43	1.29

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
960	Seriate	18.55	0.85
961	Serina	19.89	2.58
962	Solto Collina	18.73	1.07
963	Songavazzo	20.80	2.60
964	Sorisole	17.27	1.04
965	Sotto il Monte Giovanni XXIII	17.83	1.04
966	Sovere	18.04	0.83
967	Spinone al Lago	19.89	1.52
968	Spirano	17.74	0.95
969	Stezzano	17.64	1.09
970	Strozza	13.31	3.12
971	Suisio	17.96	0.99
972	Taleggio	20.61	2.64
973	Tavernola Bergamasca	17.42	0.94
974	Telgate	16.63	1.99
975	Terno d'Isola	18.13	1.47
976	Torre Boldone	17.87	0.94
977	Torre de' Roveri	16.93	0.98
978	Torre Pallavicina	18.70	1.57
979	Trescore Balneario	18.55	0.89
980	Treviglio	19.37	0.72
981	Treviolo	18.07	1.70
982	Ubiale Clanezzo	17.16	0.88
983	Urgnano	17.79	1.02
984	Valbondione	18.07	2.78
985	Breno	17.68	1.30
986	Brescia	20.07	1.34
987	Brione	17.92	0.98
988	Caino	16.78	1.24
989	Calcinato	17.99	1.28
990	Calvagese della Riviera	16.98	1.33
991	Calvisano	16.64	1.14
992	Capo di Ponte	17.54	1.49
993	Capovalle	16.84	2.32
994	Capriano del Colle	16.20	0.89
995	Capriolo	17.78	0.79
996	Carpenedolo	17.37	1.17
997	Castegnato	16.95	1.29
998	Castelcovati	15.85	1.81
999	Castel Mella	16.97	1.42
1000	Castenedolo	16.82	0.82

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
1001	Casto	17.86	1.43
1002	Castrezzato	17.12	1.70
1003	Cazzago San Martino	15.57	1.22
1004	Cedegolo	16.35	2.13
1005	Cellatica	17.14	0.99
1006	Cerveno	17.97	1.72
1007	Ceto	18.19	1.62
1008	Cevo	17.73	2.39
1009	Chiari	17.81	0.88
1010	Cigole	16.98	0.82
1011	Cimbergo	17.95	2.15
1012	Cividate Camuno	17.02	1.26
1013	Coccaglio	17.37	1.59
1014	Collebeato	17.27	0.87
1015	Collio	17.76	2.17
1016	Cologne	17.43	0.84
1017	Comezzano-Cizzago	15.76	1.68
1018	Concesio	17.05	0.88
1019	Corte Franca	17.40	1.02
1020	Corteno Golgi	19.98	2.77
1021	Corzano	15.78	1.00
1022	Darfo Boario Terme	18.17	0.95
1023	Dello	16.15	0.94
1024	Desenzano del Garda	19.77	1.37
1025	Edolo	18.62	1.79
1026	Erbusco	16.98	0.83
1027	Esine	17.35	1.06
1028	Fiesse	18.12	1.80
1029	Flero	16.49	0.93
1030	Gambara	16.46	0.91
1031	Gardone Riviera	19.45	1.57
1032	Gardone Val Trompia	17.62	0.95
1033	Gargnano	20.51	1.81
1034	Gavardo	17.85	0.92
1035	Ghedi	16.87	1.38
1036	Gianico	17.15	1.01
1037	Gottolengo	16.42	1.15
1038	Gussago	17.74	0.91
1039	Idro	18.82	2.40
1040	Incudine	19.60	2.13
1041	Irma	19.22	1.92

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
1042	Iseo	20.86	1.94
1043	Isorella	16.06	1.27
1044	Lavenone	19.12	1.57
1045	Leno	16.78	1.02
1046	Limone sul Garda	19.89	2.03
1047	Lodrino	15.94	2.08
1048	Lograto	16.68	0.85
1049	Lonato del Garda	18.27	2.09
1050	Longhena	17.07	1.35
1051	Losine	16.11	2.18
1052	Lozio	20.69	2.56
1053	Lumezzane	17.52	1.25
1054	Maclodio	17.14	1.42
1055	Magasa	19.42	2.52
1056	Mairano	15.82	1.02
1057	Malegno	18.09	1.06
1058	Malonno	16.28	1.95
1059	Manerba del Garda	19.76	2.36
1060	Manerbio	18.97	1.06
1061	Marcheno	17.31	1.19
1062	Marmentino	17.84	1.57
1063	Marone	18.44	0.95
1064	Mazzano	17.12	1.28
1065	Milzano	15.93	1.12
1066	Moniga del Garda	21.35	2.80
1067	Monno	17.53	2.11
1068	Monte Isola	17.52	1.37
1069	Monticelli Brusati	17.32	1.24
1070	Montichiari	17.28	1.54
1071	Montirone	16.50	1.70
1072	Mura	17.47	2.06
1073	Muscoline	17.98	1.70
1074	Nave	16.35	1.08
1075	Niardo	17.78	1.34
1076	Nuvolento	17.13	0.94
1077	Nuvolera	16.87	1.28
1078	Odolo	17.31	1.27
1079	Offlaga	15.56	1.28
1080	Ome	18.07	0.84
1081	Ono San Pietro	16.03	1.64
1082	Orzinuovi	15.22	1.12

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
1083	Orzivecchi	15.19	1.59
1084	Ospitaletto	17.24	0.94
1085	Ossimo	18.55	1.36
1086	Padenghe sul Garda	21.00	2.75
1087	Paderno Franciacorta	17.72	1.10
1088	Paisco Loveno	14.64	4.51
1089	Paitone	17.47	1.10
1090	Palazzolo sull'Oglio	18.09	0.62
1091	Paratico	18.17	0.67
1092	Paspardo	17.73	2.36
1093	Passirano	17.43	0.75
1094	Pavone del Mella	16.47	0.83
1095	San Paolo	16.49	1.23
1096	Pertica Alta	18.05	2.10
1097	Pertica Bassa	14.45	1.03
1098	Pezzaze	15.95	2.10
1099	Pian Camuno	18.98	2.20
1100	Pisogne	17.45	1.12
1101	Polaveno	15.79	1.36
1102	Polpenazze del Garda	19.19	1.93
1103	Pompiano	15.85	1.42
1104	Poncarale	16.68	1.09
1105	Ponte di Legno	19.86	3.04
1106	Pontevico	17.49	0.78
1107	Pontoglio	17.04	0.92
1108	Pozzolengo	17.32	1.08
1109	Pralboino	17.31	1.05
1110	Preseglie	15.26	1.59
1111	Prestine	20.36	1.95
1112	Prevalle	18.20	1.30
1113	Provaglio d'Iseo	17.25	0.77
1114	Provaglio Val Sabbia	17.61	1.33
1115	Puegnago sul Garda	16.81	1.61
1116	Quinzano d'Oglio	18.17	0.86
1117	Remedello	16.72	0.90
1118	Rezzato	17.58	0.82
1119	Roccafranca	15.31	1.94
1120	Rodengo Saiano	17.74	1.37
1121	Volciano	18.42	1.09
1122	Roncadelle	17.70	1.07
1123	Rovato	16.80	1.10

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
1124	Rudiano	17.24	1.60
1125	Sabbio Chiese	16.71	1.15
1126	Sale Marasino	19.63	1.30
1127	Sal	19.29	1.21
1128	San Felice del Benaco	18.87	1.90
1129	San Gervasio Bresciano	16.00	1.08
1130	San Zeno Naviglio	17.29	0.78
1131	Sarezzo	19.00	1.15
1132	Saviore dell'Adamello	18.37	2.93
1133	Sellero	15.67	2.20
1134	Seniga	15.23	1.87
1135	Serle	15.97	1.44
1136	Sirmione	21.64	2.82
1137	Soiano del Lago	20.69	3.02
1138	Sonico	16.38	2.42
1139	Sulzano	17.64	0.92
1140	Tavernole sul Mella	16.52	1.76
1141	Tem	20.52	3.21
1142	Tignale	19.17	2.32
1143	Torbole Casaglia	16.97	1.35
1144	Toscolano-Maderno	20.03	1.68
1145	Travagliato	17.67	0.87
1146	Tremosine	20.56	2.53
1147	Trenzano	16.02	1.77
1148	Treviso Bresciano	18.51	1.53
1149	Urago d'Oglio	17.07	0.83
1150	Vallio Terme	19.56	2.14
1151	Valvestino	19.71	2.36
1152	Verolanuova	16.79	0.87
1153	Verolavecchia	10.10	0.77
1154	Vestone	18.17	0.96
1155	Vezza d'Oglio	20.74	2.88
1156	Villa Carcina	17.69	0.74
1157	Villachiara	16.51	1.37
1158	Villanuova sul Clisi	18.36	1.29
1159	Vione	19.54	2.83
1160	Visano	17.77	1.12
1161	Vobarno	17.55	1.97
1162	Zone	19.39	1.50
1163	Piancogno	18.84	1.17
1164	Alagna	20.14	1.37

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
1165	Albaredo Arnaboldi	16.92	2.48
1166	Albonese	20.29	1.67
1167	Albuzzano	17.39	1.38
1168	Arena Po	18.42	1.45
1169	Badia Pavese	21.29	1.88
1170	Bagnaria	19.04	3.02
1171	Barbianello	19.18	1.75
1172	Bascap	19.23	1.68
1173	Bastida de' Dossi	15.57	3.34
1174	Bastida Pancarana	19.34	1.58
1175	Battuda	21.02	1.55
1176	Belgioioso	19.30	1.07
1177	Bereguardo	19.34	1.25
1178	Borgarello	18.08	1.50
1179	Borgo Priolo	18.65	2.61
1180	Borgoratto Mormorolo	20.43	2.76
1181	Borgo San Siro	19.24	1.37
1182	Bornasco	17.41	1.92
1183	Bosnasco	17.88	1.88
1184	Brallo di Pregola	18.81	4.21
1185	Breme	21.86	2.23
1186	Bressana Bottarone	19.19	1.41
1187	Broni	15.30	2.16
1188	Calvignano	19.19	3.27
1189	Campospinoso	18.27	1.30
1190	Candia Lomellina	22.15	2.15
1191	Canevino	16.80	3.29
1192	Canneto Pavese	19.21	2.20
1193	Carbonara al Ticino	19.85	1.25
1194	Casanova Lonati	19.54	1.43
1195	Casatisma	19.10	1.61
1196	Casei Gerola	19.16	1.82
1197	Casorate Primo	20.11	1.29
1198	Cassolnovo	21.04	1.20
1199	Castana	21.00	2.21
1200	Casteggio	18.34	1.85
1201	Castelletto di Branduzzo	18.92	1.44
1202	Castello d'Agogna	20.97	1.88
1203	Castelnovetto	21.17	2.01
1204	Cava Manara	19.08	1.18
1205	Cecima	20.72	3.07

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
1206	Ceranova	18.40	2.28
1207	Ceretto Lomellina	22.49	2.21
1208	Cergnago	22.29	1.74
1209	Certosa di Pavia	18.68	1.05
1210	Cervesina	18.69	2.00
1211	Chignolo Po	17.28	1.32
1212	Cigognola	16.65	2.23
1213	Cilavegna	20.57	1.24
1214	Codevilla	20.32	2.17
1215	Confienza	21.39	1.99
1216	Copiano	17.09	1.41
1217	Corana	18.56	2.86
1218	Cornale	18.46	2.46
1219	Corteolona	19.76	1.08
1220	Corvino San Quirico	18.00	2.05
1221	Costa de' Nobili	17.71	1.75
1222	Cozzo	21.81	1.89
1223	Cura Carpignano	18.45	1.59
1224	Dorno	19.73	1.23
1225	Ferrera Erbognone	20.65	1.59
1226	Filighera	18.19	1.17
1227	Fortunago	16.80	4.09
1228	Frascarolo	20.16	2.25
1229	Galliavola	18.10	2.21
1230	Gambarana	21.99	2.27
1231	Gambol	20.49	1.21
1232	Pinarolo Po	18.23	1.64
1233	Pizzale	18.40	2.17
1234	Ponte Nizza	21.49	2.83
1235	Portalbera	18.78	1.26
1236	Rea	20.27	1.78
1237	Redavalle	17.12	2.47
1238	Retorbido	20.50	1.94
1239	Rivanazzano	20.77	2.36
1240	Robbio	21.35	2.17
1241	Robecco Pavese	21.37	3.49
1242	Rocca de' Giorgi	19.68	3.72
1243	Rocca Susella	19.69	3.16
1244	Rognano	23.76	3.61
1245	Romagnese	20.63	4.04
1246	Roncaro	18.65	1.95

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
1247	Rosasco	21.11	2.42
1248	Rovescala	19.42	2.43
1249	Ruino	20.57	2.63
1250	San Cipriano Po	19.50	1.34
1251	San Damiano al Colle	18.91	2.12
1252	San Genesio ed Uniti	17.98	0.97
1253	San Giorgio di Lomellina	20.93	1.85
1254	San Martino Siccomario	19.67	1.38
1255	Sannazzaro de' Burgondi	18.71	1.57
1256	Santa Cristina e Bissone	19.73	1.21
1257	Santa Giuletta	18.05	1.99
1258	Sant'Alessio con Vialone	18.08	1.53
1259	Santa Margherita di Staffora	19.91	4.25
1260	Santa Maria della Versa	20.00	1.58
1261	Sant'Angelo Lomellina	18.46	2.14
1262	San Zenone al Po	18.42	2.32
1263	Sartirana Lomellina	19.70	2.19
1264	Scaldasole	18.85	1.42
1265	Corte de' Frati	17.08	1.34
1266	Credera Rubbiano	17.34	1.33
1267	Crema	18.68	1.75
1268	Cremona	19.26	1.22
1269	Cremosano	16.57	0.84
1270	Crotta d'Adda	17.40	1.38
1271	Cumignano sul Naviglio	16.71	1.86
1272	Derovere	18.15	1.96
1273	Dovera	17.29	0.93
1274	Drizzona	17.37	1.36
1275	Fiesco	17.02	1.43
1276	Formigara	16.85	1.37
1277	Gabbioneta-Binanuova	18.01	1.19
1278	Gadesco-Pieve Delmona	17.10	1.54
1279	Genivolta	15.88	1.76
1280	Gerre de' Caprioli	17.40	1.16
1281	Gombito	17.80	1.02
1282	Grontardo	16.46	1.42
1283	Grumello Cremonese ed Uniti	17.06	1.31
1284	Gussola	16.83	1.18
1285	Isola Dovarese	17.69	1.25
1286	Izano	15.77	1.15
1287	Madignano	17.27	1.03

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
1288	Malagnino	16.90	0.99
1289	Martignana di Po	17.28	1.19
1290	Monte Cremasco	17.72	0.90
1291	Montodine	15.66	1.33
1292	Moscazzano	17.19	1.04
1293	Motta Baluffi	15.75	1.70
1294	Offanengo	16.99	0.82
1295	Olmeneta	16.28	1.26
1296	Ostiano	18.89	1.04
1297	Paderno Ponchielli	16.79	1.48
1298	Palazzo Pignano	16.83	0.98
1299	Pandino	18.10	0.74
1300	Persico Dosimo	15.85	1.47
1301	Pescarolo ed Uniti	17.53	0.91
1302	Pessina Cremonese	16.78	1.23
1303	Piadena	18.67	0.97
1304	Pianengo	16.26	1.15
1305	Pieranica	16.59	1.31
1306	Pieve d'Olmi	16.09	1.24
1307	Pieve San Giacomo	15.91	1.20
1308	Pizzighettone	17.10	1.14
1309	Pozzaglio ed Uniti	16.87	0.89
1310	Quintano	16.96	1.14
1311	Ricengo	16.95	1.26
1312	Ripalta Arpina	14.94	1.69
1313	Ripalta Cremasca	17.08	0.85
1314	Ripalta Guerina	15.64	1.38
1315	Rivarolo del Re ed Uniti	16.72	1.38
1316	Rivolta d'Adda	17.61	0.82
1317	Robecco d'Oglio	16.01	1.46
1318	Romanengo	17.76	0.77
1319	Salvirola	15.87	1.20
1320	San Bassano	17.27	0.80
1321	San Daniele Po	16.53	1.98
1322	San Giovanni in Croce	18.93	1.13
1323	San Martino del Lago	17.00	1.75
1324	Scandolara Ravara	15.82	1.88
1325	Scandolara Ripa d'Oglio	18.06	0.93
1326	Sergnano	16.44	0.86
1327	Sesto ed Uniti	17.88	1.08
1328	Solarolo Rainerio	16.47	1.24

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
1329	Soncino	18.04	0.79
1330	Soresina	18.97	1.18
1331	Sospiro	18.10	0.89
1332	Spinadesco	17.33	1.05
1333	Spineda	18.21	1.64
1334	Spino d'Adda	18.74	1.24
1335	Stagno Lombardo	15.91	1.39
1336	Ticengo	18.01	1.49
1337	Torlino Vimercati	17.16	1.17
1338	Tornata	18.20	1.13
1339	Torre de' Picenardi	16.89	1.36
1340	Zavattarello	18.24	3.21
1341	Zeccone	17.84	1.02
1342	Zeme	20.30	1.97
1343	Zenevredo	18.40	1.36
1344	Zerbo	18.80	1.94
1345	Zerbol	18.82	1.45
1346	Zinasco	18.62	1.40
1347	Acquanegra Cremonese	16.65	1.46
1348	Agnadello	17.30	0.80
1349	Annicco	17.04	1.70
1350	Azzanello	18.81	1.56
1351	Bagnolo Cremasco	17.53	0.64
1352	Bonemerse	16.02	1.21
1353	Bordolano	16.96	1.18
1354	Ca' d'Andrea	16.80	1.69
1355	Calvatone	18.14	1.34
1356	Camisano	16.55	1.72
1357	Campagnola Cremasca	15.41	1.39
1358	Capergnanica	15.54	1.63
1359	Cappella Cantone	16.58	1.59
1360	Cappella de' Picenardi	14.99	2.60
1361	Monzambano	18.95	1.49
1362	Motteggiana	16.63	1.44
1363	Ostiglia	17.61	1.91
1364	Pegognaga	16.56	1.43
1365	Pieve di Coriano	18.30	1.96
1366	Piubega	17.61	1.65
1367	Poggio Rusco	17.86	1.95
1368	Pomponesco	18.36	1.32
1369	Ponti sul Mincio	19.97	2.07

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
1370	Porto Mantovano	17.12	1.26
1371	Quingentole	19.57	1.75
1372	Quistello	17.66	1.54
1373	Redondesco	17.74	1.41
1374	Revere	18.82	1.79
1375	Rivarolo Mantovano	18.13	1.25
1376	Rodigo	18.02	1.31
1377	Roncoferraro	16.90	1.33
1378	Roverbella	16.90	1.12
1379	Sabbioneta	18.33	1.25
1380	San Benedetto Po	17.18	1.41
1381	San Giacomo delle Segnate	16.16	1.97
1382	San Giorgio di Mantova	17.81	1.64
1383	San Giovanni del Dosso	17.96	1.83
1384	San Martino dall'Argine	17.27	1.47
1385	Schivenoglia	16.21	1.91
1386	Sermide	18.43	2.30
1387	Serravalle a Po	16.82	2.04
1388	Solferino	18.60	1.09
1389	Sustinente	18.62	1.65
1390	Suzzara	16.89	1.26
1391	Viadana	18.38	1.28
1392	Villa Poma	17.54	1.80
1393	Villimpenta	18.57	1.62
1394	Virgilio	18.02	1.57
1395	Volta Mantovana	17.69	1.17
1396	Abbadia Lariana	20.81	1.53
1397	Airuno	19.11	0.82
1398	Annone di Brianza	17.14	1.28
1399	Ballabio	20.52	2.64
1400	Barzago	19.08	0.76
1401	Barzan	17.95	0.71
1402	Barzio	21.59	2.64
1403	Bellano	21.45	1.35
1404	Bosisio Parini	19.28	0.81
1405	Brivio	18.61	0.60
1406	Bulciago	18.92	0.90
1407	Calco	17.38	0.96
1408	Calolziocorte	18.81	1.58
1409	Carenno	20.21	2.13
1410	Casargo	22.61	2.68

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
1411	Casatenovo	18.15	0.87
1412	Cassago Brianza	17.89	0.80
1413	Cassina Valsassina	22.38	3.30
1414	Castello di Brianza	19.04	0.97
1415	Cernusco Lombardone	18.54	0.84
1416	Cesana Brianza	18.26	0.95
1417	Civate	18.87	0.88
1418	Colico	19.39	2.04
1419	Colle Brianza	21.32	1.40
1420	Cortenova	19.63	1.39
1421	Costa Masnaga	19.20	0.78
1422	Crandola Valsassina	24.94	2.48
1423	Cremella	18.82	0.94
1424	Cremeno	21.39	3.02
1425	Dervio	22.05	1.58
1426	Dolzago	18.49	0.88
1427	Dorio	23.58	2.00
1428	Ello	19.71	1.50
1429	Erve	20.12	1.24
1430	Esino Lario	20.32	2.50
1431	Galbiate	18.88	0.72
1432	Garbagnate Monastero	18.23	0.84
1433	Garlate	18.64	1.03
1434	Imbersago	18.82	0.72
1435	Introbio	21.62	1.93
1436	Introzzo	23.01	2.93
1437	Lecco	21.97	1.99
1438	Lierna	22.03	2.20
1439	Lomagna	18.41	0.72
1440	Malgrate	18.94	0.71
1441	Mandello del Lario	19.11	1.00
1442	Margno	23.89	3.12
1443	Merate	18.34	0.93
1444	Missaglia	18.14	0.62
1445	Moggio	21.26	3.54
1446	Molteno	18.55	0.87
1447	Monte Marenzo	18.58	1.50
1448	Montevecchia	17.63	0.77
1449	Monticello Brianza	19.65	1.45
1450	Morterone	21.06	3.29
1451	Nibionno	19.19	0.83

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
1452	Oggiono	18.76	0.64
1453	Olgiate Molgora	18.45	0.72
1454	Olginate	18.17	0.63
1455	Oliveto Lario	21.53	2.49
1456	Osnago	19.19	0.83
1457	Paderno d'Adda	19.65	1.03
1458	Pagnona	18.54	1.36
1459	Parlasco	17.95	2.80
1460	Pasturo	20.31	2.21
1461	Perego	19.52	0.88
1462	Perledo	23.06	2.34
1463	Pescate	19.17	1.04
1464	Premana	17.07	1.89
1465	Primaluna	21.24	2.05
1466	Robbiate	18.05	0.77
1467	Rogeno	18.52	0.77
1468	Rovagnate	18.91	0.64
1469	Santa Maria Ho	18.59	1.03
1470	Sirone	17.93	0.74
1471	Sirtori	18.50	1.10
1472	Sueglio	18.33	3.60
1473	Suello	18.07	1.06
1474	Taceno	20.44	2.77
1475	Torre de' Busi	19.12	1.75
1476	Tremenico	19.23	2.33
1477	Valgreghentino	17.62	1.34
1478	Valmadrera	18.34	0.78
1479	Varenna	24.17	2.33
1480	Vendrogno	23.16	2.80
1481	Vercurago	17.35	1.01
1482	Verderio Inferiore	18.15	1.32
1483	Verderio Superiore	18.76	1.49
1484	Vestreno	21.31	2.08
1485	Vigan	18.27	0.78
1486	Abbadia Cerreto	18.41	1.58
1487	Bertonico	15.27	1.84
1488	Boffalora d'Adda	15.77	1.37
1489	Borghetto Lodigiano	17.85	1.03
1490	Borgo San Giovanni	16.34	1.56
1491	Brembio	17.49	1.13
1492	Camairago	17.58	1.97

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
1493	Casaletto Lodigiano	17.48	0.84
1494	Casalmaiocco	18.60	1.57
1495	Casalpusterlengo	18.76	1.61
1496	Caselle Landi	15.45	2.11
1497	Caselle Lurani	16.44	1.64
1498	Castelnuovo Bocca d'Adda	18.80	1.47
1499	Castiglione d'Adda	16.64	0.92
1500	Castiraga Vidardo	16.52	1.07
1501	Cavacurta	15.44	1.55
1502	Cavenago d'Adda	16.44	1.28
1503	Cervignano d'Adda	18.18	1.25
1504	Codogno	18.72	1.37
1505	Comazzo	18.51	1.55
1506	Cornegliano Laudense	19.97	1.55
1507	Corno Giovine	18.01	1.30
1508	Cornovecchio	16.95	1.52
1509	Corte Palasio	17.95	0.99
1510	Crespiatica	17.11	0.80
1511	Fombio	16.71	1.04
1512	Galgagnano	17.45	2.07
1513	Graffignana	17.15	1.31
1514	Guardamiglio	17.32	0.96
1515	Livraga	18.61	0.98
1516	Lodi	19.23	0.89
1517	Lodi Vecchio	18.61	0.80
1518	Maccastorna	16.33	3.56
1519	Mairago	18.11	0.78
1520	Maleo	17.03	1.13
1521	Marudo	16.73	1.20
1522	Massalengo	17.00	1.02
1523	Meleti	16.57	2.33
1524	Merlino	19.19	2.05
1525	Montanaso Lombardo	16.76	0.90
1526	Mulazzano	18.20	0.88
1527	Orio Litta	17.57	1.12
1528	Ospedaletto Lodigiano	17.60	1.12
1529	Ossago Lodigiano	17.75	0.84
1530	Pieve Fissiraga	19.82	2.30
1531	Salerano sul Lambro	17.86	0.90
1532	San Fiorano	18.33	1.14
1533	San Martino in Strada	16.62	0.96

Continued on next page

Table B.1 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
1534	San Rocco al Porto	17.82	1.02
1535	Sant'Angelo Lodigiano	18.87	1.04
1536	Santo Stefano Lodigiano	19.30	2.04
1537	Secugnago	17.16	1.15
1538	Senna Lodigiana	16.11	1.66
1539	Somaglia	17.69	1.14
1540	Sordio	18.08	1.17
1541	Tavazzano con Villavesco	18.23	1.22
1542	Terranova dei Passerini	16.64	1.24
1543	Turano Lodigiano	16.82	1.35
1544	Valera Fratta	17.27	1.10
1545	Villanova del Sillaro	17.64	1.49
1546	Zelo Buon Persico	17.38	1.11

Table B.2: Estimated Mean of Household Equivalised Income (x1000) (MEAN) and estimated Root Mean Squared Error of the Mean estimator (RMSE) for the Toscana Municipalities. Semi-parametric Fay and Herriot model.

	MUNICIPALITY	MEAN	RMSE
1	Aulla	16.50	1.07
2	Bagnone	16.19	1.79
3	Carrara	14.59	0.93
4	Casola in Lunigiana	14.82	1.68
5	Comano	14.52	1.85
6	Filattiera	17.03	1.93
7	Fivizzano	17.80	1.29
8	Fosdinovo	15.27	1.08
9	Licciana Nardi	15.94	1.07
10	Massa	13.04	1.15
11	Montignoso	14.37	0.98
12	Mulazzo	17.32	1.72
13	Podenzana	17.02	1.42
14	Pontremoli	14.98	1.94
15	Tresana	17.04	2.08
16	Villafranca in Lunigiana	15.54	1.23
17	Zeri	15.27	2.68
18	Altopascio	16.94	0.53

Continued on next page

Table B.2 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
19	Bagni di Lucca	16.25	0.97
20	Barga	17.15	0.72
21	Borgo a Mozzano	18.01	0.49
22	Camaiore	15.43	0.89
23	Camporgiano	15.62	0.96
24	Capannori	17.41	1.09
25	Careggine	15.86	0.99
26	Castelnuovo di Garfagnana	15.61	0.91
27	Castiglione di Garfagnana	14.94	1.19
28	Coreglia Antelminelli	17.71	0.68
29	Fabbriche di Vallico	17.28	0.99
30	Forte dei Marmi	13.08	1.56
31	Fosciandora	15.11	1.15
32	Gallicano	17.97	0.59
33	Giuncugnano	13.83	1.71
34	Lucca	19.52	1.08
35	Massarosa	17.18	0.63
36	Minucciano	14.38	1.34
37	Molazzana	15.44	0.98
38	Montecarlo	16.36	0.69
39	Pescaglia	15.98	1.23
40	Piazza al Serchio	13.33	1.20
41	Pietrasanta	13.67	1.05
42	Pieve Fosciana	15.80	0.97
43	Porcari	17.11	0.64
44	San Romano in Garfagnana	13.14	1.19
45	Seravezza	15.14	0.97
46	Sillano	13.33	1.55
47	Stazzema	13.92	1.08
48	Vagli Sotto	13.84	1.39
49	Vergemoli	20.04	2.05
50	Viareggio	15.37	1.08
51	Villa Basilica	17.61	1.26
52	Villa Collemandina	15.83	1.01
53	Abetone	12.23	2.30
54	Agliana	18.28	1.41
55	Buggiano	18.79	1.44
56	Cutigliano	15.76	1.87
57	Lamporecchio	17.81	0.59
58	Larciano	17.20	0.80
59	Marliana	16.75	1.63

Continued on next page

Table B.2 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
60	Massa e Cozzile	19.82	0.72
61	Monsummano Terme	17.36	0.58
62	Montale	17.59	0.79
63	Montecatini-Terme	18.98	0.98
64	Pescia	18.66	1.33
65	Pieve a Nievole	17.57	1.25
66	Pistoia	18.99	1.11
67	Piteglio	17.31	1.26
68	Ponte Buggianese	18.57	0.64
69	Quarrata	17.88	1.21
70	Sambuca Pistoiese	15.84	1.73
71	San Marcello Pistoiese	20.59	1.29
72	Serravalle Pistoiese	17.86	0.64
73	Uzzano	18.65	0.79
74	Chiesina Uzzanese	17.73	0.66
75	Bagno a Ripoli	20.14	1.08
76	Barberino di Mugello	17.08	0.77
77	Barberino Val d’Elsa	18.06	1.18
78	Borgo San Lorenzo	18.20	0.62
79	Calenzano	18.35	0.61
80	Campi Bisenzio	17.98	1.16
81	Capraia e Limite	19.53	0.91
82	Castelfiorentino	16.66	0.95
83	Cerreto Guidi	17.42	0.65
84	Certaldo	18.26	1.07
85	Dicomano	18.30	0.79
86	Empoli	17.63	0.85
87	Fiesole	20.05	0.82
88	Figline Valdarno	18.08	0.97
89	Firenze	22.00	1.20
90	Firenzuola	17.74	1.39
91	Fucecchio	17.70	1.28
92	Gambassi Terme	17.85	0.73
93	Greve in Chianti	18.39	0.69
94	Impruneta	18.34	0.46
95	Incisa in Val d’Arno	17.18	1.08
96	Lastra a Signa	19.98	1.32
97	Londa	18.52	1.62
98	Marradi	17.31	1.38
99	Montaione	18.92	1.00
100	Montelupo Fiorentino	19.38	0.57

Continued on next page

Table B.2 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
101	Montespertoli	18.77	0.96
102	Palazzuolo sul Senio	18.09	1.52
103	Pelago	19.40	1.33
104	Pontassieve	18.89	1.13
105	Reggello	18.43	0.96
106	Rignano sull'Arno	18.64	0.58
107	Rufina	20.05	0.78
108	San Casciano in Val di Pesa	17.63	0.54
109	San Godenzo	17.04	1.32
110	San Piero a Sieve	17.28	0.64
111	Scandicci	20.58	1.18
112	Scarperia	18.44	0.82
113	Sesto Fiorentino	20.58	1.52
114	Signa	18.67	0.53
115	Tavarnelle Val di Pesa	17.82	0.58
116	Vaglia	20.10	1.07
117	Vicchio	17.71	0.83
118	Vinci	18.35	0.63
119	Bibbona	14.57	2.41
120	Campiglia Marittima	18.24	1.25
121	Campo nell'Elba	15.46	1.71
122	Capoliveri	16.48	1.68
123	Capraia Isola	17.02	3.05
124	Castagneto Carducci	16.82	1.39
125	Cecina	19.00	1.15
126	Collesalvetti	17.92	0.66
127	Livorno	18.40	1.09
128	Marciana	14.62	2.12
129	Marciana Marina	15.77	1.66
130	Piombino	19.15	1.14
131	Porto Azzurro	15.98	1.65
132	Portoferraio	16.37	1.26
133	Rio Marina	14.10	1.57
134	Rio nell'Elba	13.62	2.12
135	Rosignano Marittimo	17.95	0.92
136	San Vincenzo	15.60	1.28
137	Sassetta	16.97	1.30
138	Suvereto	18.59	0.82
139	Bientina	17.95	0.67
140	Buti	18.63	0.69
141	Calci	18.13	0.79

Continued on next page

Table B.2 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
142	Calcinaia	18.57	1.53
143	Capannoli	17.78	0.72
144	Casale Marittimo	17.42	0.95
145	Casciana Terme	17.78	0.66
146	Cascina	18.05	0.58
147	Castelfranco di Sotto	17.80	0.56
148	Castellina Marittima	18.31	1.15
149	Castelnuovo di Val di Cecina	17.49	1.14
150	Chianni	19.01	0.82
151	Crespina	17.27	0.75
152	Fauglia	17.19	0.69
153	Guardistallo	18.89	1.17
154	Lajatico	18.36	1.28
155	Lari	17.96	1.42
156	Lorenzana	17.13	0.83
157	Montecatini Val di Cecina	15.64	1.16
158	Montescudaio	18.38	1.28
159	Monteverdi Marittimo	16.38	1.53
160	Montopoli in Val d'Arno	17.75	0.74
161	Orciano Pisano	15.03	1.16
162	Palaia	16.84	0.64
163	Peccioli	18.10	0.80
164	Pisa	18.93	1.03
165	Pomarance	17.44	1.24
166	Ponsacco	17.70	0.71
167	Pontedera	18.60	0.51
168	Riparbella	18.63	1.06
169	San Giuliano Terme	18.91	1.27
170	San Miniato	18.68	0.56
171	Santa Croce sull'Arno	17.44	0.73
172	Santa Luce	16.50	1.06
173	Santa Maria a Monte	17.54	0.73
174	Terricciola	17.78	0.73
175	Vecchiano	17.70	0.78
176	Vicopisano	18.81	0.52
177	Volterra	19.65	0.94
178	Anghiari	18.45	1.13
179	Arezzo	20.93	1.48
180	Badia Tedalda	19.48	1.74
181	Bibbiena	19.29	1.27
182	Bucine	18.72	1.10

Continued on next page

Table B.2 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
183	Capolona	19.74	0.78
184	Caprese Michelangelo	18.90	1.02
185	Castel Focognano	18.45	0.66
186	Castelfranco di Sopra	18.47	0.46
187	Castel San Niccol˜	17.04	1.06
188	Castiglion Fibocchi	19.14	0.87
189	Castiglion Fiorentino	17.34	1.43
190	Cavriglia	19.39	0.60
191	Chitignano	18.89	0.93
192	Chiusi della Verna	18.37	0.95
193	Civitella in Val di Chiana	18.70	0.81
194	Cortona	19.69	1.41
195	Foiano della Chiana	18.66	0.89
196	Laterina	19.03	0.98
197	Loro Ciuffenna	18.93	1.02
198	Lucignano	16.56	1.08
199	Marciano della Chiana	19.27	1.05
200	Montemignaio	17.38	1.47
201	Monterchi	17.53	1.46
202	Monte San Savino	18.62	0.81
203	Montevarchi	20.52	0.74
204	Ortignano Raggiolo	16.70	1.05
205	Pergine Valdarno	18.45	0.81
206	Pian di Sco	18.56	1.02
207	Pieve Santo Stefano	19.09	1.09
208	Poppi	17.25	0.47
209	Pratovecchio	20.73	1.08
210	San Giovanni Valdarno	20.48	0.70
211	Sansepolcro	21.16	1.83
212	Sestino	17.10	1.77
213	Stia	20.48	1.09
214	Subbiano	20.20	0.99
215	Talla	18.28	1.01
216	Terranuova Bracciolini	19.23	0.58
217	Abbadia San Salvatore	19.02	1.59
218	Asciano	18.87	1.01
219	Buonconvento	20.05	1.06
220	Casole d'Elsa	18.63	0.77
221	Castellina in Chianti	18.30	1.19
222	Castelnuovo Berardenga	20.19	1.04
223	Castiglione d'Orcia	17.91	1.17

Continued on next page

Table B.2 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
224	Cetona	20.94	1.56
225	Chianciano Terme	23.78	1.64
226	Chiusdino	19.27	1.01
227	Chiusi	20.34	1.22
228	Colle di Val d'Elsa	19.60	0.85
229	Gaiole in Chianti	17.63	0.92
230	Montalcino	21.46	1.18
231	Montepulciano	20.85	1.66
232	Monteriggioni	19.92	0.93
233	Monteroni d'Arbia	20.06	0.97
234	Monticiano	18.72	1.29
235	Murlo	21.31	1.13
236	Piancastagnaio	18.78	1.07
237	Pienza	20.87	1.28
238	Poggibonsi	19.19	0.67
239	Radda in Chianti	19.46	1.37
240	Radicofani	19.02	1.45
241	Radicondoli	17.45	1.33
242	Rapolano Terme	19.38	1.33
243	San Casciano dei Bagni	17.67	1.49
244	San Gimignano	18.70	0.62
245	San Giovanni d'Asso	18.53	1.11
246	San Quirico d'Orcia	21.03	1.00
247	Sarteano	21.02	1.33
248	Siena	21.34	1.33
249	Sinalunga	19.87	0.86
250	Sovicille	20.39	1.46
251	Torrita di Siena	19.27	1.53
252	Trequanda	19.38	1.11
253	Arcidosso	20.26	1.21
254	Campagnatico	18.07	0.89
255	Capalbio	15.84	1.62
256	Castel del Piano	19.65	0.95
257	Castell'Azzara	19.66	1.77
258	Castiglione della Pescaia	16.85	1.80
259	Cinigiano	20.00	1.18
260	Civitella Paganico	17.29	1.01
261	Follonica	16.98	1.02
262	Gavorrano	17.92	1.19
263	Grosseto	19.10	1.39
264	Isola del Giglio	14.55	1.46

Continued on next page

Table B.2 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
265	Magliano in Toscana	19.26	1.75
266	Manciano	21.06	1.54
267	Massa Marittima	18.75	1.04
268	Monte Argentario	14.52	1.71
269	Montieri	16.30	1.62
270	Orbetello	17.86	1.44
271	Pitigliano	19.68	1.48
272	Roccalbegna	22.03	1.75
273	Roccastrada	18.15	1.00
274	Santa Fiora	19.10	1.34
275	Scansano	18.75	1.02
276	Scarlino	19.37	1.47
277	Seggiano	19.95	1.59
278	Sorano	17.87	1.45
279	Monterotondo Marittimo	16.37	1.12
280	Semproniano	21.62	1.74
281	Cantagallo	16.88	0.92
282	Carmignano	18.20	1.37
283	Montemurlo	17.14	1.02
284	Poggio a Caiano	16.44	0.79
285	Prato	18.71	0.91
286	Vaiano	19.22	1.11
287	Vernio	18.15	0.75

Table B.3: Estimated Mean of Household Equivalised Income (x1000) (MEAN) and estimated Root Mean Squared Error of the Mean estimator (RMSE) for the Campania Municipalities. Semi-parametric Fay and Herriot model.

	MUNICIPALITY	MEAN	RMSE
1	Ailano	10.27	1.85
2	Alife	12.01	0.92
3	Alvignano	13.09	0.97
4	Arienzo	14.03	0.88
5	Aversa	13.34	1.41
6	Baia e Latina	14.18	1.15
7	Bellona	13.51	0.95
8	Caianello	9.91	1.61

Continued on next page

Table B.3 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
9	Caiazzo	10.82	1.43
10	Calvi Risorta	12.80	1.04
11	Camigliano	12.69	1.06
12	Cancello ed Arnone	12.62	1.38
13	Capodrise	12.71	1.13
14	Capriati a Volturno	12.70	1.85
15	Capua	13.54	0.72
16	Carinaro	11.30	1.30
17	Carinola	13.02	1.78
18	Casagiove	13.28	0.60
19	Casal di Principe	10.57	3.16
20	Casaluce	9.81	1.48
21	Casapulla	12.48	0.91
22	Caserta	14.06	0.83
23	Castel Campagnano	12.05	1.25
24	Castel di Sasso	11.14	2.19
25	Castello del Matese	13.89	1.11
26	Castel Morrone	12.72	1.73
27	Castel Volturno	8.79	1.30
28	Cervino	13.57	1.07
29	Cesa	10.54	1.12
30	Ciorlano	10.55	2.21
31	Conca della Campania	11.15	1.72
32	Curti	11.83	0.78
33	Dragoni	12.96	0.79
34	Fontegreca	8.11	2.57
35	Formicola	14.31	1.45
36	Francolise	12.75	1.11
37	Frignano	11.39	1.76
38	Gallo Matese	14.06	1.70
39	Galluccio	12.46	1.31
40	Giano Vetusto	13.25	1.44
41	Gioia Sannitica	10.20	1.57
42	Grazzanise	15.15	1.22
43	Gricignano di Aversa	9.72	1.55
44	Letino	11.63	1.41
45	Liberi	13.24	1.79
46	Lusciano	11.47	0.93
47	Macerata Campania	11.37	0.86
48	Maddaloni	13.81	0.88
49	Marcianise	15.15	1.42

Continued on next page

Table B.3 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
50	Marzano Appio	11.97	2.16
51	Mignano Monte Lungo	12.11	1.49
52	Mondragone	10.61	1.09
53	Orta di Atella	9.20	2.02
54	Parete	11.60	1.71
55	Pastorano	12.63	1.30
56	Piana di Monte Verna	15.32	1.31
57	Piedimonte Matese	13.92	0.97
58	Pietramelara	11.56	1.20
59	Pietravairano	13.89	1.07
60	Pignataro Maggiore	13.26	1.62
61	Pontelatone	13.48	1.19
62	Portico di Caserta	12.46	1.09
63	Prata Sannita	12.80	1.26
64	Pratella	13.99	1.28
65	Presenzano	12.02	1.48
66	Raviscanina	11.39	1.28
67	Recale	11.67	0.91
68	Riardo	12.66	1.07
69	Rocca d’Evandro	12.18	1.96
70	Roccamonfina	11.51	1.30
71	Roccaromana	13.86	1.82
72	Rocchetta e Croce	11.08	1.77
73	Ruviano	9.23	2.45
74	San Cipriano d’Aversa	10.36	3.07
75	San Felice a Cancello	11.41	1.02
76	San Gregorio Matese	12.31	1.30
77	San Marcellino	10.25	1.78
78	San Nicola la Strada	12.36	0.90
79	San Pietro Infine	12.72	1.64
80	San Potito Sannitico	10.48	1.26
81	San Prisco	12.45	0.89
82	Santa Maria a Vico	12.36	0.60
83	Santa Maria Capua Vetere	13.75	0.82
84	Santa Maria la Fossa	16.12	1.26
85	San Tammaro	10.43	1.61
86	Sant’Angelo d’Alife	12.67	1.20
87	Sant’Arpino	11.00	1.00
88	Sessa Aurunca	11.73	1.20
89	Sparanise	13.04	1.14
90	Succivo	12.08	0.81

Continued on next page

Table B.3 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
91	Teano	15.59	1.34
92	Teverola	10.20	1.55
93	Tora e Piccilli	14.00	1.49
94	Trentola-Ducenta	12.10	1.53
95	Vairano Patenora	10.94	1.38
96	Valle Agricola	10.56	1.55
97	Valle di Maddaloni	11.26	0.94
98	Villa di Briano	11.82	1.76
99	Villa Literno	8.78	2.54
100	Vitulazio	13.14	0.87
101	Falciano del Massico	13.34	1.36
102	Cellole	10.16	2.30
103	Casapesenna	9.13	4.18
104	San Marco Evangelista	10.59	1.25
105	Airola	14.63	0.81
106	Amorosi	15.56	1.11
107	Apice	12.78	1.09
108	Apollosa	13.93	1.51
109	Arpaia	10.59	1.00
110	Arpaise	10.70	1.75
111	Baselice	11.85	1.54
112	Benevento	14.77	0.81
113	Bonea	11.59	1.24
114	Bucciano	11.99	0.94
115	Buonalbergo	10.48	1.30
116	Calvi	12.41	0.95
117	Campolattaro	11.51	1.45
118	Campoli del Monte Taburno	9.77	1.66
119	Casalduni	14.25	1.15
120	Castelfranco in Miscano	12.93	1.27
121	Castelpagano	11.30	1.56
122	Castelpoto	12.69	1.24
123	Castelvenere	11.96	1.31
124	Castelvetere in Val Fortore	11.61	2.15
125	Cautano	11.25	1.87
126	Ceppaloni	12.34	1.05
127	Cerreto Sannita	12.46	0.90
128	Circello	16.07	2.12
129	Colle Sannita	14.71	1.20
130	Cusano Mutri	12.14	1.04
131	Dugenta	12.50	1.31

Continued on next page

Table B.3 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
132	Durazzano	13.62	0.98
133	Faicchio	11.88	1.18
134	Foglianise	13.63	0.79
135	Foiano di Val Fortore	13.78	1.51
136	Forchia	11.63	1.13
137	Fragneto l'Abate	8.26	2.47
138	Fragneto Monforte	12.41	1.21
139	Frasso Telesino	14.39	1.12
140	Ginestra degli Schiavoni	11.52	1.72
141	Guardia Sanframondi	11.08	1.49
142	Limatola	9.46	2.14
143	Melizzano	13.14	1.09
144	Moiano	11.99	1.24
145	Molinara	13.11	1.19
146	Montefalcone di Val Fortore	13.80	1.36
147	Montesarchio	13.13	0.57
148	Morcone	13.59	1.25
149	Paduli	11.48	1.14
150	Pago Veiano	11.98	0.80
151	Pannarano	13.92	0.91
152	Paolisi	13.63	1.00
153	Paupisi	10.84	1.65
154	Pesco Sannita	13.05	1.21
155	Pietraroja	11.87	1.37
156	Pietrelcina	12.22	1.04
157	Ponte	11.90	1.64
158	Pontelandolfo	10.88	1.35
159	Puglianello	11.87	1.22
160	Reino	12.26	1.30
161	San Bartolomeo in Galdo	11.17	1.01
162	San Giorgio del Sannio	13.75	0.88
163	San Giorgio La Molara	12.44	1.18
164	San Leucio del Sannio	12.90	0.94
165	San Lorenzello	12.50	0.95
166	San Lorenzo Maggiore	13.14	1.03
167	San Lupo	14.44	1.80
168	San Marco dei Cavoti	13.69	1.33
169	San Martino Sannita	11.70	1.01
170	San Nazzaro	13.90	1.04
171	San Nicola Manfredi	12.62	1.04
172	San Salvatore Telesino	13.60	1.96

Continued on next page

Table B.3 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
173	Santa Croce del Sannio	12.71	1.71
174	Sant'Agata de' Goti	11.27	1.42
175	Sant'Angelo a Cupolo	13.32	0.88
176	Sassinoro	12.44	2.30
177	Solopaca	14.00	0.98
178	Telese Terme	13.45	1.49
179	Tocco Caudio	11.65	1.29
180	Torrecuso	13.85	1.72
181	Vitulano	13.38	1.26
182	Sant'Arcangelo Trimonte	11.98	2.67
183	Acerra	12.02	0.73
184	Afragola	12.15	0.96
185	Agerola	11.17	0.91
186	Anacapri	12.17	1.25
187	Arzano	12.48	1.02
188	Bacoli	12.75	0.72
189	Barano d'Ischia	13.11	1.51
190	Boscoreale	12.00	0.87
191	Boscotrecase	12.43	1.11
192	Brusciano	12.75	0.67
193	Caivano	11.22	1.11
194	Calvizzano	11.35	0.76
195	Camposano	12.90	0.54
196	Capri	12.78	1.03
197	Carbonara di Nola	12.90	1.74
198	Cardito	11.34	0.86
199	Casalnuovo di Napoli	11.60	1.26
200	Casamarciano	12.82	0.58
201	Casamicciola Terme	13.58	1.51
202	Casandrino	10.66	1.05
203	Casavatore	13.47	0.88
204	Casola di Napoli	9.71	1.07
205	Casoria	13.18	0.71
206	Castellammare di Stabia	12.83	1.17
207	Castello di Cisterna	14.18	1.10
208	Cercola	12.81	0.65
209	Cicciano	11.22	0.86
210	Cimitile	12.33	0.62
211	Comiziano	11.55	1.61
212	Crispano	10.39	1.18
213	Forio	11.20	2.53

Continued on next page

Table B.3 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
214	Frattamaggiore	10.85	1.22
215	Frattaminore	12.11	0.88
216	Giugliano in Campania	11.65	1.01
217	Gragnano	14.06	0.91
218	Grumo Nevano	13.14	0.86
219	Ischia	10.74	1.91
220	Lacco Ameno	15.14	2.24
221	Lettere	10.97	1.33
222	Liveri	13.60	1.21
223	Marano di Napoli	14.30	1.49
224	Mariglianella	13.13	0.84
225	Marigliano	12.29	1.19
226	Massa Lubrense	13.15	0.89
227	Melito di Napoli	11.57	1.26
228	Meta	14.75	1.03
229	Monte di Procida	13.86	0.96
230	Mugnano di Napoli	12.57	0.81
231	Napoli	16.82	0.70
232	Nola	13.52	0.55
233	Ottaviano	13.38	0.65
234	Palma Campania	13.48	0.82
235	Piano di Sorrento	15.16	0.86
236	Pimonte	13.56	1.31
237	Poggiomarino	11.36	1.04
238	Pollena Trocchia	13.78	1.76
239	Pomigliano d'Arco	13.61	0.72
240	Pompei	14.11	1.95
241	Portici	14.25	0.81
242	Pozzuoli	14.10	1.43
243	Procida	11.54	1.60
244	Qualiano	10.96	0.99
245	Quarto	14.97	1.20
246	Ercolano	13.28	0.94
247	Roccarainola	11.41	0.73
248	San Gennaro Vesuviano	13.12	0.74
249	San Giorgio a Cremano	14.89	0.83
250	San Giuseppe Vesuviano	12.79	0.61
251	San Paolo Bel Sito	14.66	0.82
252	San Sebastiano al Vesuvio	14.95	1.02
253	Sant'Agnello	15.12	0.97
254	Sant'Anastasia	13.49	0.74

Continued on next page

Table B.3 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
255	Sant'Antimo	11.24	0.86
256	Sant'Antonio Abate	11.79	1.02
257	San Vitaliano	13.10	0.75
258	Saviano	12.24	0.77
259	Scisciano	12.84	1.01
260	Serrara Fontana	11.80	1.57
261	Somma Vesuviana	12.16	0.79
262	Sorrento	15.42	1.08
263	Striano	9.60	0.94
264	Terzigno	13.48	1.37
265	Torre Annunziata	14.87	1.40
266	Torre del Greco	14.50	1.42
267	Tufino	12.01	1.37
268	Vico Equense	10.85	1.00
269	Villaricca	12.30	0.91
270	Visciano	13.08	0.59
271	Volla	13.75	1.45
272	Santa Maria la Carita'	11.14	0.95
273	Trecase	11.88	1.19
274	Massa di Somma	12.59	0.64
275	Aiello del Sabato	13.30	1.22
276	Altavilla Irpina	11.22	1.13
277	Andretta	12.90	1.37
278	Aquilonia	11.99	1.29
279	Ariano Irpino	11.28	1.13
280	Atripalda	14.13	1.26
281	Avella	12.16	1.14
282	Avellino	17.69	1.09
283	Bagnoli Irpino	11.59	0.96
284	Baiano	13.66	1.50
285	Bisaccia	13.64	1.35
286	Bonito	13.58	0.98
287	Cairano	13.61	1.82
288	Calabritto	12.40	1.21
289	Calitri	13.22	1.17
290	Candida	10.33	1.53
291	Caposele	12.71	0.76
292	Capriglia Irpina	13.34	1.14
293	Carife	13.22	1.47
294	Casalbore	12.88	0.94
295	Cassano Irpino	13.87	1.72

Continued on next page

Table B.3 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
296	Castel Baronia	12.02	1.27
297	Castelfranci	11.57	1.35
298	Castelvetere sul Calore	11.73	1.18
299	Cervinara	12.83	0.63
300	Cesinali	12.97	1.34
301	Chianche	15.05	1.37
302	Chiusano di San Domenico	12.26	0.85
303	Contrada	13.23	0.61
304	Conza della Campania	12.81	1.47
305	Domicella	11.60	1.58
306	Flumeri	10.42	1.33
307	Fontanarosa	13.47	1.12
308	Forino	12.80	0.58
309	Frigento	12.41	0.96
310	Gesualdo	13.41	1.03
311	Greci	11.36	1.75
312	Grottaminarda	13.18	0.93
313	Grottolella	11.65	1.21
314	Guardia Lombardi	12.05	1.54
315	Lacedonia	14.67	1.17
316	Lapio	13.10	1.36
317	Lauro	15.65	1.00
318	Lioni	12.33	0.88
319	Luogosano	11.62	1.10
320	Manocalzati	11.54	0.82
321	Marzano di Nola	12.01	0.74
322	Melito Irpino	10.99	1.27
323	Mercogliano	12.15	1.44
324	Mirabella Eclano	12.60	0.99
325	Montaguto	14.53	2.47
326	Montecalvo Irpino	12.33	1.24
327	Montefalcione	11.72	0.82
328	Monteforte Irpino	11.67	1.23
329	Montefredane	12.45	0.76
330	Montefusco	14.22	1.71
331	Montella	12.48	1.80
332	Montemarano	11.16	1.37
333	Montemiletto	9.76	1.26
334	Monteverde	12.83	2.03
335	Montoro Inferiore	12.43	0.89
336	Montoro Superiore	11.99	0.91

Continued on next page

Table B.3 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
337	Morra De Sanctis	10.70	1.89
338	Moschiano	13.89	1.12
339	Mugnano del Cardinale	13.16	0.82
340	Nusco	11.02	1.22
341	Ospedaletto d'Alpinolo	12.51	1.39
342	Pago del Vallo di Lauro	12.50	1.02
343	Parolise	11.84	1.10
344	Paternopoli	13.58	1.12
345	Petruro Irpino	12.98	1.74
346	Pietradefusi	11.74	1.37
347	Pietrastornina	12.46	1.62
348	Prata di Principato Ultra	12.47	1.84
349	Pratola Serra	12.06	0.73
350	Quadrelle	11.85	1.28
351	Quindici	14.91	1.56
352	Roccabascerana	12.17	0.97
353	Rocca San Felice	9.31	1.88
354	Rotondi	12.58	0.75
355	Salza Irpina	14.37	1.42
356	San Mango sul Calore	13.67	1.41
357	San Martino Valle Caudina	12.32	0.85
358	San Michele di Serino	12.52	1.06
359	San Nicola Baronia	15.09	1.44
360	San Potito Ultra	14.21	0.90
361	San Sossio Baronia	12.03	1.22
362	Santa Lucia di Serino	13.35	1.11
363	Sant'Andrea di Conza	17.57	1.74
364	Sant'Angelo all'Esca	14.69	1.34
365	Sant'Angelo a Scala	9.94	2.36
366	Sant'Angelo dei Lombardi	11.82	1.08
367	Santa Paolina	13.16	0.91
368	Santo Stefano del Sole	11.78	0.90
369	Savignano Irpino	11.73	2.01
370	Scampitella	12.38	1.01
371	Senerchia	8.92	1.88
372	Serino	12.11	0.75
373	Sirignano	12.88	1.33
374	Solofra	13.68	1.14
375	Sorbo Serpico	11.39	1.26
376	Sperone	13.01	0.70
377	Sturno	12.24	0.90

Continued on next page

Table B.3 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
378	Summonte	13.87	0.81
379	Taurano	13.38	1.34
380	Taurasi	10.41	2.20
381	Teora	11.52	1.75
382	Torella dei Lombardi	11.62	1.07
383	Torre Le Nocelle	12.35	1.24
384	Torrioni	12.60	1.38
385	Trevico	12.02	0.96
386	Tufo	14.20	1.28
387	Vallata	13.32	1.10
388	Vallesaccarda	10.65	1.86
389	Venticano	13.29	0.99
390	Villamaina	11.61	1.09
391	Villanova del Battista	8.76	2.13
392	Volturara Irpina	11.41	1.70
393	Zungoli	13.59	1.28
394	Acerno	14.19	1.27
395	Agropoli	12.29	1.21
396	Albanella	12.79	1.07
397	Alfano	12.78	1.45
398	Altavilla Silentina	12.19	1.16
399	Amalfi	15.02	0.95
400	Angri	14.11	0.69
401	Aquara	13.28	1.70
402	Ascea	10.50	1.71
403	Atena Lucana	13.52	1.05
404	Atrani	16.84	2.22
405	Auletta	11.70	1.25
406	Baronissi	12.07	1.02
407	Battipaglia	14.14	0.88
408	Bellosguardo	13.28	1.74
409	Bracigliano	13.05	0.74
410	Buccino	11.70	1.24
411	Buonabitacolo	14.05	1.44
412	Caggiano	11.23	1.24
413	Calvanico	12.85	0.96
414	Camerota	10.88	1.70
415	Campagna	11.22	0.87
416	Campora	9.81	2.66
417	Cannalonga	11.59	1.90
418	Capaccio	11.10	1.23

Continued on next page

Table B.3 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
419	Casalbuono	15.15	1.66
420	Casaletto Spartano	9.15	2.11
421	Casal Velino	11.32	1.66
422	Caselle in Pittari	11.96	1.37
423	Castelcivita	11.30	1.70
424	Castellabate	10.25	1.59
425	Castelnuovo Cilento	12.33	1.00
426	Castelnuovo di Conza	9.47	1.84
427	Castel San Giorgio	12.99	0.82
428	Castel San Lorenzo	13.91	1.45
429	Castiglione del Genovesi	11.51	1.60
430	Cava de’ Tirreni	15.48	0.98
431	Celle di Bulgheria	12.22	1.45
432	Centola	10.16	1.63
433	Ceraso	13.72	1.29
434	Cetara	13.24	1.47
435	Cicerale	11.58	1.66
436	Colliano	11.68	1.25
437	Conca dei Marini	10.79	1.74
438	Controne	12.94	1.21
439	Contursi Terme	12.01	0.91
440	Corbara	11.20	1.15
441	Corleto Monforte	7.04	4.95
442	Cuccaro Vetere	11.73	1.79
443	Eboli	12.73	0.84
444	Felitto	13.02	1.28
445	Fisciano	12.26	0.71
446	Furore	11.44	2.65
447	Futani	11.69	1.89
448	Giffoni Sei Casali	11.92	0.89
449	Giffoni Valle Piana	12.40	0.97
450	Gioi	8.51	4.65
451	Giungano	13.08	0.82
452	Ispani	10.48	2.10
453	Laureana Cilento	11.99	1.56
454	Laurino	13.43	1.29
455	Laurito	13.48	1.66
456	Laviano	11.78	1.09
457	Lustra	13.49	1.06
458	Magliano Vetere	11.03	1.74
459	Maiori	14.47	1.21

Continued on next page

Table B.3 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
460	Mercato San Severino	12.83	0.60
461	Minori	14.12	1.08
462	Moio della Civitella	12.12	2.50
463	Montano Antilia	10.00	3.39
464	Montecorice	9.12	2.12
465	Montecorvino Pugliano	11.13	1.72
466	Montecorvino Rovella	13.10	0.94
467	Monteforte Cilento	10.08	2.18
468	Monte San Giacomo	13.27	1.66
469	Montesano sulla Marcellana	13.22	1.41
470	Morigerati	13.08	2.06
471	Nocera Inferiore	15.10	0.84
472	Nocera Superiore	13.80	0.84
473	Novi Velia	14.27	1.19
474	Ogliastro Cilento	14.65	1.34
475	Olevano sul Tusciano	13.20	0.76
476	Oliveto Citra	11.12	1.15
477	Omignano	12.13	1.10
478	Orria	13.30	1.61
479	Ottati	12.21	1.69
480	Padula	11.66	1.21
481	Pagani	14.90	0.93
482	Palomonte	7.89	2.27
483	Pellezzano	13.18	0.74
484	Perdifumo	13.16	1.25
485	Perito	15.12	1.35
486	Pertosa	14.03	1.30
487	Petina	14.34	1.00
488	Piaggine	14.85	2.06
489	Pisciotta	12.08	1.46
490	Polla	12.31	0.95
491	Pollica	11.12	1.79
492	Pontecagnano Faiano	13.33	1.11
493	Positano	12.38	1.29
494	Postiglione	11.81	1.20
495	Praiano	12.37	1.30
496	Prignano Cilento	11.51	1.22
497	Ravello	13.99	1.11
498	Ricigliano	8.72	2.62
499	Roccadaspide	11.78	1.17
500	Roccagloriosa	12.11	1.48

Continued on next page

Table B.3 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
501	Roccapiemonte	13.83	0.74
502	Rofrano	12.77	1.81
503	Romagnano al Monte	7.91	2.48
504	Roscigno	14.45	1.54
505	Rutino	13.07	0.93
506	Sacco	14.85	1.78
507	Sala Consilina	12.21	1.02
508	Salento	14.61	1.03
509	Salerno	17.88	1.67
510	Salvitelle	12.00	1.49
511	San Cipriano Picentino	12.84	1.45
512	San Giovanni a Piro	11.52	1.83
513	San Gregorio Magno	11.37	1.10
514	San Mango Piemonte	12.64	1.16
515	San Marzano sul Sarno	13.25	0.72
516	San Mauro Cilento	11.51	2.09
517	San Mauro la Bruca	11.43	2.45
518	San Pietro al Tanagro	13.95	1.26
519	San Rufo	13.07	1.01
520	Santa Marina	12.43	1.83
521	Sant'Angelo a Fasanella	13.64	1.46
522	Sant'Arsenio	14.82	1.07
523	Sant'Egidio del Monte Albino	12.99	0.71
524	Santomenna	11.15	1.87
525	San Valentino Torio	11.84	0.68
526	Sanza	12.21	1.25
527	Sapri	15.06	1.85
528	Sarno	13.83	0.70
529	Sassano	12.76	1.50
530	Scafati	11.60	1.33
531	Scala	12.16	0.87
532	Serramezzana	12.75	1.07
533	Serre	10.92	1.19
534	Sessa Cilento	13.32	1.22
535	Siano	15.94	1.20
536	Sicignano degli Alburni	12.65	0.88
537	Stella Cilento	12.71	1.34
538	Stio	12.20	1.32
539	Teggiano	13.58	1.31
540	Torchiara	9.95	2.04
541	Torraca	13.30	1.59

Continued on next page

Table B.3 – continued from previous page

	MUNICIPALITY	MEAN	RMSE
542	Torre Orsaia	12.27	1.28
543	Tortorella	14.30	2.03
544	Tramonti	12.29	1.20
545	Trentinara	11.81	1.31
546	Valle dell'Angelo	10.37	2.22
547	Vallo della Lucania	14.59	1.10
548	Valva	10.58	1.33
549	Vibonati	12.86	1.73
550	Vietri sul Mare	14.49	0.96
551	Bellizzi	13.48	1.23